ISCB44

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

MILAN, ITALY | 27 – 31 AUGUST 2023

University of Milano-Bicocca
Building U6 Piazza dell'Ateneo Nuovo 1

Final Programme & Abstract Book

Organising Secretariat
Promoest srl, Via G. Moroni 33, 20146 Milano (MI)
iscb2023@promoest.com

44th Annual Conference
of the International Society
for Clinical Biostatistics

Joint conference with the
Italian Region of the
International Biometric Society

The Local Organising Committee of the 44th Conference
of the International Society for Clinical Biostatistics (ISCB)
Joint conference with the Italian Region of the International Biometric Society (IBS)
would like to thank its sponsors for their support.

Host Partners:

Platinum sponsors:

Silver sponsors:

Bronze sponsors:

Supporter:

Welcome from LOC and SPC Chair

Welcome

FROM LOC AND SPC CHAIR

The city of Milano and the University of Milano-Bicocca are proud to host the 44th annual conference of ISCB that will be joint with the IBS-Italian Region of the International Biometric Society.
The Scientific Programme Committee (SPC) and the Local Organising Committee (LOC) have been working together to create a stimulating scientific and social programme. ISCB 2023 will span five days and will feature outstanding speakers, leading scientists and young people at the beginning of their career presenting their papers and joining discussions.

Five pre-conference courses (3 half-day and 2 full-day) given by recognized experts on the analysis of genomic data, prediction models, estimands and analyses of longitudinal continuous outcomes, innovative epidemiological designs and pseudo observations in survival analysis will start the conference activities.

In the plenary sessions two highly distinguished speakers will address topics hotly debated in modern biostatistics and epidemiology. The President's invited speaker Vanessa Didelez (University of Bremen, D) will focus on causal inference, estimands and trials in epidemiology and biostatistics, while the Keynote Speaker Lisa McShane (National Cancer Institute, US) will share the experience related to her statistical adventures in pursuit of precision medicine.

The 8 invited sessions will focus on key topics such as recent advances in high dimensional data analysis, causal inference, predictive algorithms and models, complex study protocols, efficient and innovative designs in clinical trials and observational studies, analysis of longitudinal data, quantification of effects and safety signals in clinical trials, and design and evaluation of vaccination programmes. In addition to 26 Invited presentations, more than 200 abstracts will be presented within 35 Oral Contributed Sessions on a wide range of statistical topics and more than 300 posters will be available for viewing during the entire conference.

The conference will also feature additional activities for young researchers for their networking and career development such as an early career biostatistician day, a session with the most brilliant students presenting their awarded work and a student gathering. The Early Career Researchers' Day will provide a unique opportunity for graduate students, post-doctoral fellows and biostatisticians at the beginning of their career to share experiences and advice, discuss opportunities and challenges, and practise their presentation skills in a less formal environment.

Two mini-symposia will complete the programme. One will be devoted to celebrate the 10th anniversary of the STRATOS Initiative (one day) and one will address novel approaches to cozmplex data and predictive modeling in healthcare research (half-day).

This ISCB conference will be hybrid and we have worked hard to make the conference accessible, interactive and enjoyable for both online and in-person attendance. For those attending online we are streaming the full scientific programme live through our conference platform and the interactions between (virtual and in-person) delegates will be possible through the conference app.

The conference takes place at the University of Milano-Bicocca, a young (now celebrating its 25th anniversary), dynamic, multi-disciplinary research oriented university. In-person attendees will also discover a vibrant city. Indeed, Milano, with its centuries of history and art, is a world-class economic and cultural hub, a capital of business, history, art, fashion, design and culinary delights to suit every taste. Its extraordinary aesthetic traditions have been the humus for new design projects. The city also enjoys an enviable setting in Lombardy. Filled with enchanting art cities, this is the Italian region with the largest natural parks and splendid lakes - like Como – surrounded by mountains and offers many opportunities for delegates. The in-person delegates will also enjoy the reception in the cloisters of Società Umanitaria, a former franciscan monastery, while the social dinner will be at the Villa Reale of Monza, a historical neoclassical building in the city near Milano also famous for the Italian Formula 1 Grand Prix.

The members of the SPC and the LOC welcome you to the ISCB44 conference in Milano!

MARIA GRAZIA VALSECCHI
Chair of the Local Organising Committee

STEFANIA GALIMBERTI
Chair of the Scientific Programme Committee

Index

Index

Index

ISCB44

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

MILAN, ITALY | 27 – 31 AUGUST 2023

University of Milano-Bicocca
Building U6 Piazza dell'Ateneo Nuovo 1

Organising Secretariat
Promoest srl, Via G. Moroni 33, 20146 Milano (MI)
iscb2023@promoest.com

Organisation – Committees of ISCB44 and It-IBS

Scientific Programme Committee (SPC)

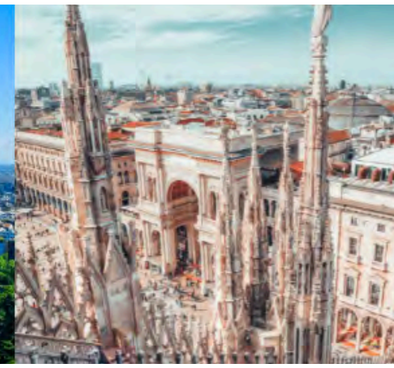
Stefania Galimberti (IT, Chair)
Claudia Angelini (IT)
Jonathan Bartlett (UK)
Stefano Calza (IT)
Clelia Di Serio (IT)
Malka Gorfine (IL)
Dario Gregori (IT)
Thomas Jaki (DE)
Thomas Lumley (NZ)
Rajarshi Mukherjee (US)
Cécile Proust-Lima (FR)
Marie Reilly (SE)
Kaspar Rufibach (CH)
Thomas Scheike (DK)
Ewout Steyerberg (NL)
Francesco Stingo (IT)
Giota Touloumi (GR)
James Wason (UK)
Emily C. Zabor (US)

Local Organising Committee (LOC)

Maria Grazia Valsecchi | Chair
Laura Antolini | Vice-chair
Paola Rebora | Treasurer
Davide Bernasconi
Valeria Edefonti
Aldo Solari

Mini-Symposia Committee (MSC)

Marco Bonetti | Chair
Federico Ambrogi
Emanuele Di Angelantonio
Stefania Galimberti
Mauro Gasparini
Maria Luisa Restaino



About ISCB

The International Society for Clinical Biostatistics (ISCB) was founded in 1978 to stimulate research into the principles and methodology used in the design and analysis of clinical research and to increase the relevance of statistical theory to the real world of clinical medicine.

Membership is open to all interested individuals who share the Aims of the Society. ISCB's membership includes clinicians, statisticians and members of other disciplines, such as epidemiologists, clinical chemists and clinical pharmacologists, working or interested in the field of clinical biostatistics.

ISCB has an Executive Committee and 6 Subcommittees: Conference Organising, Early Career Biostatisticians, Education, National Groups, Statistics in Regulatory Affairs (SiRA), Student Conference Awards.

Executive Committee

Tomasz Burzykowski (BE) | President
Thomas Jaki (UK) | Vice President
Elaine Pascoe (AU) | Secretary
Chris Metcalfe (UK) | Treasurer
Zdenek Valenta (CZ) | Past President (2023)
Charlotte Bolch (US) | Member & Media & Communications Coordinator
Kinga Salapa (PL) | Member
Laure Wynants (NL) | Member
Milada Cvancarova Småstuen (NO) | Member
Nan van Geloven (NL) | Member
Nikos Pantazis (GR) | Member
Paola Rebora (IT) | Member
Philip S Boonstra (US) | Member

Honorary Members

Michal Abrahamowicz (CA)
Harbajan Chadha-Boreham (FR)
Lutz Edler (DE)
Emmanuel Lesaffre (BE)
Michael Schemper (AT)
Martin Schumacher (DE)
Jørgen Seldrup (FR)
Stephen J. Senn (UK)
Hans C. van Houwelingen (NL)
David W. Warne (CH)
John R. Whitehead (UK)

ISCB Administrative seat Office

convin

29 K. Varnali Street, 152 33
Chalandri,
Athens, Greece
Call center | +30 210 6833600
Mobile contact | +30 6956 665669
Website | www.iscb.international
Email | office@iscb.international

About It-IBS

The Italian Region of the International Biometric Society (It-IBS) was founded in 1953 with the aim to promote the development, application and dissemination of statistical and mathematical theory and methods in the biosciences, including agriculture, biomedical science and public health, ecology, environmental sciences, and allied disciplines. The Society welcomes as members statisticians, mathematicians, biologists, and others devoted to interdisciplinary efforts in advancing biosciences. The community provides a forum for methodologic research and for discussion on the application of sound scientific methods.

The Italian Region of the International Biometric Society (SIB - Società Italiana di Biometria) now has as its main objective the cross-sectional and synergistic integration of biostatistics with biomedical and statistical disciplines. The mission of the Society is to promote the quality of educational and research initiatives in which biostatistics is at the center of a network of cooperation between Universities, IRCCS, Hospitals and the Biomedical Industry at the national and international level in order to foster the spread of biostatistical and biometric culture as a specific and highly specialized profession. This figure embraces in its most modern declinations, various fields, from statistical-computational methods in biomedicine and related sciences (e.g., environmental sciences, veterinary medicine) , to methodology for bioinformatic data in genetics and genomics. A non-secondary objective of the Italian Region is integration with other European and Mediterranean IBS regions and with scientific societies at the national level.

Executive Committee

Stefania Galimberti | President
Marialuisa Restaino | Secretary and Educational Activities
Paola Rebora | Treasurer
Stefano Calza | Webmaster
Francesco Stingo | Social Media and Communication

2023 Awards & Support Scheme Recipients

Student Conference Award (StCA)	Presentation title	Session code	University/Organisation	Country
Jiaxin Zhang	**TO3.1** Sensitivity analysis for missingness assumptions in causal inference: accommodating the substantive analysis	**TO3.1**	Murdoch Children's Research Institute	Australia
Léonie Courcoul	**TO3.2** A location-scale joint model with a time- dependent subject- specific variance of the marker and competing event	**TO3.2**	University of Bordeaux	France
Zhenwei Yang	**TO3.3** A bayesian joint modelling for misclassified interval- censoring and competing risks	**TO3.3**	Erasmus University Medical Center	Netherlands
Pedro Miranda Afonso	**TO3.4** A joint model for (un)bounded longitudinal markers, competing risks, and recurrent events using registry dat	**TO3.4**	Erasmus University Medical Center	Netherlands
Joshua Philipp Entrop	**TO3.5** Parametric estimation of the mean number of events in the presence of competing risks	**TO3.5**	Karolinska Institute	Sweden
Conference Award for Scientists (CASc)	**Presentation title**	**Session code**	**University/Organisation**	**Country**
Nofel Ahmed	**MO8.1** Gradient boosting for survival analysis with competing risks	**MO8.1**	University of Dhaka	Bangladesh
Oya Kalaycioglu	**MO11.3** Evaluating the sample size requirements of tree-based machine learning techniques for clinical risk prediction	**MO11.3**	Izzet Baysal University	Turkey
Samia Ashrafi	**MP122** Detection of multiple change points in survival analysis with narrowest significant pursuit technique	**MP122**	University of Dhaka	Bangladesh
Conference Fund for Devloping Countries (CFDC)	**Presentation title**	**Session code**	**University/Organisation**	**Country**
Mutlu Altuntas	**MP1** On detecting outliers in a circular-linear regression by cephalometric angles data	**MP1**	Sinop University	Turkey
Adnan Karaibrahimoglu	**MP3** On the comparison of classical and wrapped cauchy circular regression in patients with gastric carcinoma	**MP3**	Suleyman Demirel University	Turkey
Xiang Geng	**MP27** Sample size determination based on restricted mean time lost in the presence of competing risks	**MP27**	Southern Medical University	China
Zhiyin Yu	**MP31** Restricted mean time lost model for covariates with time-varying effects	**MP31**	Southern Medical University	China
Elvis Karanja Muchene	**MP91** Real time prediction of infectious disease outbreaks based on google trend data in africa	**MP91**	University of Nairobi	Kenya
Zijing Yang	**MP103** Dynamic prediction based on conditional restricted mean survival time for right-censored data	**MP103**	Stomatological Hospital, Southern Medical University	China
Md. Abdul Basit	**WP12** A calibrated sensitivity model for observational studies with multivalued treatments	**WP12**	Institute of Statistical Research & Training (ISRT), University of Dhaka	Bangladesh
Priom Saha	**WP23** Sensitivity analysis with matched pairs from observational studies	**WP23**	Institute of Statistical Research & Training (ISRT)	Bangladesh
Muhammad Iftakhar Alam	**WP37** A combined criterion for dose finding in phase i clinical trials	**WP37**	University of Dhaka	Bangladesh
El Badisy Imad	**WP74** Comparison of imputation methods in the presence of time-varying and non-linear covariates effects	**WP74**	University of Aix-Marseille, FR	Morocco

Keynote Speakers

Keynote Speaker

Lisa McShane, PhD

Associate Director, Division of Cancer Treatment & Diagnosis
Chief, Biometric Research Program
National Cancer Institute, NIH
Bethesda, US

Monday 28 August 2023 | 9:30-10:30
Statistical adventures in pursuit of precision medicine: secret signatures, sliding subgroups & more

President's Invited Speaker

Vanessa Didelez

Professor of Statistics with Focus on Theory and Methods for Causal Inference
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Bremen, D

Wednesday 30 August 2023 | 11:10-12:15
On causal inference, estimands and trials in epidemiology and biostatistics

Pre-conference courses

SUNDAY | 27 August 2023

	ROOM U6.1a	ROOM U6.1b	ROOM U6.1e	ROOM U6.1f
9:00-10:30		Estimands and analyses of longitudinal continuous outcomes in clinical trials Teachers: Marcel Wolbers, Alessandro Noci	Beyond classic epidemiological designs Teachers: Marie Reilly, Paola Rebora, Francesca Graziano	Pseudo observations in survival analysis Teachers: Per Kragh Andersen, Henrik Ravn
10:30-11:00	COFFEE BREAK			
11:00-12:30		Estimands and analyses of longitudinal continuous outcomes in clinical trials Teachers: Marcel Wolbers, Alessandro Noci	Beyond classic epidemiological designs Teachers: Marie Reilly, Paola Rebora, Francesca Graziano	Pseudo observations in survival analysis Teachers: Per Kragh Andersen, Henrik Ravn
12:30-13:30	LUNCH BREAK			
13:30 - 15:00	Evaluation of prediction models: from AUC to calibration and decision curve analysis Teachers: Ewout Steyerberg, Ben van Calster, Ed Bonneville	Analysis of genomic data: R Bioconductor Teacher: Davide Risso	Beyond classic epidemiological designs Teachers: Marie Reilly, Paola Rebora, Francesca Graziano	Pseudo observations in survival analysis Teachers: Per Kragh Andersen, Henrik Ravn
15:00-15:30	COFFEE BREAK			
15:30 - 17:00	Evaluation of prediction models: from AUC to calibration and decision curve analysis Teachers: Ewout Steyerberg, Ben van Calster, Ed Bonneville	Analysis of genomic data: R Bioconductor Teacher: Davide Risso	Beyond classic epidemiological designs Teachers: Marie Reilly, Paola Rebora, Francesca Graziano	Pseudo observations in survival analysis Teachers: Per Kragh Andersen, Henrik Ravn
19:00-21:00	Students/ECB gathering at Fuorimano OTBP			

Conference Days

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY
	28 AUGUST 2023	29 AUGUST 2023	30 AUGUST 2023	31 AUGUST 2023
9:00 -10:30	GREAT HALL Conference opening Keynote speaker: Lisa McShane	Invited session and parallel contributed sessions	Invited session and parallel contributed sessions	Mini-symposia 1 and 2 and ECB
10:30-11:00	COFFEE BREAK			
11:00 - 12:30	Invited session and parallel contributed sessions	Invited session and parallel contributed sessions	President's Invited Speaker: Vanessa Didelez	Mini-symposia 1 and 2 and ECB
12:30-13:30	LUNCH BREAK		LUNCH BREAK AND AGM ISCB AT ROOM U6.6	LUNCH BREAK AND AGM IT-IBS AT ROOM U6.10
13:30 - 15:00	Invited session and parallel contributed sessions	EXCURSIONS	Invited session and parallel contributed sessions	Mini-symposium 1 (STRATOS)
15:00-15:30	COFFEE BREAK		COFFEE BREAK	
15:30 - 17:00	Invited session and parallel contributed sessions		Invited session and parallel contributed sessions	Mini-symposium 1 (STRATOS)
19:00-21:00	Welcome reception **CHIOSTRO DEI PESCI** Umanitaria di Milano Via S. Barnaba, 48 Milano (MI)		Conference dinner **VILLA REALE MONZA** Viale Brianza, 1 Monza (MZ)	

social events
parallel sessions
plenary lecture

ECB: early career biostatistician
AGM: annual general meeting

Conference Days

INVITED SESSIONS	MONDAY	TUESDAY	WEDNESDAY
	28 AUGUST 2023	29 AUGUST 2023	30 AUGUST 2023
9:00 -10:30	Conference opening Keynote speaker: Lisa McShane	Evaluation of predictive algorithms and models: uncertainty and impact onmedical care	Recurrent events and their use in medical studies
10:30-11:00	COFFEE BREAK		
11:00 - 12:30	Advances on causal inference in longitudinal studies	High-dimensional inference in biostatistics	President's Invited Speaker: Vanessa Didelez
12:30-13:30	LUNCH BREAK		
13:30 - 15:00	Innovative designs for dose optimization studies	EXCURSIONS	Quantification of safety signals in clinical trials: Estimand, estimation, and how would good look like in ten years?
15:00-15:30	COFFEE BREAK		COFFEE BREAK
15:30 - 17:00	Vaccination programmes: post-implementation assessment of protection, benefits and risks		Marginal versus conditional effects in clinical trials
19:00-21:00	Welcome reception **CHIOSTRO DEI PESCI** Umanitaria di Milano Via S. Barnaba, 48 Milano (MI)		Conference dinner **VILLA REALE MONZA** Viale Brianza, 1 Monza (MZ)

social events
parallel sessions
plenary lecture

MONDAY | 28 August 2023

	ROOM U6.6	ROOM U6.7	ROOM U6.8	ROOM U6.9	ROOM U6.10	ROOM U6.11	GREAT HALL INVITED SESSIONS
9:00 - 10:30	**GREAT HALL** CONFERENCE OPENING PLE1: Plenary Lecture 1: Keynote Speaker: Lisa McShane "Statistical adventures in pursuit of precision medicine: secret signatures, sliding subgroups & more"						
10:30 - 11:00	COFFEE BREAK						
11:00 - 12:30	**MO1** Clinical Trials 1	**MO2** Survival analysis 1	**MO3** Prediction models 1	**MO4** Biomarkers	**MO5** High dimensional data 1	**MO6** Epidemiology 1	**MIN 1** Advances on causal inference in longitudinal studies
12:30 - 13:30	LUNCH BREAK						
13:30 - 15:00	**MO7** Clinical Trials 2	**MO8** Survival analysis 2	**MO9** Latent variable modelling	**MO10** Causal inference 1	**MO11** Machine learning 1	**MO12** Real world data	**MIN 2** Innovative designs for dose optimization studies
15:00 - 15:30	COFFEE BREAK						
15:30 - 17:00	**MO13** Clinical Trials 3	**MO14** Survival analysis 3	**MO15** Prediction medicine 1	**MO16** Longitudinal analysis 1	**MO17** High dimensional data 2	**MO18** Meta-Analysis	**MIN 3** Vaccination programmes: post-implementation assessment of protection, benefits and risks
19:00 - 21:00	Welcome reception – **CHIOSTRO DEI PESCI** Umanitaria di Milano						

TUESDAY | 29 August 2023

	ROOM U6.6	ROOM U6.7	ROOM U6.8	ROOM U6.9	ROOM U6.10	ROOM U6.11	GREAT HALL INVITED SESSIONS
9:00 - 10:30	**TO1** Clinical Trials 4	**TO2** Survival analysis 4	**TO3** Students Awardees	**TO4** Causal inference 2	**TO5** Machine learning 2		**TIN 1** Evaluation of predictive algorithms and models: uncertainty and impact on medical care
10:30 - 11:00	COFFEE BREAK						
11:00 - 12:30	**TO6** Clinical Trials 5	**TO7** Survival analysis 5	**TO8** Prediction models 2	**TO9** Longitudinal analysis 2	**TO10** Clinical Trials 6	**TO11** ItR-IBS & Italian Statistical Society	**TIN 2** High-dimensional inference in biostatistics
12:30 - 13:30	LUNCH BREAK						
13:30 - 18:00	EXCURSIONS						

WEDNESDAY | 30 August 2023

	ROOM U6.6	ROOM U6.7	ROOM U6.8	ROOM U6.9	ROOM U6.10	ROOM U6.11	GREAT HALL INVITED SESSIONS
9:00 - 10:30	**WO1** Clinical Trials	**WO2** Survival analysis 6	**WO3** Longitudinal analysis 3	**WO4** Causal inference 3	**WO5** Missing data	**WO6** Epidemiology 2	**WIN 1** Recurrent events and their use in medical studies
10:30 - 11:00	COFFEE BREAK						
11:00 - 13:30	**GREAT HALL** PLE2: Plenary Lecture 2: President's Invited Speaker: Vanessa Didelez **"On causal inference, estimands and trials in epidemiology and biostatistics"**						
12:30 - 13:30	LUNCH BREAK and Annual General Meeting (AGM) ISCB at ROOM U6.6						
13:30 - 15:00	**WO7** Clinical Trials 8	**WO8** 13:20-15:00 Dedicated to Prof. Ettore Marubini Survival analysis 7	**WO9** Prediction models 3	**WO10** Causal inference 4	**WO11** Machine learning 3	**WO12** Syntetic data 1	**WIN 2** Quantification of safety signals in clinical trials: estimand, estimation, and how would good look like in ten years?
15:00 - 15:30	COFFEE BREAK						
15:30 - 17:00	**WO13** Clinical Trials 9	**WO14** Miscellanea	**WO15** Prediction mdicine 2	**WO16** Longitudinal analysis 4	**WO17** High dimensional data 3		**WIN 3** Marginal versus conditional effects in clinical trials
19:00 - 21:00	Conference dinner: **VILLA REALE MONZA**, Monza (MB)						

44th Annual Conference
of the International Society
for Clinical Biostatistics

Joint conference with the
Italian Region of the
International Biometric Society

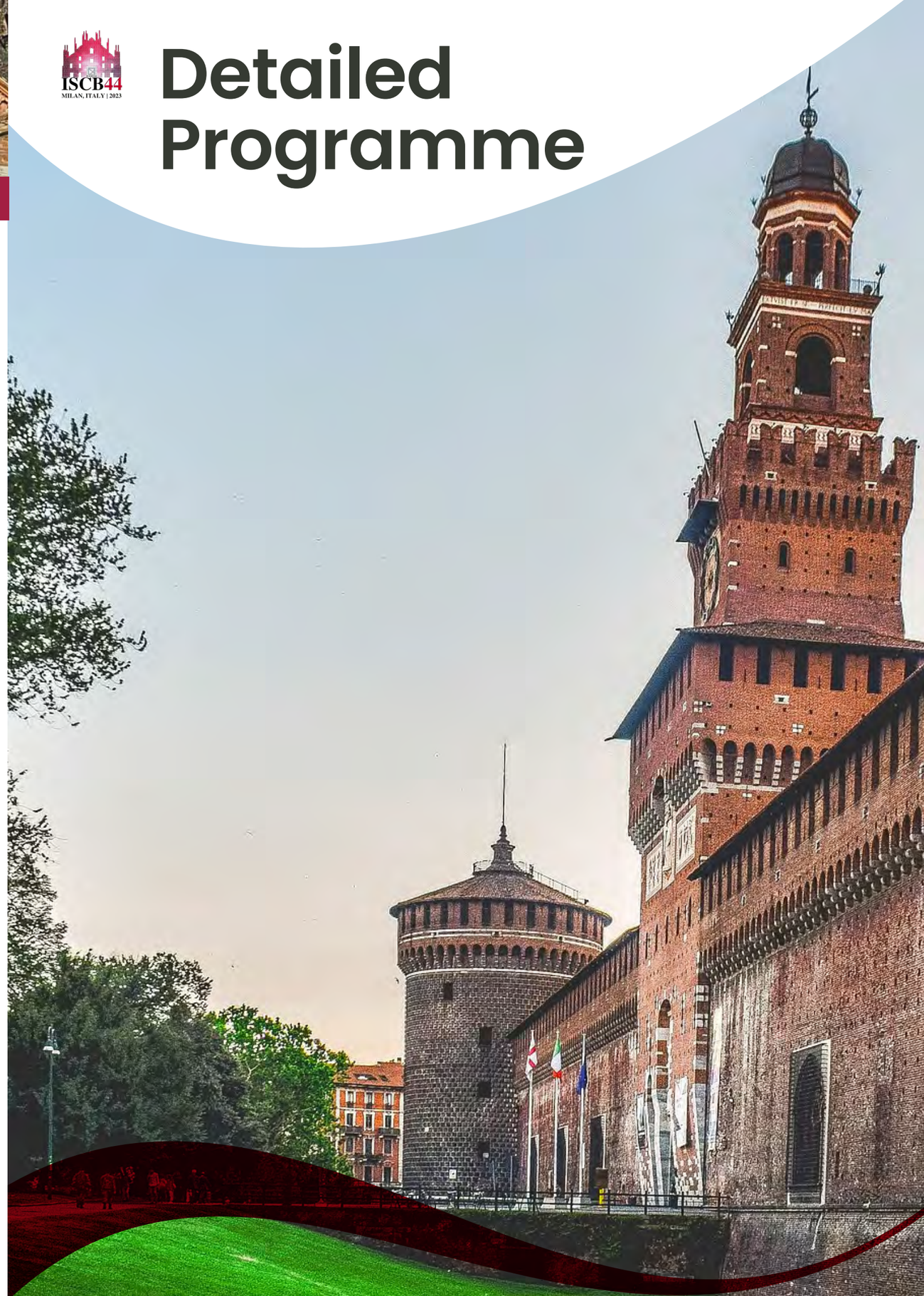
Table of contents

THURSDAY | 31 August 2023

Mini-symposia 1 and 2 and Early Career Biostatistician (ECB)

	ROOM U6.6	ROOM U6.8	ROOM U6.9
9:00 - 10:30	**TMS 1** **Mini-symposium 1:** Ten years STRATOS initiative – brief summary of progress and plans for the future	**TMS 2** **Mini-symposium 2:** Novel approaches to complex data and predictive modeling in healthcare research	**ECB** Early Career Biostatisticians' (ECB) Day
10:30 - 11:00	COFFEE BREAK		
11:00 - 12:30	**TMS 1** **Mini-symposium 1:** Ten years STRATOS initiative – brief summary of progress and plans for the future	**TMS 2** **Mini-symposium 2:** Novel approaches to complex data and predictive modeling in healthcare research	**TMS 3** Early Career Biostatisticians' (ECB) Day
12:30 - 13:30	LUNCH BREAK	ROOM U6.10 – 12:30 **GENERAL MEETING OF THE ITALIAN REGION OF IBS**	
13:30 - 15:10	**TMS 1** **Mini-symposium 1:** Ten years STRATOS initiative – brief summary of progress and plans for the future		
15:10 - 15:30	COFFEE BREAK		
15:30 - 17:00	**TMS 1** **Mini-symposium 1:** Ten years STRATOS initiative – brief summary of progress and plans for the future		

Detailed Programme



SUNDAY | 27 AUGUST 2023

ROOM U6.1a

13:30-17:00

PRE-CONFERENCE COURSE:
Evaluation of prediction models: from AUC to calibration and decision curve analysis

HALF DAY – AFTERNOON

Teachers: **Ewout Steyerberg**, LUMC Leiden
Ben van Calster, Ben Van Calster
Ed Bonneville, KU Leuven

TIMELINE		
13:30-15:00	►	**PART A**
15:00-15:30	►	COFFEE BREAK
15:30-17:00	►	**PART B**

ABSTRACT
Prediction models relate multiple patient or disease characteristics to diagnostic or prognostic outcomes. Prediction models enjoy increasing popularity with our increasing knowledge on markers, imaging and hopes for Artifical Intelligence / Machine Learning as flexible modeling tools. Validation of predictions from such models is essential, whether developed by classical statistical methods or novel AI approaches. Evaluation of performance commonly starts with an assessment of the discriminative ability of a prediction model, as often quantified by the Area under ROC curve (AUC) for binary outcomes. Other performance measures are popular among the AI community, such as the F1 score and Precision Recall Curve (PRC). In addition to discrimination, calibration is key to assess when predictions intend to be communicated to clinicians and patients, and support shared decision-making. Various summary measures are available to the analyst to complement graphical calibration assessment. Nowadays, further evaluation includes the Net Benefit, a summary measure that depends on the risk threshold for defining high versus low risk. Net Benefit is related to the threshold in a Decision Curve Analysis (DCA).

COURSE OUTLINE
1. Overview of traditional and modern performance measures for binary outcomes
2. Detailed overview of decision curve analysis to assess potential clinical utility
3. Discussion of different types of model validation, including choice of performance metrics
4. Extension of performance measures to nominal, ordinal, survival, and competing risk outcomes

LEARNING GOALS
• Obtain an up-to-date understanding of statistical performance measures for prediction models, and extensions to decision-analytic measures
• Know about measures for different types of outcomes, beyond binary outcomes: categorical; ordinal outcomes; survival; competing risks
• Learn how to estimate apparent, internally validated, and externally validated performance
• Be able to apply approaches in R

TARGET AUDIENCE
Applied biostatisticians interested in prediction research, diagnostic and prognostic modeling, validation concepts

ROOM U6.1b

9:00-12:30

PRE-CONFERENCE COURSE:
Estimands and analyses of longitudinal continuous outcomes in clinical trials

HALF DAY – MORNING

Teachers: **Marcel Wolbers**, Data & Statistical Sciences Department
Alessandro Noci, Pharma Development, Roche, Basel

TIMELINE		
9:00-10:30	►	**PART A**
10:30-11:00	►	COFFEE BREAK
11:00-12:30	►	**PART B**

ABSTRACT
For more than a decade, a mixed model for repeated measures (MMRM) approach has been the de facto standard for the primary analysis of clinical trials with longitudinal outcomes. The conventional MMRM model provides valid inference under a basic missing-at-random assumption. However, it is increasingly appreciated that more complex analysis methods may be required to fully align the analysis strategy with the targeted estimand and flexible missing data assumptions.

Course outline
This short course provides a thorough introduction to estimands and estimation strategies for clinical trials with longitudinal continuous outcomes. A particular focus will be on estimation methods based on multiple imputation of missing data and their implementation in the R package rbmi.

Specifically, the course will cover the following topics:

• Introduction to continuous longitudinal estimands, missing data assumptions (basic and extended MAR, reference-based missingness), and traditional analysis approaches.

• Estimators based on multiple imputation and their implementation in statistical software.

• Case study: Estimands and estimators in early Parkinson's disease.

After the course, attendees will understand the strengths and limitations of conventional MMRM analyses in view of the estimands framework. They will be familiar with alternative analysis methods based on multiple imputation which are of increasing importance for primary and sensitivity analyses. Finally, they will understand how such analyses can be implemented with statistical software.

TARGET AUDIENCE
Statisticians from the pharmaceutical industry and academic clinical research units. No prior knowledge except for basic familiarity with the statistical software R and the estimands framework as described in the "ICH E9(R1) addendum on estimands and sensitivity analyses in clinical trials" is required.

ROOM U6.1b

13:30-17:00

PRE-CONFERENCE COURSE:
Analysis of genomic data: R Bioconductor

HALF DAY – AFTERNOON

Teacher: **Davide Risso**, University of Padua

TIMELINE		
13:30-15:00	►	**PART A**
15:00-15:30	►	COFFEE BREAK
15:30-17:00	►	**PART B**

ABSTRACT
Omics data are quickly becoming ubiquitous in research and clinical studies, particularly in cancer, with several hospitals now routinely measuring genomic and transcriptomic profiles of cancer patients, in their efforts to move towards personalized medicine. While many bioinformatic and statistical methods exist to analyze such data, the size and complexity of omics data can be daunting for researchers approaching the field. Genomic data are characterized by high-dimensionality and complexity, technology-specific biases, and require domain-specific knowledge and bespoke informatic tools to be successfully analyzed.
This half-day course introduces participants to Bioconductor, an R-based open science and open development project for the analysis and comprehension of high-throughput biological data, such as RNA and DNA sequencing. We will introduce the Bioconductor project and how it relates to other R packages, explain how to work with DNA sequences in R, and how to analyze RNA sequencing datasets, using The Cancer Genome Atlas as an example. At the end of the course, participants should be able to explore DNA mutations and copy number alterations, and to perform and interpret a differential expression analysis.

COURSE OUTLINE
The course consists of two parts. The first session is a mix of theory and practice. Participants will be introduced to the basics of sequencing data and Bioconductor. We will also discuss how to perform a differential expression analysis and how to integrate genomic and clinical data. The second session will be hands-on, where participants apply what they have learned to a real dataset. There will be a brief rejoinder at the end of the session to discuss the analyses.
Specifically, the course will cover the following topics:
1. Data import and management in R/Bioconductor
2. Exploratory Data Analysis and Quality Control (EDA/QC)
3. Data normalization
4. Differential expression analysis
5. Integration of genomic and clinical data

By the end of this course, you should be able to:
• Have a working knowledge of DNA and RNA sequencing
• Perform an exploratory analysis of genomic and transcriptomic data
• Perform a differential expression analysis
• Interpret and visualize the results

TARGET AUDIENCE
The course is aimed at biostatisticians or medical researchers working with biological or clinical data that want to learn how to include genomics and transcriptomics data in their analyses. Participants are expected to have basic knowledge of the R statistical software; no prior knowledge of Bioconductor is required.

SUNDAY | 27 AUGUST 2023

ROOM U6.1e

9:00-17:00

PRE-CONFERENCE COURSE:
Beyond classic epidemiological designs

FULL DAY

Teachers: **Marie Reilly**, Karolinska Institutet, Sweden
Paola Rebora, University of Milano-Bicocca, Italy
Francesca Graziano, University of Milano-Bicocca, Italy

TIMELINE		
9:00-10:30	►	**PART A**
10:30-11:00	►	COFFEE BREAK
11:00-12:30	►	**PART B**
12:30-13:30	►	LUNCH BREAK
13:30-15:00	►	**PART C**
15:00-15:30	►	COFFEE BREAK
15:30-17:00	►	**PART D**

ABSTRACT
Recent years have seen an increase in population-based health registers, and clinical trial registers, as well as in biobanks, motivated by the increasing interest in investigating genomic or molecular biomarkers. While this is undoubtedly an area of great interest and expansion in medical research, the statistical analysis for exploring and developing new biomarkers often faces limited availability of biological samples. In this context, it is crucial to develop novel study designs, and associated statistical analysis methods, for a parsimonious use of available resources.
Epidemiologists and biostatisticians are familiar with matched and unmatched cohort and case-control designs, but there are numerous other subsampling designs that are useful, and perhaps more efficient, when a sample is selected from a well-defined cohort or population, such as an electronic register. Moreover, despite the availability of methods and software for implementing many of these designs, they are rarely used in practice.
The course begins with extensions to the classic designs for binary outcome from both a sampling and analysis perspective: intentional and unintentional two-stage designs, balanced and optimal sampling at the second stage. The cohort, nested case-control and case-cohort designs for time-to-event outcomes will be reviewed, their interrelationship explored through examples, and extensions to more efficient and optimal designs presented. The practical sessions will include computational exercises using statistical software (STATA and R). The course is based on the forthcoming book “Controlled Epidemiological Studies”.

COURSE OUTLINE
1. review of classic designs (cross sectional, cohort, case-control) matching, risk measures
2. two-stage design (recognition, examples and analysis)
3. balanced and optimal sampling of two-stage data in different settings
4. review of classic designs for time-to-event outcome and their interrelationship
5. novel applications of case-cohort and nested case-control designs
6. extended and optimal sampling for studies with time-to-event outcome

After this course, you should be able to:
- select & justify a suitable design for a specific research question concerning a binary and time-to-event outcome
- compare and interpret risk estimates from different sampling strategies
- recognise, analyse and design a two-stage study of a binary outcome
- implement and analyse a nested case-control or case-cohort study in various settings
- design and analyse a two-stage study of a time-to-event outcome

TARGET AUDIENCE
Applied biostatisticians / epidemiologists or graduate students familiar with logistic and Cox model.

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

SUNDAY | 27 AUGUST 2023

ROOM U6.1f

9:00-17:00 — **PRE-CONFERENCE COURSE:** Pseudo observations in survival analysis

FULL DAY

Teachers: **Per Kragh Andersen**, professor, Biostatistics, University of Copenhagen, Denmark
Henrik Ravn, senior statistical director, Novo-Nordisk A/S, Denmark

TIMELINE

9:00-10:30	►	**PART A**	**13:30-15:00**	►	**PART C**
10:30-11:00	►	COFFEE BREAK	15:00-15:30	►	COFFEE BREAK
11:00-12:30	►	**PART B**	**15:30-17:00**	►	**PART D**
12:30-13:30	►	LUNCH BREAK			

ABSTRACT

Pseudo observations (PO) have since their appearance in the biostatistical literature in 2003 been an active research field in survival analysis. PO now appear in SAS procedures, R packages, RCT protocols and textbooks.
The basic idea is that a random variable f(T) that is incompletely observed due to censoring is replaced by its PO for regression analysis of E(f(T)|Z) where T is the survival time and Z covariates. Here, the PO is obtained from an estimator of the marginal mean E(f(T)) that takes the censored data properly into account, such as the Kaplan-Meier estimator for P(T>t)=E(I(T>t)). Thereby, censoring is dealt with 'once and for all' and standard generalized estimating equations may be used with the PO as response variable
In the course, we will first give a brief recap of basic survival analysis, including the Kaplan-Meier estimator, the Cox model, and basics on competing risks and recurrent events. We will next introduce the PO and explain how they can be used for analysis of 'marginal' parameters in survival analysis. We will explain this in detail and show how the analysis may be performed using the R software. We will also briefly discuss the mathematical properties of PO methods.

COURSE OUTLINE

This full-day course will be at an intermediate level with emphasis on hands-on practicals and interpretation of analysis results.
1.1 Recap of survival data: Kaplan-Meier, hazard, Nelson-Aalen, Cox model, competing risks, cumulative incidence function, recurrent events mean value
1.2 R-exercises: Kaplan-Meier, Nelson-Aalen, Cox, competing risks on data examples
1.3 General definition and analyses of POs and specifically for I(T>t) based on Kaplan-Meier
1.4 R-exercises: Computations and analysis of PO for I(T>t) in one or more time points

2.1 PO generally: Competing risks, restricted mean survival time (RMST) and recurrent events
2.2 R-exercises: Computations and analysis of PO based on cumulative incidence function and RMST
2.3 Theoretical properties for POs

By the end of this course, you should be able to:

Knowledge. Participants should know what pseudo observations (PO) are and how they may be used for specific modeling purposes in survival analysis. This includes both practical knowledge about how to use pseudo observations when analyzing a given data set and theoretical knowledge about their mathematical properties.

Skills. Through exercises, participants will be able to compute and analyze PO using existing R packages.

Competences. Participants should be able to recognize analysis situations in which the use of PO may be beneficial and, subsequently, know how to carry out the analysis in practice.

TARGET AUDIENCE

Participants should be statisticians with a basic knowledge of 'standard' survival analysis, such as the Kaplan-Meier estimator, the Cox regression model and competing risks cumulative incidence function. Fundamental R knowledge is required, including how to install and apply new packages.

MONDAY | 28 AUGUST 2023

GREAT HALL

9:00 – 9:30 — **CONFERENCE OPENING**

Welcome from **Conference Local Organising Committee Chair**
Welcome from **University of Milano-Bicocca**
Welcome from **Lombardy Region and Milan representatives**
Presentation of **ISCB44 Conference Awards**

GREAT HALL

9:30 - 10:30 — PLE1: **KEYNOTE SPEAKER LECTURE**
Statistical adventures in pursuit of precision medicine: secret signatures, sliding subgroups & more

Introduction by: **Stefania Galimberti** (IT), *SPC Chair and It-IBS President*
Keynote Speaker: **Lisa McShane** (US) *Associate Director, Division of Cancer Treatment & Diagnosis Chief, Biometric Research Program National Cancer Institute, NIH Bethesda*

ROOM U6.6

11:00-12:30 — PARALLEL SESSION MO1: **Clinical Trials 1**

Chair: **James Wason** (UK)

11:00-11:18	**MO1.1**	Stella Jinran Zhan	*Adaptive seamless designs in the two-trial paradigm: advantages and limitations*
11:19-11:36	**MO1.2**	James Carpenter	*Power calculations for multi-arm multi-stage trials with multicomponent disease rating scales outcomes [+]*
11:37-11:54	**MO1.3**	Stephane Heritier	*Fast simulation of bayesian adaptive designs using the laplace approximation*
11:55-12:12	**MO1.4**	S. Faye Williamson	*A bayesian decision-theoretic randomisation procedure and the impact of delayed responses*
12:13-12:30	**MO1.5**	Aritra Mukherjee	*Evaluating the impact of outcome delay on the efficiency of two-arm group-sequential trials*

ROOM U6.7

11:00-12:30 PARALLEL SESSION MO2: **Survival Analysis 1**

Chair: **Marco Bonetti** (IT)

11:00-11:18	MO2.1	Angus Jennings	*Perils of rct survival extrapolation with effect waning: why marginal and conditional estimates differ*
11:19-11:36	MO2.2	Morten Valberg	*Can we trust the hazard ratio? -causal consequences of observed proportional hazards*
11:37-11:54	MO2.3	Lucia Ameis	*A non-parametric proportional risk model to assess a treatment effect in an application to survival data*
11:55-12:12	MO2.4	Marije Sluiskes	*A reduced rank proportional hazards model for age-related multimorbidity event data*
12:13-12:30	MO2.5	Ina Dormuth	*Multicasanova - multiple group comparisons for non-proportional hazard settings*

ROOM U6.8

11:00-12:30 PARALLEL SESSION MO3: **Prediction Model 1**

Chair: **Aldo Solari** (IT)

11:00-11:18	MO3.1	Matteo Rota	*Tuning the regularization parameter in penalized regression: an approach based on false selection rate*
11:19-11:36	MO3.2	Menelaos Pavlou	*Penalised regression methods with modified tuning produce better prediction models*
11:37-11:54	MO3.3	Frank Weber	*An introduction to projection predictive variable selection*
11:55-12:12	MO3.4	Nilufar Akbari	*Confidence interval estimation for selected and unselected predictors after variable selection*

ROOM U6.9

11:00-12:30 PARALLEL SESSION MO4: **Biomarkers**

Chair: **Paola Rebora** (IT)

11:00-11:18	MO4.1	Giulia Risca	*Validation of a skewed surrogate endpoint for a timeto-event outcome: the use of a zaga distribution*
11:19-11:36	MO4.2	Inna Chervoneva	*Metrics of spatial interaction between immune and cancer cells in tumor microenvironment as cancer biomarkers*
11:37-11:54	MO4.3	Chinyereugo Chikere	*Bayesian network meta-analysis of time-to-event data for evaluation of predictive biomarkers using ipd and ad*
11:55-12:12	MO4.4	Neal Alexander	*Agreement and error of titration assays*

MONDAY 28 AUGUST 2023

ROOM U6.10

11:00-12:30 PARALLEL SESSION MO5: **High dimensional data 1**

Chair: **Stefano Calza** (IT)

11:00-11:18	MO5.1	Andrea Bratsberg	*Conditional variable screening for ultra-high dimensional longitudinal data with time interactions*
11:19-11:36	MO5.2	Jeroen Goedhart	*Incorporating external information into bayesian additive regression trees using empirical bayes*
11:37-11:54	MO5.3	Nuria Senar	*A framework for interpretation and testing of sparse canonical correlation*
11:55-12:12	MO5.4	Wessel van Wieringen	*Two-dimensional fused targeted ridge estimation for health indicator prediction*
12:13-12:30	MO5.5	Kieran Baker	*Methodology for, and insight from, analysing displacement by tremor in parkinson's disease*

ROOM U6.11

11:00-12:30 PARALLEL SESSION MO6: **Epidemiology 1**

Chair: **Marie Reilly** (SE)

11:00-11:18	MO6.1	Mark Van de Wiel	*Linked shrinkage to improve the estimation of interaction effects in a regression model*
11:19-11:36	MO6.2	Apostolos Gkatzionis	*Novel insights for quantifying Selection bias using interactions On the log-additive scale;*
11:37-11:54	MO6.3	Satoshi Uno	*Investigating accuracy of disease outcome definition in pharmacoepidemiology bayesian latent class model*
11:55-12:12	MO6.4	Lola Etievant	*Increase efficiency and reduce bias when assessing hpv vaccination efficacy by using non-targeted hpv strains*
12:13-12:30	MO6.5	Francesca Ieva	*Quantifying the effect of mobility restrictions on health policies during the pandemic: a fda approach*

GREAT HALL

11:00 - 12:30 INVITED SESSION MIN1: **Advances on causal inference in longitudinal studies**

Organizer | Chair: **Cécile Proust-Lima** (FR)

11:00-11:18	MIN1.1	Mireille Schnitzer	*Longitudinal outcome-adaptive and marginal fused LASSO for confounder selection and model pooling with timevarying treatments*
11:19-11:37	MIN1.2	Eric Tchetgen Tchetgen	*Proximal Causal Inference for Separable Effects With Applications to Aging Research*
11:37-11:55	MIN1.3	Ruth Keogh	*Risk prediction under hypothetical interventions*

ROOM U6.6

13:30-15:00 PARALLEL SESSION MO7: **Clinical Trials 2**

Chair: **Thomas Jaki** (UK)

13:30-13:48	MO7.1	Erik van Zwet	*Estimating the treatment effect in a clinical trial*
13:49-14:06	MO7.2	Silvia Calderazzo	*Robust incorporation of external information in hypothesis testing;*
14:07-14:24	MO7.3	Niklas Hagemann	*Testing for similarity of multivariate mixed outcomes with application to efficacy-toxicity responses*
14:25-14:42	MO7.4	Thomas Gärtner	*Comparing statistical models to estimate causal treatment effects in aggregated n-of-1 trials*
14:43-15:00	MO7.5	Juliana Schneider	*Multimodal outcomes in n-of-1 trials: combining deep learning and statistical inference*

ROOM U6.7

13:30-15:00 PARALLEL SESSION MO8: **Survival Analysis 2**

Chair: **Maria Luisa Restaino** (IT)

13:30-13:48	MO8.1	Nofel Ahmed *	*Gradient boosting for survival analysis with competing risks.*
13:49-14:06	MO8.2	Edouard Bonneville	*Imputing missing covariates for competing risks analyses when using the fine–gray model*
14:07-14:24	MO8.3	Kathrin Möllenhoff	*Similarity of competing risks models with constant intensities in an application to healthcare pathways*
14:25-14:42	MO8.4	Harry Parr	*Developing & validating a competing risk joint model to characterise the prognosis of prostate cancer patients*
14:43-15:00	MO8.5	Amy Zhou	*A novel case-cohort analytical framework for semi-competing risks within the frequentist paradigm*

ROOM U6.8

13:30-15:00 PARALLEL SESSION MO9: **Latent Variable Modelling**

Chair: **Valeria Edefonti** (IT)

13:30-13:48	MO9.1	Letizia Orsini	*Random effects models of tumour growth and interval breast cancer - a study of incident cases*
13:49-14:06	MO9.2	Maren Hackenberg	*Latent dynamic modeling with differential equations for individual disease trajectories*
14:07-14:24	MO9.3	Hua Shen	*Unifying probability and nonprobability samples with misclassified covariate for improved inference*
14:25-14:42	MO9.4	Cécile Proust-Lima	*Correctly accounting for misclassification when linking latent groups with external variables*
14:43-15:00	MO9.5	Suzanne Keddie	*Estimation of the causes of fever using partial latent class analysis*

* CONFERENCE AWARDS FOR SCIENTISTS

MONDAY 28 AUGUST 2023

ROOM U6.9

13:30-15:00 PARALLEL SESSION MO10: **Causal Inference 1**

Chair: **Laura Antolini** (IT)

13:30-13:48	MO10.1	Jonathan Bartlett	*G-formula for causal inference via multiple imputation*
13:49-14:06	MO10.2	Daisy Shepherd	*Implementation of gcomputation in practice: a new diagnostic tool to guide outcome model specification*
14:07-14:24	MO10.3	Marta Spreafico	*Investigating positivity violations in marginal structural survival models: a study on estimator performance*
14:25-14:42	MO10.4	Shaun Seaman	*Estimating optimal dynamic treatment regime for survival time outcome using g-estimation*
14:43-15:00	MO10.5	Florian Lasch	*Handling symptomatic treatmentcin alzheimer's disease trials – estimators for a hypothetical strategy*

ROOM U6.10

13:30-15:00 PARALLEL SESSION MO11: **Machine Learning 1**

Chair: **Federico Ambrogi** (IT)

13:30-13:48	MO11.1	Giulia Capitoli	*Fuzzy sets in probability trees: a novel interpretable ai decision making model*
13:49-14:06	MO11.2	Juliette Murris	*Random survival forests for analysing survival data with recurrent events*
14:07-14:24	MO11.3	Oya Kalaycioglu*	*Evaluating the sample size requirements of tree-based machine learning techniques for clinical risk prediction*
14:25-14:42	MO11.4	Manel Rakez	*Impact of temporal breast density changes on the prediction of breast cancer in women from screening programs*
14:43-15:00	MO11.5	Giulia Lorenzoni	*Using chatgpt for classification of pediatric injuries from emergency department records*

ROOM U6.11

13:30-15:00 PARALLEL SESSION MO12: **Real World Data**

Chair: **Davide Bernasconi** (IT)

13:30-13:48	MO12.1	Caroline Chesang	*Emulating an existing trial of treatments for prostate cancer using real-world data: methods and challenges*
13:49-14:06	MO12.2	Joost van Rosmalen	*Augmenting treatment arms with data from expanded access using propensityscore weighted power priors*
14:07-14:24	MO12.3	Nicole Fontana	*Sequence analysis techniques to evaluate drugs-based diagnostic therapeutic paths in heart failure patients*
14:25-14:42	MO12.4	Jake Probert	*Developing an algorithm to identify breast cancer recurrences using routinely collected data in england*
14:43-15:00	MO12.5	Anqi Sui	*Outlier detection in clinical performance monitoring and comparison of commonly used methods*

GREAT HALL

13:30 - 15:00 INVITED SESSION MIN2: **Innovative designs for dose optimization studies**

Organizer | Chair: **Emily Zabor** (US)

13:00-13:18	MIN2.1	Alex Kaizer	*The Use of Master Protocol Designs with Dose-Optimization Studies*
13:19-14:36	MIN2.2	Ruitao Lin	*DEMO: Bayesian Adaptive Dose Exploration–Monitoring–Optimization Design based on Short, Intermediate, and Long-term Outcomes*
14:37-15:00	MIN2.3	Hakim-Moulay Dehbi	*Controlled amplification in oncology dose-finding trials*

15:00-15:30 COFFEE BREAK

ROOM U6.6

15:30-17:00 PARALLEL SESSION MO13: **Clinical Trials 3**

Chair: **Philip Boonstra** (US)

15:30-15:48	MO13.1	Ying Yuan	*Elastic priors to dynamically borrow information from historical data in clinical trials*
15:49-16:06	MO13.2	Emma Gerard	*Dynamic borrowing of heterogeneous historical controls: how to avoid cherry picking?*
16:07-16:24	MO13.3	Haiyan Zheng	*Incorporating historical data in the design and analysis of small population clinical trials*
16:25-16:42	MO13.4	Lou Whitehead	*A bayesian sample size calculation using functional mixture weights to incorporate historical data*
16:43-17:00	MO13.5	Elvira Erhardt	*Extrapolation in pediatrics using bayesian dynamic borrowing, tipping point analysis and expert elicitation*

ROOM U6.7

15:30-17:00 PARALLEL SESSION MO14: **Survival Analysis 3**

Chair: **Thomas Scheike** (DK)

15:30-15:48	MO14.1	Kaspar Rufibach	*Clinical trial design based on a multistate model that jointly models progression-free and overall survival*
15:49-16:06	MO14.2	Elena Tassistro	*Modelling the hazard of transition into the absorbing state in the illnessdeath model*
16:07-16:24	MO14.3	Sida Chen	*Bayesian blockwise inference for joint models of longitudinal and multistate processes*
16:25-16:42	MO14.4	Kaya Miah	*Model selection strategies for multistate modeling incorporating molecular data*

MONDAY 28 AUGUST 2023

ROOM U6.8

15:30-17:00 PARALLEL SESSION MO15: **Precision Medicine 1**

Chair: **Nan van Geloven** (NL)

15:30-15:48	MO15.1	Matteo Pedone	*A bayesian nonparametric approach to personalized treatment selection*
15:49-16:06	MO15.2	Pedro Cardoso	*Precision medicine in type 2 diabetes: bayesian non-parametric modelling of glucose-lowering therapy efficacy*
16:07-16:24	MO15.3	Caterina Gregorio	*Estimating optimal rules for personalized treatment decisions through functional survival analysis*
16:25-16:42	MO15.4	Dimitris Rizopoulos	*Causal effects of salvage therapy using joint models for longitudinal and time-to-event dat*

ROOM U6.9

15:30-17:00 PARALLEL SESSION MO16: **Longitudinal Analysis 1**

Chair: **Giota Touloumi** (GR)

15:30-15:48	MO16.1	Matteo Amestoy	*Combined shrinkage of fixed and random effects in linear mixed models using empirical bayes*
15:49-16:06	MO16.2	Daniel Temko	*A bayesian functional principal component analysis framework for longitudinal genome-wide association studies*
16:07-16:24	MO16.3	Nemo Fournier	*A bayesian model to study the genetic risks driving alzheimer's disease progression patterns*
16:25-16:42	MO16.4	Marco Palma	*Distributional models for the quantification of within-individual lung function variability in cystic fibrosis*
16:43-17:00	MO16.5	Tui Nolan	*A hierarchical modelling approach for principal components analysis on multiple longitudinal variables*

ROOM U6.10

15:30-17:00 PARALLEL SESSION MO17: **High Dimensional Data 2**

Chair: **Lara Lusa** (SI)

15:30-15:48	MO17.1	Erika Banzato	*A bartlett-type correction for likelihood ratio tests for testing equality of gaussian graphical models*
15:49-16:06	MO17.2	Fernando Castro-Prado	*Pearson's chi-squared meets distance and kernel tests: an application to complex disease genetics*
16:07-16:24	MO17.3	Anna Vesely	*Exploring between-subject consistency in fmri signals through partial conjunction null hypotheses*
16:25-16:42	MO17.4	Angela Andreella	*Towards a power analysis and sample size estimation for plsbased methods*
16:43-17:00	MO17.5	Livio Finos	*Selective inference in factorial designs with high-dimensional response*

MONDAY 28 AUGUST 2023

ROOM U6.11

15:30-17:00 PARALLEL SESSION MO18: **Meta-Analysis**

Chair: **Dario Gregori** (IT)

15:30-15:48	MO18.1	Stephen Walter	*Bias corrections for study weights in meta-analyses with binary outcomes*
15:49-16:06	MO18.2	Pablo Verde	*Bayesian nonparametric approaches and the bias-corrected meta-analysis model for combining disparate studies*
16:07-16:24	MO18.3	Anne lyngholm Soerensen	*Prospective and retrospective sequential meta-analysis using trial sequential analysis*
16:25-16:42	MO18.4	Alessandra Meddis	*Pseudo-values approach for quantile analysis in individual patient data meta-analysis*

GREAT HALL

15:30 - 17:00 INVITED SESSION MIN3: **Vaccination programmes: post - implementation assessment of protection, benefits and risks**

Organizer | Chair: **Marie Reilly** (SE)

15:30-15:48	MIN3.1	Heather Whitaker	*Statistical methods for the epidemiological evaluation of vaccine safety;*
15:49-16:06	MIN3.2	Susan Hahne	*Monitoring and evaluating Covid-19 vaccination programmes: real world challeges;*
16:07-16:25	MIN3.3	Andrea Callegaro	*Statistical methods to assess immunological surrogate endpoints for vaccines;*

TUESDAY 29 AUGUST 2023

ROOM U6.6

9:00-10:30 PARALLEL SESSION TO1: **Clinical Trials 4**

Chair: **Stefania Galimberti** (IT)

09:00-09:18	TO1.1	Pavel Mozgunov	*How (not) to conduct a simulation study for a trial design: a case of dosefinding clinical trials*
09:19-09:36	TO1.2	Xijin Chen	*Using ctdna as a novel biomarker of efficacy for dose-finding trials in oncology*
09:37-09:54	TO1.3	Dario Zocholl	*Estimating the similarity between adult and pediatric dose-toxicity curves to inform pediatric dosefinding*
09:55-10:12	TO1.4	Anais Andrillon	*Incorporating patientreported outcomes in dose-finding clinical trials with continuous patientenrollment*
10:13-10:30	TO1.5	Emily Alger	*Designing patient-centred dose-finding trials with patient-reported outcomes: opportunities and challenges*

ROOM U6.7

9:00-10:30 PARALLEL SESSION TO2: **Survival Analysis 4**

Chair: **Laura Antolini** (IT)

09:00-09:18	TO2.1	Riccardo De Santis	*Sign-Flip Test For Coefficients In The Cox Regression Model*
09:19-09:36	TO2.2	Annabel Webb	*Penalized Likelihood Estimation Of Cox Models With Doubly Truncated And Interval Censored Survival Times*
09:37-09:54	TO2.3	Guilherme Wang de Faria Barros	*Impact Of Non-Informative Censoring On Propensity Score Based Estimates Of Marginal Hazard Ratios*
09:55-10:12	TO2.4	Alexandra Strobel	*Impact Of Omitted Covariates On Treatment Estimates In Propensity Score Matched Studies*
10:13-10:30	TO2.5	Ákos Ferenc Pap	*Comparing Overall Benefit/Risk Of Treatments By Weighted Cox Model On Ordering Scores For Relevant Events*

ROOM U6.8

9:00-10:30 PARALLEL SESSION TO3: **Students Award**

Chair: **Jonathan Bartlett** (UK)

09:00-09:18	TO3.1	Jiaxin Zhang	*Sensitivity analysis for missingness assumptions in causal inference: accommodating the substantive analysis*
09:19-09:36	TO3.2	Léonie Courcoul	*A location-scale joint model with a timedependent subjectspecific variance of the marker and competing event*
09:37-09:54	TO3.3	Zhenwei Yang	*A bayesian joint modelling for misclassified intervalcensoring and competing risks*
09:55-10:12	TO3.4	Pedro Miranda Afonso	*A joint model for (un)bounded longitudinal markers, competing risks, and recurrent events using registry data*
10:13-10:30	TO3.5	Joshua Philipp Entrop	*Parametric estimation of the mean number of events in the presence of competing risks*

ROOM U6.9

9:00-10:30 PARALLEL SESSION TO4: **Causal Inference 2**

Chair: **David Glidden** (US)

09:00-09:18	TO4.1	Jingyi Xuan	*Is Inverse Probability Of Censoring Weighting A Safe Alternative To Per-Protocol Analysis?*
09:19-09:36	TO4.2	Ilaria Prosepe	*Combining Sequential Stratification And Iptw Weights To Estimate The Survival Benefit Of Liver Transplantation*
09:37-09:54	TO4.3	Rik van Eekelen	*Reducing Time-Lag Bias When Comparing Treated Patients To Controls With A Different Start Of Follow-Up*
09:55-10:12	TO4.4	Hans de Ferrante	*Basing Discrete Event Simulators For Organ Allocation On Counterfactual Mortality Risks*
10:13-10:30	TO4.5	Kateline Le Bourdonnec	*Continuous-Time Mediation Analysis For Repeated Mediators And Outcomes*

ROOM U6.10

9:00-10:30 PARALLEL SESSION TO5: **Machine Learning 2**

Chair: **Stefano Calza** (IT)

09:00-09:18	TO5.1	Davide Bernasconi	*Ensemble Algorithm Based On Shapley Values Beyond Binary Classification: Simulations And Clinical Application*
09:19-09:36	TO5.2	Yao Chen	*Comparison Of Classification Methods For Multiplex Digital Pcr Data*
09:37-09:54	TO5.3	Camilo Broc	*Comparative Analysis Of Supervised Integrative Methods For Multi-Omics Data*
09:55-10:12	TO5.4	Marie Breeur	*Optimal Transport For Automatic Alignment Of Non-Targeted Metabolomic Data*
10:13-10:30	TO5.5	Corrado Lanera	*Artificial Intelligence For The Prediction Of Weaning Readiness Outcome In Mechanically Ventilated Patients*

GREAT HALL

9:00 - 10:30 INVITED SESSION TIN1: **Evaluation of predictive algorithms and models: uncertainty and impact on medical care**

Organizer | Chair: **Ewout Steyerberg** (NL)

09:00-09:18	TIN1.1	Ben van Calster	*Sources of uncertainty in clinical prediction models;*
09:19-09:36	TIN1.2	Paula Dhiman	*Reporting and Methodological quality of machine learning prediction model studies: an overview of results;*
09:37-09:55	TIN1.3	Laure Wynants	*Measuring clinical utility: uncertainty in Net Benefit;*

TUESDAY 29 AUGUST 2023

10:30-11:00 COFFEE BREAK

ROOM U6.6

11:00-12:30 PARALLEL SESSION TO6: **Clinical Trials 5**

Chair: **Kaspar Rufibach** (CH)

11:00-11:19	TO6.1	James Wason	*Sample Size Estimation For Clinical Trials Using Complex Responder Endpoints*
11:19-11:36	TO6.2	Carolin Herrmann	*Sample Size Adaptations In Clinical Trials Comparing Restricted Mean Survival Times – Advantages And Drawbacks*
11:37-11:54	TO6.3	S. Faye Williamson	*Hybrid Sample Size Calculations For Cluster Randomised Trials Usingì Assurance*
11:55-12:12	TO6.4	Samuel Sarkodie	*A Hybrid Approach To Sample Size Reestimation In Cluster Randomized Trials With Continuous Outcomes*
12:13-12:30	TO6.5	Judith ter Schure	*The Anytime-Valid Logrank Test For Flexible Collaborative Meta-Analysis And Platform Trials*

ROOM U6.7

11:00-12:30 PARALLEL SESSION TO7: **Survival Analysis 5**

Chair: **Thomas Scheike** (DK)

11:00-11:19	TO7.1	Birzhan Akynkozhayev	*Flexible Parametric Accelerated Failure Time Models With Cure*
11:19-11:36	TO7.2	Serigne Lo	*Penalized Likelihood Approach For Mixture Cure Model With Interval Censoring - An Application To Thin Melanoma*
11:37-11:54	TO7.3	Benoit Liquet	*Mixture Cure Semi-Parametric Accelerated Failure Time Models With Partly Intervalcensored Data*
11:55-12:12	TO7.4	Thomas Klausch	*A Flexible Bayesian Prevalenceincidence Mixture Model For Screening Data*

ROOM U6.8

11:00-12:30 PARALLEL SESSION TO8: **Prediction Models 2**

Chair: **Ewout Steyerberg** (NL)

11:00-11:19	TO8.1	Haoyue Wang	*Comparing Uncertainty In Individual Probability Predictions With Various Models And Model Average*
11:19-11:36	TO8.2	Zhujie Gu	*Evaluating The Uncertainty Of The Risk Predicted From The Two-Stage Landmarking Model*
11:37-11:54	TO8.3	Ewan Carr	*A Simulation Approach To Calculating Minimum Sample Sizes For Prediction Modelling: The Pmsims Package For R*
11:55-12:12	TO8.4	Johanna Munoz	*Synthesis Calibration Curves*
12:13-12:30	TO8.5	Bart Mertens	*Accounting For Missing Values In The Calibration And Application Of Prediction Models*

ROOM U6.9

11:00-12:30 PARALLEL SESSION TO9: **Longitudinal Analysis 2**

Chair: **Cécile Proust-Lima** (FR)

11:00-11:19	TO9.1	Stefano Renzetti	*A Weighted Quantile Sum Regression With Penalized Weights And Two Indices*
11:19-11:36	TO9.2	Balint Tamasi	*Flexible Parametric Regression For Correlated Data With Transformation Models*
11:37-11:54	TO9.3	Zsolt Lang	*Comparison Of Conditional And Marginal Means In Distribution Based Marginalized Multilevel Models*
11:55-12:12	TO9.4	Melanie Prague	*Use Of Priors In Automated Model Building Strategies For Non Linear Mixed Effects Models*
12:13-12:30	TO9.5	Mariella Gregorich	*Variable Selection With 'Too' Many Zero-Inflated Predictors: A Nonnegative Garrote Approach*

ROOM U6.10

11:00-12:30 PARALLEL SESSION TO10: **Clinical Trials 6**

Chair: **Thomas Jaki** (UK)

11:00-11:19	TO10.1	Jennifer Thompson	*Marginal Odds Ratios For Cluster Randomised Trials: A Novel Analysis Method*
11:19-11:36	TO10.2	Floriane Le Vilain--Abraham	*Pseudo-Values Regression For Restricted Mean Survival Time In Small Sample Cluster Randomized Trials*
11:37-11:54	TO10.3	Kelsey Grantham	*Optimal Staircase Designs And When To Use Them*
11:55-12:12	TO10.4	Ehsan Rezaei	*Finding Cost-Efficient Incomplete Stepped Wedge Designs Using An Iterative Approach*
12:13-12:30	TO10.5	Antoine Pitoy	*Joint Modelling For Phase Iii Clinical Trial Primary Endpoint Estimation: Simulation Study And Application*

TUESDAY 29 AUGUST 2023

ROOM U6.11

11:00-12:30 PARALLEL SESSION TO11: **ItR-IBS & Italian Statistical Society**

Chair: **Fulvia Mecatti** (IT)

11:00-11:19	TO11.1	Chiara Brombin	*Exploring The Relationship With The Self-Image In The Digital Era: Integrating Model-Based Clustering And Graphical Model Approaches*
11:19-11:36	TO11.2	Marco Bonetti	*The Average Uneven Mortality Index: Building On The "Edagger" Measure Of Lifespan Inequality*
11:37-11:54	TO11.3	Aldo Solari	*Simultaneous Directional Inference*
11:55-12:12	TO11.4	Clelia Di Serio	*Treatment Effectassessment In Observational Studies: A Propensity Score Method Based On Bayesian Networks*

GREAT HALL

11:00 - 12:30 INVITED SESSION TIN2: **High-dimensional inference in biostatistics**

Organizer | Chair: **Francesco C. Stingo** (IT)

11:00-11:18	TIN2.1	Veera Baladandayuthapani	*Spacex: gene co-expression network estimation for spatial transcriptomics;*
11:19-11:36	TIN2.2	Manuela Zucknick	*Bayesian hierarchical models for large-scale pharmacogenomic screens of drug combinations;*
11:37-11:55	TIN2.3	Paul Kirk	*Outcome-guided multi-view bayesian clustering for integrative omic data analysis*

12:30-13:30 LUNCH BREAK

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

WEDNESDAY 30 AUGUST 2023

ROOM U6.6

9:00-10:30 PARALLEL SESSION WO1: **Clinical Trials**

Chair: **Philip Boonstra** (US)

09:00-09:19	WO1.1	Shirin Golchi	*Methods For Assessment Of Frequentist Operating Characteristics In Bayesian Trials*
09:19-09:36	WO1.2	Nico Bruder	*Optimal Adaptive Designs For Time-To-Event Data: A Simulation Study*
09:37-09:54	WO1.3	Nigel Stallard	*Confirmatory Adaptive Enrichment Designs With A Normally Distributed Outcome*
09:55-10:12	WO1.4	Abigail Burdon	*Adaptive Enrichment Clinical Trial Designs Using Joint Modelling Of Longitudinal And Time-To-Event Data*
10:13-10:30	WO1.5	Luke Ouma	*A Two-Stage Bayesian Adaptive Umbrella Design Borrowing Information Over The Control Data*

ROOM U6.7

9:00-10:30 PARALLEL SESSION WO2: **Survival Analysis 6**

Chair: **Marco Bonetti** (IT)

09:00-09:19	WO2.1	Maximilian Bardo	*The Shape Of The Relative Frailty Variance Induced By Discrete Random Effects In Time-To-Event Models*
09:19-09:36	WO2.2	Maria Veronica Vinattieri	*Family History In Breast Cancer Development*
09:37-09:54	WO2.3	Jordache Ramjith	*Flexible Time-To-Event Models For Double-Interval-Censored Data With A Competing Event*
09:55-10:12	WO2.4	Jacobo de Uña-Álvarez	*Model Assessment In Regression With A Doubly Truncated Response*
10:13-10:30	WO2.5	Caroline Weibull	*Modelling Excess Mortality Comparing To A Control Population: A Combined Additive And Relative Hazards Model*

ROOM U6.8

9:00-10:30 PARALLEL SESSION WO3: **Longitudinal Analysis 3**

Chair: **Lara Lusa** (SI)

09:00-09:19	WO3.1	Paola Rancoita	*How Resampling Methods Can Improve Variable Selection In Longitudinal Models*
09:19-09:36	WO3.2	Taban Baghfalaki	*Dynamic Prediction Of An Event Using Multiple Longitudinal Markers: A Model Averaging Approach*
09:37-09:54	WO3.3	Anaïs Rouanet	*Non-Parametric Clustering Of Multivariate Longitudinal Data: Identifying Sub-Phenotypes Of Alzheimer's Disease*
09:55-10:12	WO3.4	Christos Thomadakis	*Shared-Parameter Modelling Of Longitudinal Data Allowing For Possibly Informative Visiting Process And Dropout*
10:13-10:30	WO3.5	Graham Wheeler	*Impact Of Partial Information In Longitudinal Group-Sequential Designs On Probability Of Success Calculations*

ROOM U6.9

9:00-10:30 PARALLEL SESSION WO4: **Causal Inference 3**

Chair: **Valeria Edefonti** (IT)

09:00-09:19	WO4.1	Claus Ekstrøm	*Data-Driven Model Building For Life-Course Epidemiology*
09:19-09:36	WO4.2	Solène Cadiou	*Outcome- Versus Exposure-Wide Framework In Molecular Epidemiology: False Positive Findings Due To Correlation*
09:37-09:54	WO4.3	Jasmin Rühl	*Resampling-Based Confidence Intervals And Bands For The Average Treatment Effect In Time-To-Event Data*
09:55-10:12	WO4.4	Josef Fritz	*Simulating Collider Stratification Bias And An Application To The Inverse Obesity Paradox In Prostate Cancer*
10:13-10:30	WO4.5	Jun Yin	*Evaluate Application Of Causal Machine Learning To Adaptive Enrichment Clinical Trials*

ROOM U6.10

9:00-10:30 PARALLEL SESSION WO5: **Missing Data**

Chair: **Jonathan Bartlett** (UK)

09:00-09:19	WO5.1	James Carpenter	*Substantive Model Compatible Multilevel Multiple Imputation: A Joint Modeling Approach*
09:19-09:36	WO5.2	Haoxiang Gao	*Handling Missing Data In Binary Variables With Low Prevalence*
09:37-09:54	WO5.3	Daniel Bratton	*Advanced Bayesian Joint Modelling For Time-To-Event Subgroup Analysis With Partially Missing Subgroup Status*
09:55-10:12	WO5.4	Doranne Thomassen	*Imputation Of Longitudinal Patientì Reported Outcomes In The Presence Of Death And Other Intercurrent Events*
10:13-10:30	WO5.5	Elinor Curnow	*The Midoc R Package: Providing Expert Guidance And Methodology For Multiple Imputation*

ROOM U6.11

9:00-10:30 PARALLEL SESSION WO6: **Epidemiology 2**

Chair: **Giota Touloumi** (GR)

09:00-09:19	WO6.1	Marta Rossi	*A Case-Control Study To Evaluate Blood Bacterial Dna In The Intestinal Adenoma-Carcinoma Sequence*
09:19-09:36	WO6.2	Elena Colicino	*Integrating Data Across Multiple Sites To Examine Associations Between A Metal Mixture And Child Cognition*
09:37-09:54	WO6.3	Loukia Spineli	*Hierarchical Clustering For The Evaluation Of Transitivity Assumption In A Network Of Interventions*
09:55-10:12	WO6.4	Augustine Wigle	*Bayesian Unanchored Additive Models For Component Network Meta-Analysis*
10:13-10:30	WO6.5	Audrey Beliveau	*Generalized Fused Lasso For Treatment Pooling In Network Meta-Analysis*

GREAT HALL

11:00 - 12:30 INVITED SESSION WIN1:

Recurrent events and their use in medical studies

Organizer | Chair: **Thomas Scheike** (DK)

09:19-09:36	WIN1	Mouna Akacha	*Estimans for recurrent event endpoints*
09:37-09:54	WIN2	Giuliana Cortese	*Estimating the marginal and conditional means of recurrent events in presence of terminal events*
09:55-10:13	WIN3	Per Kragh Andersen	*Dealing with competing risks in the analysis of recurrent event*

10:30-11:00 COFFEE BREAK

GREAT HALL

11:00 – 11:15 INTRODUCTION

Introduction by: **Tomasz Burzykowski** (BE) ISCB President

GREAT HALL

11:00 - 12:30 PLE2: **PRESIDENT'S INVITED SPEAKER LECTURE**

On causal inference, estimands and trials in epidemiology and biostatistics

President's Invited Speaker: **Vanessa Didelez** *Professor of Statistics with Focus on Theory and Methods for Causal Inference Leibniz Institute for Prevention Research and Epidemiology - BI PS Bremen, D*

12:30 - 13:30 LUNCH BREAK and Annual General Meeting (AGM) ISCB at ROOM U6.6

ROOM U6.6

13:30-15:00 PARALLEL SESSION WO7: **Clinical Trials 8**

Chair: **Stefania Galimberti** (IT)

13:30-13:48	WO7.1	Michael LeBlanc	*Utilizing Co-Primary Endpoints To Test For Clinically Significant Differences In Progressionfree Survival*
13:49-14:06	WO7.2	Mollie Payne	*Analysis Of Multicentre Trials: Limiting The Effect Of Centre Heterogeneity On The Marginal Treatment Effect*
14:07-14:24	WO7.3	Alessandra Serra	*Determining The Minimum Duration Of Treatment In Tuberculosis: An Orderrestricted Non-Inferiority Design*
14:25-14:42	WO7.4	Susan Halabi	*On The Design Of Biomarkerdriven Trials With Measurement Error For Time To Event Outcomes*
14:43-15:00	WO7.5	Azzolina Danila	*A Superlearner-Enforced Approach For The Estimation Of Treatment Effect In Pediatric Trials*

ROOM U6.7

13:30-15:00 PARALLEL SESSION WO8: **Survival Analysis 7**

Chair: **Maria Grazia Valsecchi** (IT)

13:20 - DEDICATED TO PROF. ETTORE MARUBINI			
13:30-13:48	WO8.1	Terry Therneau	*Competing Risks, The Finegray Model, And Pseudovalues*
13:49-14:06	WO8.2	Matteo Di Maso	*Analyzing Restricted Mean Survival Time Curves Using Pseudo-Values And Machine Learning*
14:07-14:24	WO8.3	Morten Overgaard	*A Comparison Of Kaplan-- Meier-Based Inverse Probability Of Censoring Weighted Regression Methods*
14:25-14:42	WO8.4	Léa Orsini	*Quadratic Inference Functions As A New Approach To Analyze Pseudo Observations In Survival Analysis*
14:43-15:00	WO8.5	Giacomo Biganzoli	*Clinical Impact And Disease Dynamics In Competing Risks: An Analysis Of Two Historical Clinical Trials*

ROOM U6.8

13:30-15:00 PARALLEL SESSION WO9: **Prediction Models 3**

Chair: **Laure Wynants** (NL)

13:30-13:48	WO9.1	Nan van Geloven	*Causal Blind Spots In Risk-Based Decision Making*
13:49-14:06	WO9.2	Kim Luijken	*Risk-Based Decision Making:Formulating Estimands For Prediction Under Hypothetical Interventions*
14:07-14:24	WO9.3	Max Behrens	*Improving Local Prediction Models Using Similarity Based Data Pooling*
14:25-14:42	WO9.4	Maryam Farhadizadeh	*Similarity Quantification For Small Data*
14:43-15:00	WO9.5	Celina Gehringer	*Predicting Response Under Interventions In Patients With Rheumatoid Arthritis: A Methodological Exploration*

ROOM U6.9

13:30-15:00 PARALLEL SESSION WO10: **Causal Inference 4**

Chair: **David Glidden** (US)

13:30-13:48	WO10.1	Mats Stensrud	*From Data To Decisions: How Effects Of Intervening Variables Can Guide Policies*
13:49-14:06	WO10.2	Rushani Wijesuriya	*Methodology For Systematic Identification And Analysis Of Multiple Biases In Causal Inference*
14:07-14:24	WO10.3	Amirhossein Kazemi	*Instrumental Variable Analysis With Categorical Treatment And Ordinal Instrumen*
14:25-14:42	WO10.4	Laura Guedemann	*Just Wha; T The Doctor Ordered: An Evaluation Of Provider Preference-Based Instrumental Variable Methods*
14:43-15:00	WO10.5	Daniel Nevo	*Practical Considerations Of Using Negative Control Exposures To Detect Residual Confounding*

ROOM U6.10

13:30-15:00 PARALLEL SESSION WO11: **Machine Learning 3**

Chair: **Davide Gregori** (IT)

13:30-13:48	WO11.1	Andrea Sottosanti	*(Co-)Clustering Models For Spatial Transcriptomics*
13:49-14:06	WO11.2	Emilie Eliseussen	*Bayesian Rank-Based Clustering Via Mallows Mixtures With Covariates For Cancer Subtyping*
14:07-14:24	WO11.3	Marc Delord	*A Clustering Approach To Multiple Time-To-Event Data And Application To Multimorbidity Associated With Stroke*
14:25-14:42	WO11.4	Ilaria Gandin	*Identification Of Novel Dilated Cardiomyopathy Sub-Phenotypes: Unsupervised Clustering For Mixed-Data Type*

ROOM U6.11

13:30-15:00 PARALLEL SESSION WO12: **Syntetic Data 1**

Chair: **Clelia Di Serio** (IT)

13:30-13:48	WO12.1	Tim Friede	*From Clinical Trial Simulations To In-Silico Trials*
13:49-14:06	WO12.2	Luca Carmisciano	*The Challenges, Feasibility And Limits Of Statistical Analysis On Purely Synthetic Biomedical Data*
14:07-14:24	WO12.3	Marian Mitroiu	*Data-Generating Models Of Longitudinal Continuous Outcomes And Intercurrent Events To Evaluate Estimands*
14:25-14:42	WO12.4	Kiana Farhadyar	*A Simple Yet Effective Approach For Synthetic Clinical Data Generation With Realistic Marginal Distributions*
14:43-15:00	WO12.5	Andreas Ziegler	*A Simple-To-Use R Package For Mimicking Study Data By Simulations*

GREAT HALL

13:30 - 15:00 INVITED SESSION WIN2: **Quantification of safety signals in clinical trials: Estimand, estimation, and how would good look like in ten years?**

Organizer | Chair: **Kaspar Rufibach** (CH)

13:30-13:48	WIN2.1	Kaspar Rufibach	*Principled approach to time-to-event endpoints with competing risks, with a focus on analysis of aes;*
13:49-14:06	WIN2.2	Anja Loos	*Estimands for safety – one size fits all?;*
14:07-14:24	WIN2.3	Laura Antolini	*Adverse events with survival outcomes: from clinical questions to methods for statistical analysis;*
14:25-14:43	WIN2.4	Kit Roes	*Regulatory perspective on the analysis of safety in clinical trials and beyond;*

15:00-15:30 COFFEE BREAK

ROOM U6.6

15:30–17:00 PARALLEL SESSION WO13: **Clinical Trials 9**

Chair: **Emily Zabor** (US)

15:30-15:48	WO13.1	Lukas Baumann	*Basket Trial Designs Based On Power Priors That Incorporate Overall Heterogeneity*
15:49-16:06	WO13.2	Lukas D. Sauer	*Application Of Constrained Optimization Techniques To Bayesian Basket Trial Designs*
16:07-16:24	WO13.3	Satoshi Hattori	*Frequentist Analysis Of Basket Trials With One-Sample Mantel-Haenszel Procedures*
16:25-16:42	WO13.4	Ian Marschner	*Non-Concurrent Controls In Platform Trials: Separating Randomised And Nonrandomised Information*

ROOM U6.7

15:30–17:00 PARALLEL SESSION WO14: **Miscellanea**

Chair: **Stefania Galimberti** (IT)

15:30-15:48	WO14.1	Willi Sauerbrei	*Improve Clinical And Methodological Research By Adherence To Reporting Guidelines And Structured Reporting*
15:49-16:06	WO14.2	John Ferguson	*Confidence Intervals Using Approximate Propagation Of Imprecision*
16:07-16:24	WO14.3	DEBAJYOTI SINHA	*Nonparametric Bayesian Analysis Of Survival Data With Spatially Correlated Cluster Effects Using Soft-Bart*
16:25-16:42	WO14.4	Katherine Thomas	*Survey Sampling Methods For Partial Verification Bias In Diagnostic Evaluation Studies*
16:43-17:00	WO14.5	Francesco Innocenti	*Optimal Two-Stage Sampling For Mean Estimation In Multilevel Populations When Cluster Size Is Informative*

ROOM U6.8

15:30–17:00 PARALLEL SESSION WO15: **Precision Medicine 2**

Chair: **Francesco C. Stingo** (IT)

15:30-15:48	WO15.1	J. Jack Lee	*Optimizing Information Borrowing For Bayesian Hierarchical Model In Subgroup Analysis*
15:49-16:06	WO15.2	Laura Bondi	*Prioritising The Outcome In Bayesian Profile Regression: An Application To Osteoarthritis Proteomic Data*
16:07-16:24	WO15.3	Valentin Vinnat	*Bayesian Sequential Design For Identifiying And Ranking Of Subgroups Based On Biomarkers In Sepsis's Patients*
16:25-16:42	WO15.4	Christos Nakas	*Statistical Inference For Roc Curves After The Box-Cox Transformation And Use Of The R Package `Rocbc'*
16:43-17:00	WO15.5	Marcel Wolbers	*Treatment Effect Estimation For Time-Toevent Outcomes In Overlapping Subgroups Based On Shrinkage Methods*

WEDNESDAY 30 AUGUST 2023

ROOM U6.9

15:30–17:00 PARALLEL SESSION WO16: **Longitudinal Analysis 4**

Chair: **Cécile Proust-Lima** (FR)

15:30-15:48	WO16.1	Dimitris Rizopoulos	*Extended Joint Models Under The Bayesian Approach Using Jmbayes2*
15:49-16:06	WO16.2	Denis Rustand	*Bayesian Inference For Joint Models Of Longitudinal And Survival Data With Dynamic Risk Prediction*
16:07-16:24	WO16.3	Tiphaine Saulnier	*Joint Analysis Of Disease Progression Markers And Death Using Individual Temporal Recalibration*
16:25-16:42	WO16.4	Hadrien Charvat	*A Lambert Function-Based Procedure To Fit Joint Models For Multivariate Longitudinal And Time-To-Event Data*
16:43-17:00	WO16.5	Mohadeseh Shojaei Shahrokhabadi	*A Novel Platform For Analyzing Semi-Continuous Medical Cost And Survival Data*

ROOM U6.10

15:30–17:00 PARALLEL SESSION WO17: **High Dimensional Data 3**

Chair: **Stefano Calza** (IT)

15:30-15:48	WO17.1	Giulia Capitoli	*Co-Clustering Matrix Trifactorization: Spatial And Features Constraints*
15:49-16:06	WO17.2	Daniela Corbetta	*Procrustes Analysis For Spatial Transcriptomics Data*
16:07-16:24	WO17.3	Said El Bouhaddani	*Statistical Integration Of Multiomics And Drug Screening Data From Cell Lines*
16:25-16:42	WO17.4	Connie Musisi	*Analysis Of Compositional Microbiome Data With Bias Correction Using Poisson Framework*
16:43-17:00	WO17.5	Patric Tippmann	*Quantifying Uncertainty In Deep Generative Synthesis Of Tabular Medical Data With Bayesian Inference*

GREAT HALL

15:30 - 17:00 INVITED SESSION WIN3: **Marginal versus conditional effects in clinical trials**

Organizer | Chair: **Jonathan Bartlett** (UK)

15:49-16:06	WIN3.1	David Benkeser	*A value system for evaluating estimands in randomized trials*
16:07-16:24	WIN3.2	Michael Rosenblum	*Conditional vs. marginal effects in randomized trials: tradeoffs*
16:25-16:42	WIN3.3	Stephen Senn	*Why do we worry about marginal inference?*
16:43-17:00	WIN3.4	Ewout Steyerberg	*Covariate adjustment and exploiting ordinality: simulations of power and a rieview of neurological trials*

ROOM U6.6

9:00-16:30 MINI SYMPOSIUM 1: Ten years STRATOS initiative – brief summary of progress and plans for the future

Coordinators: **Willi Sauerbrei** (DE)
Federico Ambrogi (IT)

SESSION 1 - 9.00-10.40

TMS1

9.00-9.50	TMS1.1	Willi Sauerbrei (DE)	*Experience and progress with developing guidance for the analysis of key topics in observational research*
9:50-10.15	TMS1.2	Katherine Lee (UK) for TG1	*Level 1 guidance on conducting and reporting sensitivity analyses for missing data*

TOPIC GROUP: "MISSING DATA" (TG1)

10.15-10.40	TMS1.3	Kim Luijken (NL)	*Aims of the new Open Science panel*

PANEL: "OPEN SCIENCE"

10:40 COFFEE BREAK

THURSDAY 31 AUGUST 2023

SESSION 2 -11.00-12.40

11.00-11.25	TMS1.4	Georg Heinze (AT)	*Ongoing research towards state-of-the-art in variable and functional form selection for statistical models*

TOPIC GROUP: "SELECTION OF VARIABLES AND FUNCTIONAL FORMS IN MULTIVARIABLE ANALYSIS " (TG2)

11.25-11.50	TMS1.5	Cécile Proust-Lima (FR)	*How to include time-varying exposures prone to measurement error in survival analyses*

TOPIC GROUP: "MEASUREMENT ERROR AND MISCLASSIFICATION" (TG4)

11.50-12.15	TMS1.6	Michal Abrahamowicz (CA)	*Evaluating the impact of covariate measurement error on functional form estimation in regression modelling*

A COLLABORATIVE PROJECT OF TOPIC GROUPS: "SELECTION OF VARIABLES AND FUNCTIONAL FORMS IN MULTIVARIABLE ANALYSIS" (TG2) AND "MEASUREMENT ERROR AND MISCLASSIFICATION" (TG4)

12.15-12-40	TMS1.7	Federico Ambrogi (IT)	*Statistical analysis of high-dimensional biomedical data: A gentle introduction to analytical goals, common approaches and challenges*

TOPIC GROUP: "HIGH-DIMENSIONAL DATA (TG9)"

13:00 LUNCH BREAK

THURSDAY 31 AUGUST 2023

SESSION 3 - 13.30-15.10			
13.30-13.55	**TMS1.8**	Lara Lusa (IT)	*The slowly changing landscape of predictive modeling in biomedicine*
13.55-14.20	**TMS1.9**	Nan Van Geloven N (NL)	*Counterfactual prediction for personalized healthcare using observational data*
TOPIC GROUP: A COLLABORATIVE PROJECT OF TOPIC GROUPS "EVALUATING DIAGNOSTIC TESTS AND PREDICTION MODELS" (TG6) AND "CAUSAL INFERENCE" (TG7)			
14.20-14.45	**TMS1.10**	Saskia Le Cessie (NL)	*Recommendations to handle patient reported outcome data in oncology cancer trials*
WORK PACKAGE 3 OF THE SISAQOL-IMI CONSORTIUM			
14.45-15.10	**TMS1.11**	Els Goetghebeur (BE)	*Comparing quality of life – while alive – between treatment and (external) controls: methods for real world analysis*
WORK PACKAGE 3 OF THE SISAQOL-IMI CONSORTIUM			
15:10-15:30		**COFFEE BREAK**	
SESSION 4 - 15.30-16.30			
		Chair: **James Carpenter** (UK)	*PANEL DISCUSSION ABOUT THE FUTURE OF STRATOS*

ROOM U6.8

9:00-12:15 MINI SYMPOSIUM 2: Novel approaches to complex data and predictive modeling in healthcare research

Coordinators: **Emanuele Di Angelantonio** (IT)
Francesca Ieva (IT)

09:00-09:15		Chairs: **Emanuele Di Angelantonio** (IT) **Maria Grazia Valsecchi** (IT)	*Introduction*
TMS2			
09:15-9:45	**TMS2.1**	**Catalina Vallejos** (UK)	*Predicting Emergency Admissions In Scotland*
09:45-10:15	**TMS2.2**	**Mihaela van der Schaar** (UK)	*Time: The Next Frontier In Machine Learning For Healthcare*
10:15-10:45		**COFFEE BREAK**	
10:45-11:15	**TMS2.3**	**Marteen van Smeden** (NL)	*Rage Against The Machine Learning*
11:15-11:45	**TMS2.4**	**Davide Bernasconi** (IT)	*Regression And Ml Approaches For Evaluation Of Biomarkers With Application To Primary Biliary Cholangitis*
11:45 – 12:15		Chair: **Francesca Ieva** (IT)	*PANEL DISCUSSION / Q&A*

THURSDAY 31 AUGUST 2023

ROOM U6.9

9:00-12:45 — **Early Career Biostatistician Day (ECB)**

Coordinators: **Early Career Biostatistician Committee**

PART A

09:00-09:45	ECB.1	**Valeria Edefonti** (IT) (invited speaker)	*Sustaining A Culture Of Reproducibility In Research: A Personal Credo For Early Career Biostatisticians*
09:45-10:00	ECB.2	**Rushani Wijesuriya**	*A Network Of Mentors: Leveraging The Power Of Networking And Mentoring To Accelerate Your Career*
10:00-10:15	ECB.3	**Judith ter Schure**	*Unreported Rct Results: Should Biostatisticians Care?*
10:15-10:30	ECB.4	**Bethany Hillier**	*Are Statistical And Scientific Assessments Of Rapid Self-Test Diagnostics Reliable?*

10:30-10:45 — COFFEE BREAK

PART B

11:00-11:15	ECB.5	**Elena Albu**	*Challenges In Extracting And Processing Ehr Data For Dynamic Prediction Models*
11:15-11:30	ECB.6	**Alexandra Hunt**	*Early Career Biostatisticians*
11:30-11:45	ECB.7	**Autumn O'Donnell**	*Tackling Imposter Syndrom*
11:45-12:30	ECB.8	**Katherine Lee** (Invited Speaker)	*Navigating The World Of Biostatistics*
12:30-12:45		Chair: **Camila Olarte Parra** (UK)	*PANEL DISCUSSION / Q&A*

CONFERENCE DAYS

Invited and Parallel Sessions

Keynote Speaker

PLE1 Statistical adventures in pursuit of precision medicine: secret signatures, sliding subgroups & more

Mcshane L.M.*
Biometric Research Program, National Cancer Institute, National Institutes of Health, USA ~ Rockville (Maryland) ~ United States of America

A goal of precision medicine is to deliver the right drug to the right patient at the right time and dose to achieve a favorable benefit-to-risk balance. Frequently this relies on use of predictive (treatment selection) biomarker-based tests to identify patients most likely to benefit from a therapy; consequently, evaluation requires not only assessment of an overall effect of an investigational therapy, but also the ability of the biomarker test to identify patients most likely to benefit. In oncology, these predictive biomarkers (broadly termed "molecular signatures") are typically assessed on tumors, using either single-analyte laboratory assays or more comprehensive omics technologies. Although sometimes the drug-signature pairing is quite natural (e.g., drug targets product of a gene that is mutated), mechanism of action for some drug classes (e.g., immunotherapy) is not straightforwardly tied to a single molecular entity. Moreover, because predictive biomarker test development is often not carried out synchronously, or with the same rigor, as for drugs, a completely standardized, robust predictive biomarker test might not be ready when the definitive therapy trial is initiated.
Several real-world examples are described to highlight some statistical challenges in planning, conducting, and analyzing precision medicine clinical trials. Specific examples of complexities include a) molecular signatures that cannot be identified, b) poor reproducibility of biomarker assays across laboratories, and c) continuous biomarker scores that lack decision cut-points for trial enrichment or stratification, or drug labelling. Discussion will focus on how these challenges have affected design and analysis of some recent oncology clinical trials and how, when not adequately addressed, confusing clinical guidance and drug labelling results.
Development of precision medicine approaches involves increased logistical and statistical complexity compared to conventional drug development. Attention to biomarker test readiness and statistical approaches to manage biomarker uncertainty and subgroup testing are essential to achieve reliable and interpretable evidence for the value of precision medicine approaches.

[1] Harris LN, Blanke CD, Erba HP, et al. The New NCI Precision Medicine Trials, 2023, in press.
[2] McShane LM, Rothmann MD, Fleming TR. Finding the (biomarker-defined) subgroup of patients who benefit from a novel therapy: no time for a game of hide and seek, Clinical Trials 2023, in press.

President's Invited Speaker

PLE2 On causal inference, estimands and trials in epidemiology and biostatistics

Didelez V.*
Leibniz Institute for Prevention Research and Epidemiology - BIPS ~ Bremen ~ Germany

Over the last decades, causal inference has addressed a variety of causal effect estimands. In many settings, relevant exposures are not binary point treatments but sustained or time-varying. Hence, the estimand should be formulated in view of what sustained, time-varying or adaptive treatment strategies we wish to compare. To be explicit, it is useful to formulate a target trial to be emulated with observational data. Interestingly, a duality has emerged: Causal inference approaches have attracted attention in the context of actual RCTs with intercurrent events (cf. ICH E9 Addendum). These alter either the intended meaning of treatment or outcome, which the chosen estimand must take into account.
I will give an overview using examples from epidemiological practice, ranging from estimating the (side)effects of therapies, effectiveness of cancer screening, to evaluating lifestyle recommendation preventing childhood obesity. While key assumptions on sufficient confounder information and overlap often appear hard to defend, approaches to strengthen their plausibility have been developed and are increasingly adopted. Implications and lessons-to-be-learned from observational causal inference for the design and analyses of actual RCTs with intercurrent events will be discussed, taking us back to dynamic or adaptive treatment strategies. However, sometimes notions of direct effects are invoked, despite being problematic to interpret; I will illustrate how separable effects might be an interesting alternative. Finally, the role of formulating a target trial will be considered. Target trial emulation is an important principle especially for eliciting actionable estimands, but also for avoiding self-inflicted biases in the analysis of observational data.
On the one hand, the analysis of RCTs with intercurrent events looks to learn from observational causal inference; on the other hand, the causal analysis of observational data should be guided by certain design principles of randomized trials. However, the emulation of a target trial with observational data and the planning and analysis of actual RCTs, have in common that the intended practical use of the results for decision making (by individuals, physicians or public health authorities) is crucial for the choice of causal estimand.

1) Braitmaier M, Kollhorst B, ... Haug U, Didelez V. Effectiveness of mammography screening on breast cancer mortality - A study protocol for emulation of target trials using German health claims data. Clinical Epidemiology. 2022;14:1293-1303.
2) Braitmaier M, ... Didelez V, Haug U. Screening colonoscopy similarly prevented distal and proximal colorectal cancer. A prospective study among 55-69-year-olds. Journal of Clinical Epidemiology. 2022;149:118-126
3) Didelez V. Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis. 2019;25(4):593-610.
4) Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable effects for causal inference in the presence of competing events. Journal of the American Statistical Association. 2022;117(537):175-183.

Invited Sessions | Monday 28 August 2023

MIN1 INVITED SESSION

ADVANCES ON CAUSAL INFERENCE IN LONGITUDINAL STUDIES

ORGANIZER | CHAIR: CÉCILE PROUST-LIMA

MIN1.1 Longitudinal outcome-adaptive and marginal fused lasso for model selection with time-varying treatments

Schnitzer M.*[1], Ertefaie A.[2], Talbot D.[3], Wang G.[4], Berger D.[1], O''Loughlin J.[1], Sylvestre M.[1]
[1]Université de Montréal ~ Montréal ~ Canada, [2]University of Rochester ~ Rochester ~ United States of America, [3]Université Laval ~ Québec ~ Canada, [4]Harvard University ~ Boston ~ United States of America

Data sparsity is a common problem when conducting causal inference with time-varying binary treatments, especially when treatment can change over many time-points. Many methods involve weighting by the inverse of the probability of treatment, which requires modeling the probability of treatment at each time point. Under sparsity, it is possible to pool these models over time, but when correlations between covariates and treatment vary over time, this can lead to bias. Furthermore, with a large covariate space assumed to be a non-minimal sufficient adjustment set, reducing the adjustment set can greatly improve the variance of the estimator. We consider a novel approach to longitudinal confounder selection using a longitudinal outcome adaptive fused LASSO that will data-adaptively select covariates and collapse the treatment model parameters over time-points with the goal of improving the efficiency of the estimator while minimizing confounding bias. We provide theory to justify the approach and demonstrate the estimator's finitesample performance in simulation studies.

MIN1.2 Proximal causal inference for separable effects with applications to aging research

Tchetgen Tchetgen E.*
The Wharton School ~ Philadelphia ~ United States of America

In the analysis of causal effects of modifiable exposures on dementia related outcomes such as cognitive decline in aging populations, a common challenge is that participants outcomes are subject to censoring or truncation by death. It is well known that censoring by death can lead to substantial selection bias if not appropriately addressed. We propose a general framework to address this important problem, which entails: (i) Formally defining causal effects adopting the recently proposed separable effects formulation; (ii) leveraging proxies of unmeasured common causes of mortality and cognitive decline to account for significant survival bias typically present in these settings
Separable effects hypothesize that exposures in view can be decomposed into components that only impact survival but not cognitive decline, and components that only impact cognitive decline but not survival. Such an assumption is established to not only be sufficient but essentially necessary for reasoning about causal effects in presence of truncation by death. For identification, we avoid making standard but unrealistic ignorabiliy types of assumptions by adopting and extending the proximal causal inference framework of Miao et al (2018) and Tchetgen Tchetgen et al (2020) to the truncation by death setting. The approach relies on available proxies of unmeasured common causes of cognitive decline and survival, which although informative about the latter, cannot be accounted for by standard regression analysis but instead require a more nuanced analysis. We give identification conditions and study semiparametric efficiency theory for estimation of causal effects in this important setting using such proxies when available. We establish new proximal causal inference methodology to account for truncation by death, obviating the need for standard ignorability conditions by leveraging available proxies of factors inducing selection bias.
Chan Park, Mats Stensrud, Eric J Tchetgen Tchetgen. "Proximal Causal Inference for Truncation by Death"

MIN1.3 Risk prediction under hypothetical interventions

Keogh R.*[1], Weir D.[2], Van Geloven N.[3]
[1]London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom, [2]Utrecht University ~ Utrecht ~ Netherlands, [3]Leiden University Medical Center ~ Leiden ~ Netherlands

Clinical risk prediction models enable predictions of a person's risk of an outcome given their observed characteristics. It is often of interest to use risk predictions to inform whether a person should initiate a particular treatment. However, when standard clinical prediction models are developed in a population in which patients follow a mix of treatment strategies, they are unsuitable for informing treatment decisions [1]. Risk prediction under hypothetical interventions aims to address this problem by providing estimates of what a person's risk would be if they were to follow a particular treatment strategy, and requires causal inference concepts and methods [2]. This talk will give an overview of methods for development of models for prediction under interventions and will present new methods for their validation. The methods will be illustrated using a new open-access synthetic data resource which we have designed to enable researchers to assess and compare methods for addressing different types of causal questions. The focus will be on use of longitudinal observational data, such as from electronic health records, to predict risk under interventions that are longitudinal treatment strategies. Several methods for development of interventional prediction models will be discussed, with a more detailed example focusing on marginal structural models. An essential step in development and reporting of any prediction model is to validate its performance. However, methods for validating standard clinical risk prediction models do not apply to interventional prediction models. I will present newly developed methods for assessing the predictive performance of interventional prediction models, including measures of discrimination and calibration. The methods will be illustrated using a synthetic data resource mimicking real world data based on a case-study about treatments for type-2 diabetes. Its features include multiple treatments, time-dependent confounders, time-to-event outcomes, and counterfactual outcomes under different interventions. Risk prediction under hypothetical interventions is a key goal in personalised healthcare. Causal inference methods for obtaining predictions under interventions exist and new techniques are emerging. Validation of predictions under interventions is a particular challenge, and this work presents general methods for validation.

[1] M. Sperrin, G.P. Martin, A. Pate, et al. Statistics in Medicine, 28, 2018, 4142-4154.
[2] N. van Geloven, S. Swanson, C.L. Ramspek, et al. European Journal of Epidemiology, 7, 2020, 619-630.

MIN2 INVITED SESSION

INNOVATIVE DESIGNS FOR DOSE OPTIMIZATION STUDIES

ORGANIZER | CHAIR: EMILY ZABOR

MIN2.1 The use of master protocol designs with dose-optimization studies

Kaizer A.*[1], Zabor E.[2]
[1]University of Colorado Anschutz Medical Campus ~ Aurora ~ United States of America, [2]Cleveland Clinic ~ Cleveland ~ United States of America

Master protocol designs, such as basket, umbrella, and platform trials, have seen rapid growth in the interest and application to biomedical research problems. For example, oncology has been at the forefront of exploring the use of these designs across a range of indications. These master protocol designs are being considered to address limitations of more traditional trial designs that may need a series of standalone studies or separate protocols to answer related or identical research questions. In this talk, we explore the potential use of these master protocol designs for dose-optimization studies. One such approach is the potential of sharing information across indications with the same hypothesized molecular targets, across dose-levels if treatment response is exchangeable because of the lack of increasing response with increasing dose, or by incorporating prior studies that may include preliminary dose-response information or safety details. Master protocols may also be used to expand promising doses to expansion cohorts rather than closing out one study to initiate a later trial. Additionally, master protocols may facilitate the seamless transition from an earlier phase to a later phase research study, with the ability to collect longer-term safety data. The properties and performance of these designs will be summarized through simulation studies with comparison to traditional approaches to dose-optimization and subsequent research studies that are not related to master protocol designs.

Invited Sessions | Monday 28 August 2023

MIN2.2 Demo: bayesian adaptive dose exploration–monitoring–optimization design based on short, intermediate, and long-term outcomes

Lin R.*
The University of Texas MD Anderson Cancer Center ~ Houston ~ United States of America

The conventional "more-is-better" paradigm in oncology has been challenged with the advancement of targeted therapy. The US Food and Drug Administration (FDA) launched Project Optimus and released the draft guideline to reform the dose selection and optimization process. Because of the deadly nature of cancer, the gold-standard endpoint to evaluate a cancer treatment is a long-term survival endpoint, such as the overall survival or progression-free survival. An optimal dose recommended for late phases should ultimately possess a promising enough survival profile. Existing early-phase trial designs that rely on short-term toxicity and efficacy data may identify a dose that may result in suboptimal long-term survival benefits. However, using long-term survival endpoint in trial monitoring requires a longer follow-up time, and thus results in a prolonged trial duration. To address these challenges, a generalized dose optimization procedure consisting of three seamlessly connected stages is proposed. In the first stage, the short-term pharmacodynamics biomarker and toxicity endpoints are utilized to screen out overly toxic doses and doses that do not yield bioactivity. In the second stage, patients are treated at admissible doses identified in the first stage to collect accumulating toxicity and efficacy data for dose monitoring. In the third stage, patients are randomized to the doses in the refined admissible dose set to identify the optimal dose through restricted mean survival time.
Results from our simulation study indicate the proposed dose optimization design outperforms conventional approaches and is robust to changes in prior specifications. The new method is examplified using a real-world application derived from the DREAMM-1, -2, and -3 trials.
Yang, C. H., Thall, P. F., and Lin, R. (2023). DEMO: Bayesian Adaptive Dose Exploration–Monitoring– Optimization Design based on Short, Intermediate, and Long-term Outcomes. In preparation.

MIN2.3 Controlled amplification in oncology dose-finding trials

Dehbi H.*[1], O"Quigley J.[1], Iasonos A.[2]
[1]University College London ~ London ~ United Kingdom, [2]Memorial Sloan Kettering Cancer Center ~ New York ~ United States of America

In oncology clinical trials the guiding principle of model-based dose-finding designs for cytotoxic agents is to progress as fast as possible towards, and identify, the dose level most likely to be the MTD. Recent developments with non-cytotoxic agents have broadened the scope of early phase trials to include multiple objectives. The ultimate goal of dose-finding designs in our modern era is to collect the relevant information in the study for final RP2D determination.
While some information is collected on dose levels below and in the vicinity of the MTD during the escalation (using conventional tools such as the Continual Reassessment Method for example), designs that include expansion cohorts or backfill patients effectively amplify the information collected on the lower dose levels. This is achieved by allocating patients to dose levels slightly differently during the study in order to take into account the possibility that "less (dose) might be more". The objective of this paper is to study the concept of amplification. Under the heading of controlled amplification we can include dose expansion cohorts and backfill patients among others.
We make some general observations by defining these concepts more precisely and study a specific design that exploits the concept of controlled amplification.
Dehbi, Hakim-Moulay, John O'Quigley, and Alexia Iasonos. "Controlled backfill in oncology dosefinding trials." Contemporary Clinical Trials 111 (2021): 106605.

MIN3 INVITED SESSION

VACCINATION PROGRAMMES: POST - IMPLEMENTATION ASSESSMENT OF PROTECTION, BENEFITS AND RISKS

ORGANIZER | CHAIR: MARIE REILLY

MIN3.1 Statistical methods for the epidemiological evaluation of vaccine safety

Whitaker H.*
UK Health Security Agency ~ London ~ United Kingdom

Post-marketing vaccine safety surveillance will be covered in brief before introducing statistical methods used for the epidemiological evaluation of vaccine safety signals or concerns that have been identified. The aim is to give the audience an overview of methods used in this context.
The self-controlled case series (SCCS) method is commonly used to estimate the incidence of adverse events during hypothesized post vaccination risk time windows relative to the incidence of events outside these time windows, before and after vaccination. The SCCS method will be introduced; inference is within-individual, cases act as their own controls and any fixed confounders are controlled for explicitly, but there are several limiting assumptions. SCCS will be contrasted with traditional methods such as cohort and case-control designs, with examples. Some examples of methodology used in recent studies that evaluate the safety of COVID-19 vaccines will be discussed. The merits and limitations of methods used in vaccine safety evaluation will be recapped.
The self-controlled case series website: sccs-studies.info

MIN3.2 Monitoring and evaluating covid-19 vaccination programmes: real world challenges

Hahné S.*, De Gier B., Van Werkhoven H., Ainslie K., Van Den Hof S., Knol M., De Melker H.
RIVM ~ Bilthoven ~ Netherlands

In late November 2020, the first estimates of COVID-19 vaccine efficacy from clinical trials were published. These efficacy estimates were very important to provide the first information on the protection conferred by COVID-19 vaccines. However, a range of observational epidemiological studies is required to provide evidence on the benefits and risks of the vaccination programme in real life. In this, a number of methodological and other challenges were encountered. In this session I will initially outline the public health importance of monitoring and evaluating vaccination programmes throughout their lifetime. Subsequently, I will go through a number of real world challenges encountered while assessing the Dutch COVID-19 vaccination programme :
1. Legal and other impediments to linking and accessing data ;
2. Balancing scientific rigour and timeliness of results;
3. Large and rapidly changing time effects: change of SARS-CoV-2 variant predominance, large changes in incidence of infection, build up of natural immunity, establishment and waning of vaccine induced immunity;
4. Misclassification of vaccination and disease outcome status;
5. Confounding by differential behaviour between vaccinated and unvaccinated individuals.
I will discuss the context of these challenges, several examples and, where possible, an assessment of their impact and solutions. Regardless of information available from vaccine trials, monitoring and evaluating vaccination programmes is essential throughout their lifetime . To obtain real world estimates of key indicators such as vaccine coverage and vaccine effectiveness, several challenges have to be faced, which can sometimes partly be overcome. International collaboration and rapid exchange of information is crucial to provide decision makers with the best evidence possible to optimize the programme.

MIN3.3 Statistical methods to assess (immunological) surrogate endpoints for Vaccines

Callegaro A.*
GSK ~ Rixensart ~ Belgium

A correlate of protection (CoP) is an immune marker that can be used to reliably predict a vaccine's level of efficacy in preventing a clinically relevant outcome. The identification of a CoP constitutes an important success because the availability of substitute (surrogate) endpoints are important for vaccine development, licensure and effectiveness monitoring. Many statistical methods have been proposed for the evaluation of CoPs, such as the Prentice framework [1], the meta-analytical framework [2] and more recently methods from causal inference. In this talk, we present an overview of definitions, statistical methods, and challenges. For illustration, we consider examples from Covid-19 vaccines. The use of immunological markers as surrogate endpoints for vaccine evaluation is important, but complex. It is not straightforward to identify such markers or to ensure that the immunological prediction is accurate. Statistics and statisticians play a key role on this key aspect of vaccinology.

[1] Prentice, R. L., Statistics in Medicine, 8 (4), 1989, 431–40.
[2] Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys, Biostatistics, 1, 2000, 49–67.

PARALLEL SESSION MO1: CLINICAL TRIALS 1

MO1.1 Adaptive seamless designs in the two-trial paradigm: advantages and limitations

Zhan S.J.*[1], Stallard N.[1], Kunz C.U.[2]
[1]University of Warwick ~ Coventry ~ United Kingdom, [2]Boehringer Ingelheim Pharma GmbH & Co. KG ~ Biberach/Riss ~ Germany

Adaptive seamless phase II/III designs have emerged as one of the most common adaptive designs to accelerate drug development while reducing the number of patients needed in phase III trials. The benefits of these designs have been widely discussed, especially in the context of a single trial. However, regulatory requirements typically mandate at least two significant independent pivotal trials as substantial evidence for a new drug's effectiveness. Incorporating adaptive seamless designs in this conventional two-trial paradigm is a challenge. Our work aims to explore various approaches to conduct one or two pivotal seamless phase II/III trials. Commonly, one pivotal trial is designed as a seamless phase II/III trial with a second pivotal trial starting after the phase II stage with a conventional design. However, this approach does not reduce the overall development time. Therefore, questions arise on how best to conduct two adaptive seamless phase II/III trials in parallel. As the two-trial paradigm remains a controversial topic, we discuss the pros and cons of the approaches, focusing on whether combining the trials or/and phases would be appropriate under different scenarios. We compare the operating characteristics of the methods with the aim of reducing the time to approval of a new drug and the number of patients. Our results indicate that the size of phase II plays an important part and a tailored approach is needed for each trial scenario. Our study provides insights into the implementation of adaptive seamless phase II/III designs in the two-trial paradigm and suggests that combining trials or phases may not always be suitable.

[1] S. J. Zhan, C. U. Kunz, N. Stallard, Pharmaceutical Statistics, vol. 22(1), 2023, pp. 96-111

MO1.2 Power calculations for multi-arm multi-stage trials with multicomponent disease rating scales outcomes [+]

Burnell M.[1], Carpenter J.*[2]
[1]MRC Clinical Trials Unit at UCL ~ London ~ United Kingdom, [2]London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom

In many diseases, multi-component rating scores are central to assessing severity and rate of disease progression. Specifically, in Parkinson's disease, we consider the first three components of the Movement Disorder Society United Parkinson's Disease Rating Scale (MDS-UPDRS). Parts IA (clinician administered), IB and II (completed by people with Parkinson's and their carers) assess respectively non-motor and motor experiences of daily living; part III is completed by a clinician following a motor examination. In this work, building on [1, 2], we describe an analytic approach to power calculations for large Phase III multi-arm multi-stage (MAMS) trials which use such scores as the primary outcome measure. Two key components of MAMS designs are that (i) early analyses, using a mix of outcome measures, allow the possibility to discontinue an arm if there is no evidence of benefit and (ii) towards the end of the study we have the option of stopping early for efficacy on the primary outcome. In particular, we use a random slope model for each component of the score over time, and describe power calculations targeting a common proportional reduction in the rate of change. We further show how these calculations can be adapted to allow for a non-constant recruitment and withdrawal rate, and how we can calculate the correlation between successive estimates of the treatment effect over time. We consider the power of our design in the event that the treatment effect is non-proportional. We illustrate our results with the design of a multi-arm multi-stage trial in Parkinson's disease, using the MDS-UPDRS as the outcome measure. We believe our approach provides a practical, flexible method for calculating sample sizes for MAMS trials targeting a reduction in the rate of increase of multi-component disease rating scales.
[+] Presented on behalf of the Trial Design Working Group of the EJS-ACT-PD collaboration (https://ejsactpd.com/) comprising Tom Barber, Roger Barker, Yoav Ben-Schlomo, Camille Carroll, Caroline Clarke, Carl Counsell, Mark Edwards, Thomas Foltynie, Anna Jewell, Cristina Robes, Dorothy Salathiel, Anette Schrag, Sue Whipps, Alan Whone and Marie Zeissler

[1] Frost C, Kenward MG and Fox NC (2008) Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? Statistics in Medicine 28: 3717-3731.
[2] Nash S, Morgan KE, Frost C and Mulick A. (2021) Power and sample-size calculations for trials that compare slopes over time: introducing the slopepower command. The Stata Journal, 21:3, 575-601

MO1.3 Fast simulation of bayesian adaptive designs using the laplace approximation

Heritier S.*[1], Ryan E.[1], Jaki T.[2], Couturier D.[3]
[1]*School of Public Health and Preventive Medicine, Monash University ~ Melbourne ~ Australia,* [2]*Chair for Computational Statistics, Faculty of Informatics and Data Science, University of Regensburg ~ Regensburg ~ Germany,* [3]*MRC Biostatistics Unit and Cancer Research UK – Cambridge Institute, University of Cambridge ~ Cambridge ~ United Kingdom*

The use of Bayesian adaptive designs has been hindered by the lack of software readily available to statisticians. Part of the problem is due to the burden generated by Markov Chain Monte Carlo (MCMC) methods, which are typically used to compute posterior distributions. In this work, we follow a different approach based on the Laplace approximation to circumvent MCMC. The main aim of this project is to provide a flexible structure for the fast simulation of Bayesian adaptive designs. Integrated nested Laplace approximation (INLA)[1] allows one to perform approximate Bayesian inference in latent Gaussian models including generalised linear models and survival analysis models. We developed the BATS package (Bayesian Adaptive Trials Simulator) by taking advantage of the general structure provided by INLA[2] and its computationally efficient implementation in R. We focus our attention on Bayesian multi-arm multi-stage (MAMS) designs as a first step. We show that BATS is an effective tool to study the operating characteristics of a such designs for various endpoints and most common adaptations: stopping arms for efficacy or futility, fixed or response-adaptive randomisation, with user-defined rules. Other important features include: parallel processing, customisability, use on a cluster computer or PC/Mac, adjustment for independent covariates and ANCOVA. The use of INLA in Bayesian adaptive designs is promising. The BATS package provides a flexible simulation framework for Bayesian adaptive designs. It has already been used successfully to design trials accepted by Australian funding bodies such as the NHMRC or MRFF.

[1] Rue H, Martino S, Chopin N (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. JRSS-B, 71(2), pp 319-392.
[2] Van Niekerk J, Bakka H, Rue H, Schenk O (2021). New Frontiers in Bayesian Modeling Using the INLA Package in R. Journal of Statistical Software, 100, 1–28.

MO1.4 A bayesian decision-theoretic randomisation procedure and the impact of delayed responses

Williamson S.F.*[1], Jacko P.[2], Jaki T.[3]
[1]*Biostatistics Research Group, Newcastle University ~ Newcastle ~ United Kingdom,* [2]*Lancaster University ~ Lancaster ~ United Kingdom,* [3]*Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom*

The design of sequential experiments and, in particular, randomised controlled trials involves a trade-off between operational characteristics such as statistical power, estimation bias and patient benefit. A randomisation procedure based upon a dynamic programming solution, referred to as Constrained Randomised Dynamic Programming (CRDP), can be used to balance these competing objectives [1]. This is particularly well-suited to rare disease trials, which typically include a substantial proportion of all patients with the disease and therefore seek to improve patient benefit within the trial whilst maintaining sufficient power. CRDP, as with most response-adaptive randomisation procedures, hinges on the limiting assumption of patient responses being available before allocation of the next patient. This forms one of the greatest barriers to their use in practice. Therefore, motivated by the existing gap between the theory of response-adaptive randomisation (which is abundant with proposed methods in the immediate response setting) and clinical practice (in which responses are typically delayed), we evaluate the performance of CRDP in the presence of fixed and random delays. To explore the impact of delay on CRDP, we use simulation to evaluate its performance in a range of scenarios for different delay lengths. Results show that CRDP performs well in the presence of delayed responses with slight gains in power and some loss in patient benefit as delay increases. However, when compared to alternative randomisation procedures with delay, CRDP continues to offer patient benefit gains [2]. To compensate for a fixed delay, we suggest an adjustment to the time horizon used in the optimisation objective and illustrate its performance. We also consider extending the underlying Markov decision process (MDP) to incorporate state variables representing information on the pipeline patients. The generalisation of CRDP highlights its inherent flexibility to adapt to a variety of practicalities that may be encountered during the design of clinical trials. This work has provided insight into the important – and commonly asked – question of how CRDP performs when responses are delayed. In particular, we have demonstrated that CRDP is relatively robust to delayed responses. When pipeline information is incorporated into the MDP model, there are further patient benefit gains.

[1] S.F. Williamson, P. Jacko, S. S. Villar, T. Jaki, A Bayesian adaptive design for clinical trials in rare diseases. Computational Statistics & Data Analysis, Volume 113, 2016, 136–153
[2] S. F. Williamson, P. Jacko, T. Jaki, Generalisations of a Bayesian decision-theoretic randomisation procedure and the impact of delayed responses, Computational Statistics & Data Analysis, Volume 174, 2022, 107407

Parallel Sessions | Monday 28 August 2023

MO1.5 Evaluating the impact of outcome delay on the efficiency of two-arm group-sequential trials

Mukherjee A.*, Grayling M., Wason J.
Newcastle University ~ Newcastle Upon Tyne ~ United Kingdom

Adaptive designs(AD) are a broad class of designs that allow modifications to be made to a trial as patient data is accrued. ADs can offer improved efficiency and flexibility. However, a delay in observing the primary outcome variable can potentially harm the efficiency added by ADs. Principally, in the presence of such delay, we may have to make a choice as to whether to (a) pause recruitment until requisite data is accrued for the interim analysis, leading to a longer trial completion period; or (b) continue to recruit patients, which may result in a large number of participants who do not benefit from the interim analysis. Little work has been conducted to ascertain the size of outcome delay that results in the realised efficiency gains of ADs being negligible compared to classical fixed-sample alternatives. We perform such work here for twoarm group-sequential designs with different numbers of interim analyses. We measure the impact of outcome delay by developing formulae for the number of `overruns' (or `pipeline' patients) in two-arm group-sequential trials with normal data, assuming different recruitment models. Typically, the efficiency of a group-sequential trial is measured in terms of the expected sample size (ESS), with group-sequential designs generally reducing the ESS compared to a design without interim analyses. Our formulae enable us to measure the efficiency gain from the group-sequential design in terms of ESS reduction that is lost due to outcome delay. We assess whether careful choice of design (e.g. choosing the spacing of the interim analyses) can help recover the advantages of group-sequential designs in the presence of outcome delay. We similarly analyse the efficiency of group-sequential design with respect to time to complete the trial.
A delay in observing the treatment outcome is harmful to the efficiency of a group-sequential design. Also, for unequally spaced interims, pushing the first and subsequent interims towards the latter end of the trial can potentially be harmful for the design in presence of delay. However, if the measure of efficiency is the time to complete a trial, a group-sequential design most likely to provide benefit compared to a single-stage design.

[1] Hampson, L.V. and Jennison, C. (2013), Group sequential tests for delayed responses (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75: 3-54. https://doi.org/10.1111/j.1467-9868.2012.01030.x
[2] A. Mukherjee, J. M. S. Wason, and M. J. Grayling, "When is a two-stage single-arm trial efficient? An evaluation of the impact of outcome delay," Eur J Cancer, vol. 166, pp. 270–278, May 2022, doi: 10.1016/j.ejca.2022.02.010.

Parallel Sessions | Monday 28 August 2023

PARALLEL SESSION MO2: SURVIVAL ANALYSIS 1

MO2.1 Perils of rct survival extrapolation with effect waning: why marginal and conditional estimates differ

Jennings A.*[1], Rutherford M.[1], Latimer N.[2], Sweeting M.[3], Lambert P.[1]
[1]*University of Leicester ~ Leicester ~ United Kingdom,* [2]*University of Sheffield ~ Sheffield ~ United Kingdom,* [3]*AstraZeneca ~ Cambridge ~ United Kingdom*

A long-term, constant, protective treatment effect is a strong assumption when extrapolating survival from RCT (randomised controlled trial) end to a lifetime horizon. To alleviate this, sensitivity to treatment effect waning is commonly assessed for economic evaluations. Simply forcing a hazard ratio (HR) to 1, however, does not necessarily estimate loss of individual-level treatment effect accurately, due to the inherent selection bias and non-collapsibility of the HR[1,2]. A simulation study was designed to (1) explore the behaviour of the marginal HR under a waning conditional (individual-level) treatment effect and (2) investigate the impact of forcing a marginal HR to 1 when the estimand is treatment effect with individual-level waning. Data were simulated under 2 conditional treatment effect waning scenarios (instant/steady) with 4 parameter combinations each (varying the prognostic strength of heterogeneity and treatment effect). Flexible parametric time-varying conditional/marginal HR estimates were calculated for visual inspection. Restricted mean survival time (RMST) differences, estimated having constrained the marginal HR to 1, were compared to true values to assess the bias in 'treatment effect under individual-level treatment effect waning' estimates induced by the constraint of a marginal estimate. Under loss of individual treatment effect, the marginal HR took a value greater than 1 due to post-randomisation covariate imbalances in the surviving samples. Constraining this value to exactly 1 with weak prognostic strength of heterogeneity lead to small bias in RMST difference but this bias increased up to 0.9 years (69% increase) with strong prognostic strength in heterogeneity. Effect size inflation also increased with the size of protective treatment effect. Important differences exist between extrapolations that assume marginal versus conditional treatment effect waning. How waning is modelled should be carefully considered. Conditional HR constraints more closely assess loss of individual-level treatment effect, with methods available to return to marginal estimates if preferred.

[1] M. Hernán, The hazards of hazard ratios, 21(1), 2010, 13.
[2] R. Daniel, J. Zhang, D. Farewell, Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets, 63(3), 2021, 528-557.

MO2.2 Can we trust the hazard ratio? -Causal consequences of observed proportional hazards

Valberg M.*, Brathovde M., Aalen O.O.
University of Oslo ~ Oslo ~ Norway

The hazard ratio (HR) is by far the most frequently used effect measure in applied time-to-event analyses. However, the causal interpretation of the HR is not necessarily simple, even in the context of randomized controlled trials (RCTs). We demonstrate that the HR has a built-in selection bias, even when all assumptions of Cox' proportional hazard model are satisfied. Furthermore, we investigate the direction of the bias in both a single time-to-event setting and in a competing risks setting. We use a frailty modelling approach to investigate the implications of observed proportional hazards on the causal interpretability of the HR. In this setting, we show that an observed, timeconstant HR will be an attenuated version of the treatment/exposure effect in individuals, even in a simple two-armed RCT. In the competing risk setting, the causal interpretation of the observed HR is even more challenging. If the same underlying factors affect the different competing risks, then we show that the observed cause-specific HR can be both overestimating and underestimating the treatment/exposure effect. Thus, treatments that are harmful or beneficial, respectively, may appear as having the opposite effect if the observed cause-specific HRs are used as the measures of effect. Making causal conclusions based on HRs are challenging, even in RCTs and even when all assumptions of the Cox model are satisfied.

MO2.3 A non-parametric proportional risk model to assess a treatment effect in an application to survival data

Ameis L.*[1], Kuß O.[2], Hoyer A.[3], Möllenhoff K.[1]
[1]*Mathematical Institute, Heinrich Heine University Düsseldorf ~ Düsseldorf ~ Germany,* [2]*German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich Heine University Düsseldorf, Institute for Biometrics and Epidemiology ~ Düsseldorf ~ Germany,* [3]*Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University ~ Bielefeld ~ Germany*

Time-to-event analysis often relies on prior parametric assumptions or, if a non-parametric approach was chosen, Cox's proportional hazards model that is inherently tied to an assumption of proportional hazards. This limits the quality of the results in case of any violation of these assumptions. Especially the assumption of proportional hazards was recently criticized for being rarely verified. In addition, most interpretations focus on the hazard ratio, that is often misinterpreted as the relative risk and comes with the restriction of being a conditional measure. Our approach introduces an alternative to the proportional hazard assumption and allows for a direct estimation of the relative risk as well as the absolute measure of the number needed to harm. In this talk, we propose a new non-parametric estimator to assess the relative risk of two groups to experience an event under the assumption that the risk is constant over time, namely the proportional risk assumption. Precisely, we first estimate the respective cumulative distribution functions of both groups by means of the Kaplan-Meier estimator and second combine their ratio at different time points to estimate the mean relative risk. We then combine the result with one of the estimated cumulative distribution functions to assess the number needed to harm. This offers the possibility to interpret the treatment effect solely based on a Kaplan-Meier estimator and offers a flexible alternative to Cox's model if the proportional hazard assumption is violated. We demonstrate the validity of the approach by means of a simulation study and present an application to mortality data of mice from a study investigating the long-term carcinogenicity of piperonyl butoxide [1]. Our approach allows to directly estimate the relative risk as well as the number needed to harm and therefore provides the possibility of an easy and holistic interpretation.

[1] National Toxicology Program. Bioassay of piperonyl butoxide for possible carcinogenicity (CAS No. 51-06-6 / NCI-CG-TR-120). Technical report, 1979.

MO2.4 A reduced rank proportional hazards model for age-related multimorbidity event data

Sluiskes M.*, Goeman J., Putter H., Rodríguez--Girondo M.
Leiden University Medical Center ~ Leiden ~ Netherlands

The identification of biomarkers of aging is an important biomedical research theme. Most current statistical methods that aim to capture the aging process either use chronological age or time-to-mortality as the outcome of interest. There is however a shift in the field towards the study of health span and patterns of age-related multi-morbidity, as aging entails more than lifespan duration alone. Several large epidemiological studies, such as the UK Biobank and the Leiden Longevity Study, have recently incorporated detailed age-at-disease-onset profiles, obtained from electronic health records. The availability of these data opens new analytical possibilities. Nevertheless, analyses conducted thus far oversimplify the complexity of multi-morbidity patterns, for instance by ignoring information on age-at-disease-onset or by failing to acknowledge that age-related diseases are likely driven by a shared set of underlying factors. We propose a new methodological framework for the analysis of age-related multi-morbidity data, based on multiple-outcome survival modelling. Specifically, we propose to use a reduced rank proportional hazards model. This model can be fitted on the (possibly right-censored and left-truncated) age-at-disease-onset of several age-related diseases simultaneously. It assumes that there is a set of shared latent factors that drive all age-related diseases considered, thereby reducing the dimensionality of the problem and providing additional insight into different facets of the aging process. As there is a large interest in the use of high-dimensional omics data as potential biomarkers of aging, we also propose some ideas to include penalization in the reduced rank proportional hazards framework. The use and intuitive interpretation of the reduced rank proportional hazards model is illustrated by applying it to age-related multimorbidity and mortality data from the UK Biobank, using metabolomics data as predictor variables. Comparison of the reduced rank model to simpler alternative models is shown. The reduced-rank proportional hazards model is a useful statistical framework with clear added value in the context of age-related multimorbidity event data.

Parallel Sessions | Monday 28 August 2023

MO2.5 Multicasanova - multiple group comparisons for non-proportional hazard settings

Dormuth I.*[1], Pauly M.[1], Konietschke F.[3], Ditzhaus M.[2], Herrmann C.[3]
[1]TU Dortmund University ~ Dortmund ~ Germany, [2]Otto-von-Guericke- Universität ~ Magdeburg ~ Germany, [3]Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin ~ Berlin ~ Germany

When comparing multiple groups in clinical trials, strong assumptions such as proportional hazards are violated even more often than when we compare two groups. In the two-group context, some more robust alternatives to the log-rank test have already been developed and investigated. A promising group of tests for the two-sample case is the so-called combination tests such as the maxCombo or CASANAOVA [1] test. An additional challenge in multi-group comparisons is the distinction between global and local differences. We usually want to know not only whether there is a difference between the groups but also where this difference is. Hence, we are testing multiple individual hypotheses. Simple corrections, such as Bonferroni, control the familywise type I error when testing multiple null hypotheses but are usually conservative. We propose a new multiple contrast test based on the CASANOVA approach. The new group of multiCASANOVA tests uses maximum statistics to combine multiple weighted log-rank tests. These tests are more robust towards violations of the proportional hazards assumption and makes p-value corrections obsolete. We evaluate the performance of the tests with the Bonferroni corrected log-rank test, in extensive Monte-Carlo Simulation studies. These simulations cover proportional as well as crossing and non-proportional but not crossing scenarios. The results show a meaningful gain in power when using the multiCASANOVA approach particularly in non-proportional hazard situations. We present a new test for multiple testing problems for time-to-event data, which does not rely on proportional hazards assumptions and does not require p-value adjustment.

[1] Ditzhaus, M., Genuneit, J., Janssen, A., & Pauly, M. (2021). CASANOVA: Permutation inference in factorial survival designs. Biometrics.

PARALLEL SESSION MO3: PREDICTION MODEL 1

MO3.1 Tuning the regularization parameter in penalized regression: an approach based on false selection rate

Renzetti S., Rota M.*, Sandri M., Vezzoli M., Calza S.
University of Brescia ~ Brescia ~ Italy

Variable selection is one of the most pervasive problems in statistical applications. Many pattern recognition and regression algorithms were originally not designed to cope with large amounts of non-informative variables. Combining these methods with variable selection has become necessary in many applications. The most commonly used methods for selecting the penalty parameter are cross-validation, information criteria (AIC, BIC, Mallow's Cp) and Bayesian methods. The aim of our study is to propose a novel criterion for tuning the penalty parameter by using a criterion based on the rate of non-informative variables (regression parameter equal to 0) that are erroneously selected (the false selection rate, FSR) for methods like lasso, elastic net and random forest. The method involves estimating the curve of the FSR as a function of the penalty, and then selecting the penalty that corresponds to the maximum allowable FSR (determined according to the researcher's preferences and priorities). An essential element of the FSR estimator is the rate θ that uninformative variables enter the model as the penalty parameter decreases. One way to estimate this rate involves generating a set of pseudo-variables that mimic the unknown noninformative variables present in the original data set, as proposed in [1]. By observing the selection rate of these pseudo-variables for various penalty values, it is possible to estimate the rate θ and then the FSR function for different values of the penalty. The proposed algorithm comprises the following steps: (a) initial estimation of the number of unimportant variables; (b) repeated generation of pseudo-variables and fitting of penalized regression models on the augmented design matrices; (c) estimation of the rate θ for different penalty values using the mean rate of selected pseudo-variables. The proposed FSR-based approach to penalty estimation shifts the focus from prediction to association, making it particularly suitable for studies seeking to identify relationships between variables and the outcomes of interest. Extensive numerical simulations showed that the method effectively controls the FSR across a broad range of penalized regression models and data generating processes.

[1] Y. Wu, D.D. Boos, L.A. Stefanski. Controlling variable selection by the addition of pseudovariables. Journal of the American Statistical Association, 102, 2007, 235–243.

MO3.2 Penalised regression methods with modified tuning produce better prediction models

Pavlou M.*, Omar R., Ambler G.
University College London ~ London ~ United Kingdom

Risk prediction models fitted using maximum likelihood estimation (MLE) tend to be overfitted when the sample size is small, resulting in predictions that are too extreme (too close to 0 or 1) and an estimated calibration slope (CS) much less than 1. Penalised methods, such as Ridge and Lasso, have been suggested as a solution to this problem as they tend to shrink regression coefficients towards zero resulting in predictions closer to the average. The amount of shrinkage is regulated by a tuning parameter, λ, commonly selected via cross-validation. Though penalised methods have been found to improve calibration on average, they often over-shrink and exhibit large variability in the selected λ and hence, the estimated CS. Furthermore, they have been found to be worse than MLE when both the bias and variance of the CS are considered, especially for small sizes but also at sample sizes recommended to minimise overfitting. We investigate a modified tuning method to improve selection of λ and hence, improve the performance of the produced models. We consider whether these problems are partly due to selecting λ using cross-validation with 'training' datasets of reduced size compared to the original development sample, resulting in an over-estimation of λ and hence, excessive shrinkage. For example, in the commonly used 10- fold cross-validation, the size of each the cross-validation 'training' sets is 10% smaller than the original dataset. For small development datasets, this reduction in size can be non-negligible when it comes to optimising the tuning parameter. We propose a modified tuning method to circumvent these problems. We evaluated our tuning method for Ridge and Lasso regression using simulated and real data with binary outcomes. We found that the modified tuning method substantially reduced the variability in the selected λ and estimated CS for both Ridge and Lasso, compared to the standard tuning method. Penalised methods with the modified tuning can result in improved performance of the produced risk models compared to the standard tuning method. They can also result in improved risk models compared to MLE at recommended sample sizes to minimise overfitting.

1. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. 2019; 38: 1276-1296. DOI: 10.1002/sim.7992.
2. Martin GP, Riley RD, Collins GS, et al. Developing clinical prediction models when adhering to minimum sample size recommendations: The importance of quantifying bootstrap variability in tuning parameters and predictive performance. Statistical methods in medical research 2021; 30: 2545-2561. 2021/10/09. DOI: 10.1177/09622802211046388.
3. Van Calster B, van Smeden M, De Cock B, et al. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. 2020; 29: 3166-3178. DOI: 10.1177/0962280220921415. Bayesian theory and computation

MO3.3 An introduction to projection predictive variable selection

Weber F.*[1], Vehtari A.[2], Glass Ä.[1]
[1]Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center ~ Rostock ~ Germany, [2]Department of Computer Science, Aalto University ~ Espoo ~ Finland

The projection predictive variable selection for Bayesian regression models, implemented in the R package projpred [1], has been shown to possess excellent properties in terms of the tradeoff between predictive performance and sparsity (see, e.g., [2]). It also allows for valid postselection inference by retaining all uncertainty inherent to the so-called reference model, also known as the actual belief model, that serves as a yardstick in terms of predictive performance and guides the projection, thereby filtering noise from the observed response values. Unfortunately, projection predictive variable selection still seems to be applied only rarely in practice. Thus, the aim of this paper is to introduce projection predictive variable selection to a wider audience. By going through a real-world example step-by-step, we illustrate projection predictive variable selection in practice. Thereby, we also highlight advantages and pitfalls of the methodology. Our real-world example allows applied researchers with basic knowledge of the R programming language to conduct a projection predictive variable selection and associated post-selection analyses themselves.

[1] J. Piironen, M. Paasiniemi, A. Catalina, F. Weber, A. Vehtari. projpred: Projection predictive feature selection. R package, version 2.4.0, 2023. https://mc-stan.org/projpred/
[2] J. Piironen, A. Vehtari. Comparison of Bayesian predictive methods for model selection.

Parallel Sessions | Monday 28 August 2023

MO3.4 Confidence interval estimation for selected and unselected predictors after variable selection

Akbari N.*[2], Heinze G.[1]
[1]*Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23 ~ Vienna 1090 ~ Austria, 2Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Biometry and Clinical Epidemiology, Charitéplatz 1 ~ Berlin, 10117 ~ Germany*

This study aimed to investigate if a modified sandwich estimator of the covariance could help achieve valid confidence intervals for regression coefficients after variable selection. In regression models, predictors can either be pre-specified or selected through data-driven methods. Valid confidence intervals for regression coefficients can be estimated for prespecified predictors based on established theory. However, for data-driven methods, there is no widely accepted way to compute confidence intervals that encompass the uncertainty of selection. In particular, no methods have been proposed which estimate confidence intervals for non-selected predictors, and the uncertainty in selection is commonly ignored. We performed a simulation study in the setting of linear regression with five true and five noise predictors, using two correlation structures between predictors and five gradually increasing values for the residual variance. We compared five methods for obtaining confidence intervals for the regression coefficients of all candidate predictors, regardless of selection. These methods included using the coefficients and model-based variance of the full model, the final selected model with model-based variance and collapsed confidence intervals for non-selected predictors, the final selected model combined with the sandwich method modified to supply confidence intervals for all candidate predictors, a bootstrap variance method, and bootstrap percentile confidence intervals. The study evaluated two correlation structures between the candidate predictors, each combined with five gradually increasing residual variances. Results showed that coverage rates for the full model attained the claimed levels, while those for the selected model with model-based variances were much too small, particularly for correlated predictors. The bootstrap methods performed well for non-predictors and strong predictors but slightly worse for weak predictors. The sandwich method's coverage rates were close to nominal for non-predictors and slightly too small for true predictors. In any case, the modified sandwich estimator improved over the common practice of ignoring uncertainty induced by variable selection, but more research is needed to refine the method. This work was supported through the joint German-Austrian DFG and FWF project [DFG: RA- 2347/8-1] to Geraldine Rauch, [DFG: BE-2056/22-1] to Heiko Becher and [FWF: I-4739-B] to Daniela Dunkler. SANDWICH ESTIMATOR in Alt, F. B., Kotz, S., & Johnson, N. L. (1985). Encyclopedia of statistical sciences. New York, NY: John Wiley and Sons

PARALLEL SESSION MO4: BIOMARKERS

MO4.1 Validation of a skewed surrogate endpoint for a time-to-event outcome: the use of a zaga distribution

Risca G.*[2], Capitoli G.[2], Rotolo F.[1], Galimberti S.[2], Valsecchi M.G.[2]
[1]*Sanofi R&D, Oncology Biostatistics, Biostatistics and Programming Department ~ Montpellier ~ France,* [2]*School of Medicine and Surgery, University of Milano-Bicocca ~ Milano ~ Italy*

A surrogate endpoint offers an early evaluation of the true endpoint in clinical trials that require a long follow-up. The meta-analytic approach for the validation of a surrogate endpoint proposed by Burzykowski et al. [1], , which assesses surrogacy at both patient and trial levels, is now recognized as standard. In this context, no specific method exists to deal with a continuous variable characterized by a spike in zero as a surrogate endpoint for a survival outcome. The motivating clinical context is the evaluation of Minimal Residual Disease (MRD) as an early surrogate endpoint for event free survival (EFS) in childhood acute lymphoblastic leukaemia (ALL). Indeed, MRD, which quantifies the number of circulating leukemic cells, has not yet been formally validated as a surrogate endpoint, whilst it is a well-established prognostic biomarker in ALL. This marker has the peculiarity of having a highly skewed distribution characterized by a high proportion of zero values. The use of the copula model, one of the hallmark of the meta-analytic approach, is very useful in surrogate validation as allows the joint modelling of different types of endpoints to assess their associations. Here, we explored the appropriateness of a mixed discrete-continuous distribution, i.e. the Zero Adjusted Gamma (ZAGA), for the surrogate endpoint, while a Weibull distribution was used for the time-to-event endpoint. We also explored the use of rotated copulas to better capture the nature of the association between the two marginal survival distributions. The results of the extended simulation study indicated that the estimation of the individual level parameter is always unbiased and precise, while there is an underestimation of the trial level association when the distribution is heavily skewed.
The use of a ZAGA distribution might be useful in the validation of a highly skewed surrogate endpoint with a spike in zero. Rotated copulas should be also considered as an option in the process of copula specification.

[1] Burzykowski T., Molenberghs G., Buyse M., Renard D., and Geys H., Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints, 2001, Applied Statistics, 50, 405–422.

Parallel Sessions | Monday 28 August 2023

MO4.2 Metrics of spatial interaction between immune and cancer cells in tumor microenvironment as cancer biomarkers

Chervoneva I.*[1], Maisel B.[1], Yi M.[1], Peck A.[2], Rui H.[2]
[1]Thomas Jefferson University ~ Philadelphia ~ United States of America, [2]Medical College of Wisconsin ~ Milwaukee ~ United States of America

The studies of the tumor microenvironment (TME) are increasingly important for optimizing treatment strategies in oncology [1]. There is emerging evidence that spatial interactions between cancer and immune cells are essential predictors of disease progression and response to treatments. We aimed to develop prognostic biomarkers that quantify spatial interactions utilizing distributions of the nearest neighbor distances between cancer and immune cells. This work was motivated by studies of TME in a tissue microarray (TMA) of surgical specimens from a large cohort of breast cancer patients. Analysis of interaction in TMA data presents additional challenges since the size of a typical TMA tissue core is small and some cores may have very low or zero counts of immune cells of a particular type. Thus, some TMA cores may not yield a sufficient number of cells for the estimation of K-function and its variants previously developed for quantification of spatial interactions in the framework of the multitype marked point process. We propose novel metrics of spatial interactions based on (i) entire distributions of the nearest neighbor distances (NNDs) between cancer and immune cells and (ii) localized immune cell densities in the short juxtracrine distance and paracrine communication distance to cancer cells [2]. The proposed Functional Regression NND (FR-NND) biomarker is defined as the integral of the NND quantile function multiplied by the weight function represented by a penalized spline and estimated by fitting a linear functional regression model for a clinical outcome (progressionfree survival, PFS, for breast cancer data). That is, the weight function is optimized so that the FR-NND biomarker has the highest power to predict a relevant clinical outcome. The NNDbased and localized density metrics of spatial interaction were considered between cytotoxic CD8+ T-lymphocytes and cancer cells and between CD163+ tumor-associated macrophages (TAMs) and breast cancer cells. The NND-based metrics provided stronger predictors of PFS than usually considered counts or densities of CD8+ or CD163+ immune cells. The proposed NND-based and localized density metrics of spatial interaction between immune and cancer cells provide new promising cancer biomarkers suitable for evaluation in limited TMA tissue and in whole tissue data.

[1] Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., ... & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nature medicine, 24(5), 541-550.
[2] Maisel, B. A., Yi, M., Peck, A. R., Sun, Y., Hooke, J. A., Kovatich, A. J., ... & Chervoneva, I. (2022). Spatial metrics of interaction between CD163-positive macrophages and cancer cells and progression-free survival in chemo-treated breast cancer. Cancers, 14(2), 308.

MO4.3 Bayesian network meta-analysis of time-to-event data for evaluation of predictive biomarkers using ipd and ad

Umemneku--Chikere C.*[1], Owen R.[2], Wheaton L.[1], Poad H.[1], Cuevas Andrade I.[1], Ray D.[1], Khan S.[3], Tappenden P.[4], Abrams K.[5], Bujkiewicz S.[1]
[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester ~ Leicester ~ United Kingdom, [2]Swansea University Medical School, Swansea University ~ Swansea ~ United Kingdom, [3]Leicester Cancer+ Research Centre, Robert Kilpatrick Clinical Sciences Building, University of Leicester ~ Leicester ~ United Kingdom, [4]School of Health and Related Research, University of Sheffield ~ Sheffield ~ United Kingdom, [5]Department of Statistics, University of Warwick ~ Warwick ~ United Kingdom

Understanding the efficacy of cancer therapies among patients with specific biomarkers facilitates personalised cancer medicine resulting in improved patients' outcome. Randomised controlled trials (RCTs) are the gold standard for estimating treatments effects. However, subgroup analyses, in patient populations defined by a biomarker of interest, are often not reported. If the proportion of patients within the biomarker subgroup is known a network metaregression could be considered. Furthermore, evidence from individual participant data (IPD) can be combined with aggregate level data (AD) [1]. When access to IPD from RCTs is unavailable, IPD from other sources may be explored. We built on a work by Proctor et al [2]; thus, extending the existing model to allow for modelling of time-to-event data and estimation of treatment effects in both subgroups of biomarker positive and negative patients, using a mixture of IPD and AD.
We develop and applied a one-stage Bayesian hierarchical network meta-analysis (NMA) to estimate the mean treatment effects. We illustrated how IPD can be obtained either from electronic health records (EHRs) or by digitalizing Kaplan-Meier curves. The methodology is illustrated with two case studies; exploring the effect of the addition of chemotherapy (C) to taxane (X) in breast cancer with hormone receptor biomarker, and exploring the effects of therapies targeted on Vascular Endothelial Growth Factor (VEGF) and Epidermal Growth Receptor (EGFR) in colorectal cancer with Kristen Rat Sarcoma (KRAS) biomarker. The treatment effect for overall survival for CX versus X in the breast cancer example, was 1.00 (0.67,1.29) and 1.00 (0.65,1.76) in positive (HR+ve) and negative (HR-ve) subgroups respectively. In the colorectal cancer example, the treatment effect of VEGF+C versus EGFR+C was 0.93 (0.57,1.36) and 1.11 (0.82,1.54) in the KRAS wild-type and KRAS mutate biomarker subgroups respectively. Generally, the addition of IPD had added-value compared to using AD alone by reducing uncertainty in the estimated treatment effects. For example, there were 49% and 33% reduction in the uncertainty in the HR+ve and HR-ve subgroups respectively. We developed a model for estimating treatment effectiveness within biomarker subgroups which uses evidence from IPD and AD. Inclusion of IPD resulted in estimates with improved precision.

[1] Saramago P, Chuang L-H, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data 2014;14:1–11.
[2] Proctor T, Jensen K, Kieser M. Integrated evaluation of targeted and non-targeted therapies in a network meta-analysis 2020;62:777–89.

Parallel Sessions | Monday 28 August 2023

MO4.4 Agreement and error of titration assays

Alexander N.*, Schmidt W.
London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom

Quantification of assay performance is required by international standards such as GCP [1] and GCLP [2]. Titration assays can be used to define positivity either in terms of a change over time, i.e. seroconversion, or relative to a fixed threshold. The operating characteristics of these definitions depend on the precision of the assay. We present a deconvolution approach to estimating the distribution of errors, at the level of a single replicate, from the distribution of within-pair agreement. When the maximum replicate-level error is one dilution, a simple probability argument is used, with estimation by method of moments. The resulting expression for the standard error can, in conjunction with a prior estimate of precision, be used to estimate the required sample size. For the more general case, a discretized Gaussian model is used, with maximum likelihood estimation. These models fit well to eight published datasets. The discretized Gaussian model also allows the potential performance of alternative dilution factors to be assessed. For influenza hemagglutination-inhibition, the approach is compared to a previous Markov chain Monte Carlo data augmentation model. These methods allow the estimation of the underlying error distribution from observed betweenreplicate differences under repeatability conditions. The results can be used to guide the choice of the fold change necessary to infer seroconversion. Finer dilution factors, e.g. 1.5 rather than 2, could facilitate a better balance between the sensitivity and specificity of titration assays.

[1] ICH, 1996. Validation of analytical procedures: text and methodology. In: Report Q2 (R1). International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.
[2] Ezzelle J et al, 2008. Guidelines on good clinical laboratory practice: bridging operations between research and clinical research laboratories. J. Pharm. Biomed. Anal. 46, 18–29.

PARALLEL SESSION MO5: HIGH DIMENSIONAL DATA 1

MO5.1 Conditional variable screening for ultra-high dimensional longitudinal data with time interactions

Bratsberg A.*[1], Ghosh A.[2], Thoresen M.[1]
[1]University of Oslo ~ Oslo ~ Norway, [2]Indian Statistical Institute ~ Kolkata ~ India

In recent years we have been able to gather large amounts of genetic data at a fast rate, creating situations where the number of variables greatly exceeds the number of observations. In these situations, most models that can handle a moderately high dimension will now become computationally infeasible. Hence, there is a need for a pre-screening of variables to reduce the dimension efficiently and accurately to a more moderate scale. There has been much work to develop such screening procedures for independent outcomes. However, much less work has been done for high-dimensional longitudinal data, in which the observations can no longer be assumed to be independent. In addition, it is of interest to capture possible interactions between the genetic variable and time in many of these longitudinal studies. This calls for the development of new screening procedures for high-dimensional longitudinal data where the focus is on possible interactions with time. In this work, we propose a novel screening procedure that ranks variables according to the likelihood value at the maximum likelihood estimates in a semi-marginal linear mixed model, where the gene variable and its interaction with time is included in the model. This is to our knowledge the first conditional screening approach for clustered data. We prove that this approach enjoys the sure screening property, and assess the finite sample performance of the method, with a comparison of an already existing screening approach based on generalized estimating equations. We show that for capturing groups of interaction parameters, screening on the likelihood yields the best recovery rate of the true interactions, across several settings. Finally, we apply the proposed method to real data from a longitudinal study on measured serum triglyceride over the course of six hours, with measured mRNA on a targeted set of genes as our high-dimensional set of covariates. Emre Barut, Jianqing Fan, and Anneleen Verhasselt. Conditional sure independence screening. Peirong Xu, Lixing Zhu, and Yi Li. Ultrahigh dimensional time course feature selection. Patrik Hansson, Kirsten B Holven, Linn K L Øyri, Hilde K Brekke, Anne S Biong, Gyrd O Gjevestad, Ghulam S Raza, Karl-Heinz Herzig, Magne Thoresen, and Stine M Ulven. Meals with Similar Fat Content from Different Dairy Products Induce Different Postprandial Triglyceride Responses in Healthy Adults: A Randomized Controlled Cross-Over Trial.

MO5.2 Incorporating external information into bayesian additive regression trees using empirical bayes

Goedhart J.*, Klasusch T., Van De Wiel M.
Amsterdam University Medical Centers ~ Amsterdam ~ Netherlands

One of the promises of omics data is to improve the diagnosis of cancer and to find relevant biomarkers that may be used for therapy. However, omics data is typically high-dimensional, which, combined with the complicated interaction patterns between the measured omics covariates, poses significant challenges for prediction and feature selection. To improve prediction and feature selection, we propose to incorporate co-data, i.e. external information on the measured covariates, into Bayesian additive regression trees (BART) [1], a sum-of-trees prediction model that utilizes priors on the tree parameters to prevent overfitting. BARTs ability to model nonlinearities combined with co-data to guide the search for relevant variables may serve as an interesting tool for omics-based prediction models. To incorporate the co-data, we develop an Empirical Bayes (EB) framework that estimates, assisted by co-data, the hyperparameters which determine prior covariate weights in the BART model. Our proposed method can handle multiple types and sources of co-data, whereas most existing methods only allow co-data in the form of groups [2,3]. Furthermore, our proposed EB framework enables the estimation of the other hyperparameters of BART as well. Hyperparameters of BART are typically estimated using cross-validation. Empirical Bayes avoids using an arbitrary grid and may therefore render more refined hyperparameter estimates. We show that our method renders both improved predictions and variable selection compared to default BART in simulations. Moreover, it enhances prediction and variable selection stability in an application to diffuse large B-cell lymphoma diagnosis based on mutations, translocations, and DNA copy number data. Furthermore, our method is competitive to state-of-the-art co-data learners such as ecpc [4] and corf [5]. Our method is able to utilize co-data to improve the performance and variable selection of Bayesian additive regression trees.

[1] Chipman, H. A. et al. (2010). BART: Bayesian additive regression trees. Ann. Appl. Stat., 4, 266-298
[2] Du, J. & Linero, A. R. (2019). Incorporating Grouping Information into Bayesian Decision Tree Ensembles. PMLR, 97, 1686-1695
[3] Velten, B. & Huber, W. (2019). Adaptive penalization in high-dimensional regression and classification with external covariates using variational Bayes Biostatistics, 22, 348-364
[4] van Nee, M. M. et al. (2021). Flexible co-data learning for high-dimensional prediction. Stat Med, 40, 5910-5925
[5] te Beest, D. E. et al. (2017). Improved high-dimensional prediction with Random Forests by the use of co-data. BMC Bioinform, 18, 584 MO5. 2_24 9

MO5.3 A framework for interpretation and testing of sparse canonical correlations

Senar N.*, Van De Wiel M., Zwinderman A., Hof M.
Amsterdam UMC ~ Amsterdam ~ Netherlands

In clinical and biomedical research, it has become common to collect high-throughput omics data. This type of data often contains hundreds or thousands of variables per individual, while the number of patients are comparatively low, leaving the number of variables to exceed the number of observations. Data integration approaches linking datasets contribute to improvements in diagnostics and understanding of biological mechanisms. Extracting the signals connecting these datasets requires penalised multivariate methods. Canonical Correlation Analysis (CCA) is a multivariate method which has regained popularity for exploration of integrated genomics data. Sparse alternatives use penalties for dimension reduction in high-dimensional settings. In such cases, it is difficult to estimate the optimal penalty as there likely exist many competitive options. Furthermore, both the estimation of the penalty and the canonical weights are computationally burdensome. Additionally, the penalty may impact the later assessment of the estimated correlations. To deal with these challenges, we present a sparse CCA method based on the NIPALS algorithm. We impose sparsity using soft-thresholding on the canonical vectors, where sparsity is defined as the number of non-zero weights. Rather than estimating a penalty parameter, this approach grants control over the number of non-zero weights, simplifying the choice of penalty, easing interpretation and keeping computational times of permutation tests or bootstrapping procedures low. This structure allows for multiple sparsity levels to be computed simultaneously facilitating the search of penalty. With simulations, we showed that our new method outperforms existing sparse CCA approaches in terms of stability of variable selection and accuracy of the capture signal as well as maintaining orthogonality between components. Moreover, we applied our method on drug sensitivity and gene expression from the Genomics of Drug Sensitivity in Cancer database to study and compare the performance of our method on real data. The correlations were then evaluated via permutation testing. This approach leads to efficient estimation and reliable assessment of the canonical correlations through permutation testing. Compared to the currently existing methods, our method captured new information for different components better. Additionally, choosing the number of non-zeros helped reduce dimension so as to yield results which are easily interpreted.

Parallel Sessions | Monday 28 August 2023

MO5.4 Two-dimensional fused targeted ridge estimation for health indicator prediction

Van Wieringen W.*, Lettink A., Chinapaw M.
Amsterdam UMC ~ Amsterdam ~ Netherlands

Modern epidemiological and clinical studies gather, on top of the classical covariates, information generated by high-throughput techniques. The resulting data are then highdimensional, i.e. having more covariates than observations. To accommodate the highdimensionality of the data, regression models are fitted in regularized fashion by means of ridge and lasso estimation techniques. Here we present two-dimensional fused targeted ridge regularization to incorporate i) quantitative prior knowledge regarding the parameter and ii) structural relationships among the elements of the parameter. We derive estimators for the linear and logistic regression models. The estimators are accompanied by a computational efficient cross-validation procedure to choose the penalty parameters. The latter's search is further sped up by imposing a lower bound on penalty parameters. The bound prevents overfitting by limiting the degrees of freedom the estimator may consume. We illustrate the proposed procedure on the large epidemiological NHANES study, that annually collects health indicators of US citizens. We concentrate on the two years that involved over two thousand samples each. Additionally, to the `usual' personal characteristics and health parameters, 24x& accelerometer data are available. The latter data comprise, after preprocessing, the proportion of visits to the each entry of a large rectangular grid of bout length vs. acceleration level. Visits of neighboring grid points are expected to have a similar effect on health. From the data we aim to predict, e.g., Body Mass Index, where we ask ourselves whether the accelerometer information aids to this cause. Hereto we employ the developed twodimension fused targeted ridge estimator to encourages neighboring grid points to be associated with comparable parameter values. While the employed estimator yields a slight increase in predictive performance over standard competitors, the largest gain is in the interpretation of the estimate facilitated by the fusion over the two-dimensional grid. Lettink A, Chinapaw M, van Wieringen WN] (2023). "Two-dimensional fused targeted ridge regression for health indicator prediction from accelerometer data". Under revision.

MO5.5 Methodology for, and insight from, analysing displacement by tremor in parkinson's disease

Baker K.*[1], Gilmour S.[1], Umamahesan C.[1], Taylor D.[1], Charlett A.[2], Dobbs S.[1], Dobbs J.[1], WellerC.[1]
[1]King's College London ~ London ~ United Kingdom, [2]UK Health Security Agency ~ London ~ United Kingdom

Tremor is a cardinal sign of Parkinson's Disease (PD) and is usually measured using accelerometers, with inference being made with the accelerometer data. The conversion of acceleration to displacement via integration is theoretically simple, but non-trivial in practice due to various sources of noise and accumulation of errors in the integration process (Thong et. al. 2004). We aim to explore this conversion to displacement and extend existing methodology, or develop new methodology, to improve the accuracy of this process. We intend to use these displacement signals to develop key metrics of tremor, which can then be used in further modelling to improve understanding of the disease. A review of numerical integration methods for movement data was conducted, and identified methods tested using mechanically simulated data. A linear mixed model is used to understand the error structure of the different methods, and consequently select the best-performing technique for numerical integration. A novel methodology has been developed which aims to improve accuracy by identifying sections of the signal that are noise-like, and removing those prior to integration, minimising the amount of noise in the integration process. This methodology utilises a functional outlier detection windowing method with p-value correction for multiple testing. Preliminary data collected from 59 participants with PD, 38 of their spouses and 36 controls indicates that, whilst displacement by tremor was discriminatory of diagnosed-PD, it may also have potential for quantifying distance down-the-pathway to PD in those without diagnosed PD. Our methodology, also applicable to other types of sensor data, demonstrates how we can improve the numerical integration process to achieve more accurate displacement signals via noise detection and omission. The translation of accelerometer data to displacement and calculation of clinically defined metrics is a tool allowing better understanding of tremor associated with PD.
Y.K. Thong, M.S. Woolfson, J.A. Crowe, B.R. Hayes-Gill, D.A. Jones, Numerical double integration of acceleration measurements in noise, Measurement, Volume 36, Issue 1, 2004,

PARALLEL SESSION MO6: EPIDEMIOLOGY 1

MO6.1 Linked shrinkage to improve the estimation of interaction effects in a regression model

Van De Wiel M.*
Amsterdam university medical centers ~ Amsterdam ~ Netherlands

Adding interactions to a regression model is a classical problem in statistics. This may lead to interesting insights on the joint effects of covariates and may render predictive performance competitive to that of less well-interpretable machine learners. It comes at a price though, as the number of regression coefficients increases quadratically when adding two-way interactions. Our aim is to develop an estimator that adapts well to this increased dimension, thereby providing accurate estimates and appropriate inference.
We shortly review some existing selection strategies to overcome the dimensionality problem, such as the hierarchical lasso [1] and the two-step approach, which considers only interactions between significant main effects.
Our suggested approach links shrinkage of the interactions to that of the main effects using a hierarchical model.
It is a variation on the nowadays popular local shrinkage model, which stabilizes the shrinkage parameters from a ridge model by endowing them with a half-Cauchy prior [2]. We empirically show that borrowing strength between the amount of shrinkage for main effects and their interactions renders an adaptive approach that may strongly improve estimation of the regression coefficients. In addition, we evaluate the potential of the model for appropriate inference of those coefficients, which is notoriously hard for selection strategies.
The potential benefits are demonstrated on realistic dimensions of clinical or epidemiological studies, such as n=1,000 and number of covariates p = 15, using large-scale cohort data. Comparisons with the aforementioned methods and other shrinkage strategies are provided for a fairly wide spectrum of settings. The method is practical in use, as computation time is usually limited to several minutes, and implementation is fairly straightforward. Finally, extensions to interpretable machine learning and potential caveats of the approach are also discussed.
The new method is a good alternative for existing strategies to handle interactions in epidemiological and/or clinical studies, as it is accurate, provides appropriate inference and is of practical use.

[1] Lim M, and T Hastie (2015). Learning interactions via hierarchical group-lasso regularization. J Comp Graph Stat, 24 (3), 627-654.
[2] Gelman A. (2006). Prior distributions for variance parameters in hierarchical models. Bayesian analysis, 1, 515-534.

Parallel Sessions | Monday 28 August 2023

MO6.2 Novel insights for quantifying selection bias using interactions on the log additive scale

Gkatzionis A.*[1], Seaman S.[2], Hughes R.[1], Tilling K.[1]
[1]*MRC Integrative Epidemiology Unit, University of Bristol ~ Bristol ~ United Kingdom,* [2]*MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom*

Selection bias is a common concern in genetic and epidemiologic studies. If selection into a study/analysis is represented by a binary indicator S, then selection bias arises if S is a collider (common effect) of the analysis model's exposure X and outcome Y. When detecting and adjusting for selection bias, it is common to use logistic regression to identify variables associated with selection and conclude that associations between these variables in the selected sample will be biased. However, this provides little information about the magnitude of bias, and many examples exist where the exposure and outcome are strongly associated with selection but the bias in the estimated exposure-outcome association is small.
We propose an alternative approach based on a log-additive model for selection and show how this can quantify the magnitude of selection bias in a range of simple analysis models. For an analysis model using logistic regression with binary Y and X, it is known that the magnitude of selection bias in the exposure-outcome regression coefficient is proportional to the strength of interaction d3 between X and Y in a log-additive model for selection: $P(S = 1 \mid X, Y) = \exp\{d0 + d1\ X + d2\ Y + d3\ X\ Y\}$. We prove that a similar result holds under a linear regression model for a continuous outcome, or under a Poisson regression model for a count outcome, regardless of the type of the exposure variable. We also show by simulation that even if a log-additive model is not the true model for S, the interaction term in such a model is still informative about the magnitude of selection bias.
The magnitude of selection bias induced an applied analysis depends on interactions between analysis variables in their effects on selection into the study on the log-additive scale. The relationship is linear for a range of simple statistical models, including linear, logistic and Poisson regression. Our results can prove useful to applied researchers conducting sensitivity analyses or implementing inverse probability weighting to adjust for selection bias. Paper currently in writing, reference will be available by the time of the conference. Relevant literature: Jiang, Z. and P. Ding (2017). The directions of selection bias. Statistics & Probability Letters 125, 104-109. Bartlett, J. W., O. Harel, and J. R. Carpenter (2015). Asymptotically unbiased estimation of exposure odds ratios in complete records logistic regression. American Journal of Epidemiology 182 (8), 730-736.

MO6.3 Investigating accuracy of disease outcome definition in pharmacoepidemiology: bayesian latent class model

Uno S.*[1], Tango T.[2]
[1]*Astellas Pharma Inc. / Center of Medical Statistics / The Graduate University for Advanced Studies (SOKENDAI) ~ Tokyo ~ Japan,* [2]*Center of Medical Statistics ~ Tokyo ~ Japan*

Large electronic database studies have been widely used in recent years but can be affected by biases due to a lack of information. To address these issues, validation studies have been conducted to evaluate the accuracy of disease diagnoses defined from databases. However, these validation studies can be limited by the possibility of misclassification in the reference data and the dependence between diagnoses from the same data source.
The latent class model is a statistical method that can adjust these biases caused by imperfect diagnostic tests and dependence. This manuscript applies the fixed-effect Bayesian latent class models regarding the presence/absence of an assumption of the gold standard (GS) for registry diagnosis and conditional independence. The four models are distinguished by whether they assume GS and conditional independence, only one or the other, or neither. Along with the posterior distribution of each parameter, we calculate the widely applicable information criterion (WAIC) for model comparison. In addition, we performed a simulation study to demonstrate the accuracy of the four models proposed in the previous section when applied to simulated datasets associated with each data-generating scenario. The model that assumed conditional dependence and non-GS reference showed the best predictive performance among the four models in terms of WAIC. It also showed discrepant results from previous findings because the disease prevalence rate was slightly higher and the sensitivities were estimated to be significantly lower than those of the other models. Finally, the model consistently provided good results in a simulation, with the lowest Pearson-type statistics among the four models and a WAIC, irrespective of data-generating scenarios.
We found that current assessments of outcome validation can introduce bias into DB diagnoses based on data from registries. The proposed approach should gain prominence and be widely used in validation studies.
Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. Mar 2001;57(1):158-67. Sato I, Yagata H, Ohashi Y. The Accuracy of Japanese Claims Data in Identifying Breast Cancer Cases. Biological and Pharmaceutical Bulletin. 2015;38(1):53-57.

MO6.4 Increase efficiency and reduce bias when assessing hpv vaccination efficacy by using non-targeted hpv strains

Etievant L.*, Sampson J., Gail M.
Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute ~ Rockville, Maryland ~ United States of America

Studies of human papilloma virus (HPV) vaccine efficacy often obtain data on vaccine-targeted virus strains 16/18 and on non-targeted strains. However, standard analyses ignore the non- targeted strains. Assuming non-targeted strains are unaffected by HPV vaccination, and noting they can be regarded as "negative control outcomes"[1], our objective is to show how using data on the non-targeted HPV strains can (i) increase precision of the estimated vaccine effect on targeted strains in randomized trials, and (ii) reduce bias due to unmeasured confounders in observational studies.
For objective (i), we adapt an existing method [2] and use non-targeted strains as if they were baseline covariates. More specifically, we augment estimating equations by adding a function of the incidence of non-targeted HPV infections. For objective (ii), we estimate the vaccine effect on non-targeted strains and use this non-null effect to remove confounding bias in the estimated vaccine effect on the targeted strains. We present assumptions ensuring that the estimated effect on the non-targeted strains can be used to markedly reduce confounding bias from the effect on the targeted strains. A key assumption is that unmeasured confounders affect the risk of targeted and non-targeted strains proportionally, as is plausible for unmeasured sexual activity for example. We evaluate the utility of these approaches in simulations based on HPV data and in studies of HPV vaccines that only target a few highly carcinogenic strains. We find modest increases in precision in HPV vaccine trials, but our use of non-targeted HPV strains in observational studies substantially reduces confounding bias.
Using non-targeted strains known not to be affected by the HPV vaccine can improve inference on targeted strains, especially in observational studies. We believe our paper is the first to suggest a valid method for using post-randomization events to gain precision in randomized trials. In observational studies, our use of non-targeted HPV strains substantially reduces confounding bias, and such adjustments can render estimates from an observational study more useful, alone or in combination with other observational studies.

[1] Shi X., Miao W. and Tchetgen Tchetgen E., Current Epidemiology Reports, 7, 2020, 190-202.
[2] Zhang, M., Tsiatis, A., Davidian, M., Biometrics, 64, 2008, 707–715.

Parallel Sessions | Monday 28 August 2023

MO6.5 Quantifying the effect of mobility restrictions on health policies during the pandemic: a fda approach

Ieva F.*, Mazzola V., Bonaccorsi G., Secchi P.
Politecnico di Milano ~ Milano ~ Italy

In response to the COVID-19 pandemic, Italy adopted restrictive mobility measures to contain the spread of the virus. Nevertheless, a systematic/quantitative assessment of the effectiveness of such measurements over the pandemic spread and related mortality is still missing. This work was motivated by the interest in analysing mobility data and in providing decision makers with actionable tools for assessing mobility restriction policies. We aimed at developing novel strategies for assessing the relationship between mobility and overall mortality, leveraging the use of functional data analysis.
To characterize the effect of mobility restrictions, we analysed a large-scale collection of near real-time data provided by the Facebook platform [1]. We first identified clusters of areas with similar mobility density behaviours [2], applying a functional K-means informed by a Wasserstein distance. We then spotted areas with peculiar mortality density behaviour with respect to the surroundings with LISA maps, exploiting the Local Moran's I statistics whose weights account for spatial correlations. Then the Spearman correlation of mortality and mobility data is computed to assess the lag of significant association between the two phenomena. We found negative correlations between mortality and mobility densities, consistent with the fact that when high mortality concentrations were detected, the Italian Government adopted more restrictive measures and therefore allowed less in terms of mobility, and vice versa. This is particularly evident for the provinces most affected by the epidemic, especially concerning the first wave, and in cases where also the spatial component during clustering was considered. Moreover, the time lags between mobility reduction and actual results in limiting and controlling the epidemic were found to be longer in the provinces that suffered most from the COVID-19 pandemic in the period under consideration.
We exploited Facebook data to measure mobility networks in the country, with province granularity, over the weeks of the two years of pandemic. We defined two quantitative criteria (based on Spearman's correlation) to determine when mobility restriction measures can be considered effective in containing the COVID-19 spread in terms of overall mortality.

[1] Meta. Data for Good program. Accessed March 2022 https://dataforgood.facebook.com/dfg/tools/movement-maps
[2]Scimone, R., Menafoglio, A., Sangalli, L.M., Secchi, P.: A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities. Spatial Statistics. (2021) doi: 10.1016/j.spasta.2021.1005

PARALLEL SESSION MO7: CLINICAL TRIALS 2

MO7.1 Estimating the treatment effect in a clinical trial

Van Zwet E.*
Leiden University Medical Center ~ Leiden ~ Netherlands

It is nearly three quarters of a century since what is generally regarded as the first modern randomized clinical trial, the UK Medical Research Council study of the effectiveness of streptomycin in tuberculosis. Since then, tens of thousands of randomized controlled trials (RCTs) have been conducted. Our goal is to study this wealth of information, and to make use of it. We have obtained the primary efficacy results of more than 20,000 randomized controlled trials (RCTs) from the Cochrane Database of Systematic Reviews (CDSR). We summarize the result of each trial as a triple (beta,b,s) where beta is the "true" effect of the treatment, b is the unbiased, normally distributed estimator, and s is the standard error of b. We do not observe beta, but we do observe the pair (b,s). We define the z-value z=b/s and the signal-to-noise ratio SNR=beta/s. Despite the fact that the true effects (beta) are not observed, it is possible estimate the joint distribution of the z-values and the SNRs across the CDSR. This provides much insight into the statistical properties of RCTs. In particular, we can compute the expected degree of overestimation when a trial reaches statistical significance. This is sometimes called the winner's curse or the Type M error. We propose a shrinkage estimator to address this problem. We evaluate and compare the performance of the usual (unbiased) estimator and the novel shrinkage estimator on average across the CDSR. We find that the shrinkage estimator not only counters the winner's curse, but also reduces the mean squared error by nearly 50% compared to the unbiased estimator.
This talk is based on joint work with Simon Schwab, Stephen Senn, Lu Tian and Robert Tibshirani. van Zwet, E., Schwab, S. and Senn, S., 2021. The statistical properties of RCTs and a proposal for shrinkage. Statistics in Medicine, 40(27), pp.6107-6117.

MO7.2 Robust incorporation of external information in hypothesis testing

Calderazzo S.*, Wiesenfarth M., Weru V., Kopp--Schneider A.
German Cancer Research Center ~ Heidelberg ~ Germany

When designing a clinical trial, external information on the treatment and/or control arm is often available. The Bayesian approach allows borrowing of such external information through the adoption of informative prior distributions. It is well known that borrowing can improve the trial's test error rates if external information is consistent with the current trial's data-generating process, while losses can be severe otherwise. Several robust approaches have been proposed to limit the impact of potentially heterogenous external information. However, it has been shown that no power gains are possible if strict control of type I error rate is desired [1]. Moreover, such approaches require the choice of tuning parameters and/or distributions which are often not intuitively related to their induced frequentist operating characteristics. We propose a method which explicitly aims at achieving a compromise between full and no borrowing in terms of type I error rate, while enforcing an upper bound on its maximum tolerated inflation. The compromise is achieved by tuning test decision thresholds, and is applicable to both one and two-arm trials. Simulations are performed to show the properties of the approach under various prior-data conflict and prior informativeness configurations. The proposed method provides a rationale for type I error rate inflation when robust incorporation of external information is desired. In one-arm studies, such an inflation can be analytically related to the proposed method's borrowing weight, by exploiting the known duality between prior probabilities and test error costs [2]. In two-arm studies, such a correspondence is only approximate due to the data-dependence of the required test decision threshold. However, an explicit upper bound on type I error rate can still be enforced. Such an upper bound may be of advantage in a regulatory setting and may improve the transparency in communicating the trial design.

[1] A. Kopp-Schneider, S. Calderazzo, M. Wiesenfarth, Biometrical Journal, 62(2):361–374.
[2] J. O. Berger, Statistical decision theory and Bayesian analysis; 2nd ed. Springer Series in Statistics. Springer, New York, 1985.

MO7.3 Testing for similarity of multivariate mixed outcomes with application to efficacy-toxicity responses

Hagemann N.*[1], Marra G.[2], Bretz F.[3], Möllenhoff K.[1]
[1]Mathematical Institute, Heinrich-Heine-University Düsseldorf ~ Düsseldorf ~ Germany, [2]Department of Statistical Science, University College London ~ London ~ United Kingdom, [3]Statistical Methodology, Novartis Pharma AG ~ Basel ~ Switzerland

A common problem in clinical trials is to test whether an effect of an explanatory variable on the response, e.g. the effect of the dose of a compound on efficacy, is similar between two groups. In this context, similarity is defined as equivalence up to a pre-specified threshold specifying the accepted deviation between the groups. Such question is usually assessed by testing whether the (marginal) effects of the explanatory variable on the response are similar, based on, for example, confidence intervals for differences, or, to mention another example, the distance between two parametric models. These approaches typically assume a univariate continuous or binary outcome variable. An approach for associated bivariate binary response variables, based on the Gumbel model, has been recently introduced [1]. In this talk, we propose a flexible extension of such methodology that builds on a generalized joint regression framework with Gaussian copula. Compared to existing approaches, this allows for various scales of the outcome variables (e.g. continuous, binary, categorical, ordinal) including mixed outcomes as well as responses with more than two dimensions. We demonstrate the validity of our approach by means of a simulation study. An efficacy-toxicity case study demonstrates the practical relevance of the approach. Our suggested approach is widely applicable and, therefore, overcomes the requirement of the responses being bivariate and binary. Simulations show its validity as well as its comparably high power.

[1] Möllenhoff, K., Dette, H., and Bretz, F. (2021). Testing for similarity of binary efficacy-toxicity responses. Biostatistics, 23(3), 949–966.

Parallel Sessions | Monday 28 August 2023

MO7.4 Comparing statistical models to estimate causal treatment effects in aggregated n-of-1 trials

Gärtner T.*, Schneider J., Arnrich B., Konigorski S.
Digital Health Center, Hasso Plattner Institute for Digital Engineering, University of Potsdam ~ Potsdam ~ Germany

The aggregation of a series of N-of-1 trials presents an innovative and efficient study design, as an alternative to traditional randomized clinical trials. N-of-1 trials are multi-crossover controlled trials, where each patient is their own control group. Challenges for the statistical analysis arise when there is carry-over or complex dependencies of the treatment effect of interest. In this study, we evaluate and compare methods for the analysis of aggregated N-of-1 trials in different scenarios with carry-over and complex dependencies of treatment effects on covariates. For this, we simulate data of a series of N-of-1 trials for Chronic Nonspecific Low Back Pain based on assumed causal relationships parameterized by directed acyclic graphs. We simulate 1000 patients and test the model performances on different subsamples with sizes between 5 and 100 patients. In addition to existing statistical methods such as regression models, Bayesian Networks, and G-estimation, we introduce a carry-over adjusted parametric model (COAPM). The results show that all evaluated existing models were able to identify the treatment effect when there is no carry-over and no treatment interaction with a covariate. When there is carryover, COAPM yields unbiased treatment effect estimates and narrower confident intervals while all other methods underestimate the simulated treatment effect. When there is known treatment dependence, all approaches that are capable to model it yield unbiased estimates. Finally, the performance of all methods decreases slightly when there are missing values, and the bias in the effect estimates can also increase. This study presents a systematic evaluation of existing and novel approaches for the statistical analysis of a series of N-of-1 trials. We derive practical recommendations which methods may be best in which scenarios Comparison of Bayesian networks, G-estimation and linear models to estimate causal treatment effects in aggregated N-of-1 trials Thomas Gärtner, Juliana Schneider, Bert Arnrich, Stefan Konigorski medRxiv 2022.07.21.22277832; doi: https://doi.org/10.1101/2022.07.21.22277832

MO7.5 Multimodal outcomes in n-of-1 trials: combining deep learning and statistical inference

Schneider J.*, Gärtner T., Konigorski S.
Digital Health Center, Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany ~ Potsdam ~ Germany

N-of-1 trials are randomized multi-crossover trials in single participants with the purpose of investigating the possible effects of one or more treatments. These effects can be modeled individually or on an aggregate level to estimate population effects. In N-of-1 trials, a participant alternates (possibly randomly) between periods of treatment and non- or alternative treatment. This trial design is especially useful for rare diseases, chronic diseases and personalized analyses. Research in the field of N-of-1 trials has primarily focused on scalar outcomes. However, with the increasing use of digital health applications, we propose to adapt this design to multimodal outcomes that later could easily be collected by the trial participants on their personal mobile devices. Multimodal N-of-1 trials are a variant of N-of-1 trials that use multimedia such as audio, video or image data instead of metric scales to measure the outcome. In a first multimodal N-of- 1 trial, Fu et al.[1] recently investigated the effect of creams on acne severity by tracking the outcome with images of the affected areas and applying supervised deep learning models in a series of 5 individual N-of-1 trials. We present an approach for analyzing multimodal N-of-1 trials by combining unsupervised deep learning models with statistical inference. First, we trained a Variational Autoencoder (VAE) on the skin images to create a lower dimensional embedding for representation. In a dimensionality reduction step, the embeddings were then reduced to a single dimension by extracting the first principle component. Finally, a univariate t-test was applied on this component in order to determine whether treatment and non-treatment periods showed mean differences. All these steps were performed for each individual separately, and we compared the results to an expert analysis that would rate the pictures directly with respect to their acne severity and apply a t-test on those scores. The results indicated a treatment effects for one individual in the expert analysis. This effect was replicated with the proposed deep learning approach. In future analyses with further experiments on hyperparameter tuning for the VAE, different methods for dimensionality reduction, and statistical tests adjusted for confounding and time, we expect to generate better representations of the images and improve as well as the statistical efficiency of the tests.
Jingjing Fu, Shuheng Liu, Siqi Du, Siqiao Ruan, Xuliang Guo, Weiwei Pan, Abhishek Sharma, and Stefan Konigorski. Multimodal n-of-1 trials: A novel personalized healthcare design, 2023.

PARALLEL SESSION MO8: SURVIVAL ANALYSIS 2

MO8.1 Gradient boosting for survival analysis with competing risks

Ahmed Nofel (WINNER OF ISCB44 CONFERENCE AWARD FOR SCIENTISTS), Khan M.H.R.
Institute of Statistical Research and Training, University of Dhaka ~ Dhaka ~ Bangladesh

For analyzing high-dimensional time-to-event data with competing risks, sophisticated methods are required that consider censoring and the event of interest. The sparsity associated with high-dimensional data introduces even more complexity. For low-dimensional data, the subdistribution hazard model or cause-specific hazard model has been proposed. Gradient boosting decision tree (GBDT) is an ensemble machine learning method that has the potential to identify risk factors associated with micro-array data due to its ability to handle high dimensional data. The efficiency and scalability of GBDT training have recently been greatly improved by Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting (LightGBM) algorithms. In this study, we proposed a method to analyze high-dimensional competing risk data with XGBoost and LightGBM ensemble machine learning techniques under the Fine-Gray subdistribution hazard model. The proposed methods implemented with several simulation studies under a variety of settings–low and high-dimensional data with correlated and uncorrelated covariates, medium and high censoring rates, and two types of competing risks and with a real data set called non-muscle invasive bladder cancer data to demonstrate the performance of the proposed approach and compared them with the existing boosting technique known as CoxBoost. The models were assessed based on their discriminative ability through Harrel's Cindex and Uno's C-index. We found that under a high-dimensional setting, the LightGBM and XGBoost were significantly efficient compared to the CoxBoost but for a low-dimensional setting, they performed equivalently. Real data analysis also suggested that LightGBM outperformed CoxBoost and XGBoost. We found that without any parameter tuning gradient boosting methods performed equivalently with CoxBoost. With the addition of parameter tuning, even better results were achieved. Modern gradient boosting techniques are much more useful and computationally viable tools for analyzing high-dimensional data with competing risks. LightGBM can be successfully implemented in analysis of biomarker data.

MO8.2 Imputing missing covariates for competing risks analyses when using the fine–gray model

Bonneville E.F.*[1], Beyersmann J.[2], Keogh R.H.[3], Bartlett J.W.[3], Morris T.P.[4], De Wreede L.C.[1], Putter H.[1]
[1]Department of Biomedical Data Sciences, Leiden University Medical Center ~ Leiden ~ Netherlands, [2]Institute of Statistics, Ulm University ~ Ulm ~ Germany, [3]Department of Medical Statistics, London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [4]MRC Clinical Trials Unit at UCL ~ London ~ United Kingdom

The Fine–Gray (FG) model for the subdistribution hazard is often used for the development of prognostic models in competing risks settings. When there are missing values in the covariates included in a given model (at model development), researchers may wish to multiply impute them. While previous literature has addressed the use of multiple imputation (MI) in the context of cause-specific (CS) Cox proportional hazards models [1], no such guidance exists for the FG model. In particular, whether a substantive-model-compatible approach (known as SMC-FCS, see [2]) can be extended to the FG context is an open question. Assuming interest lies in estimating the risk of only one of the competing events (henceforth referred to as 'cause 1'), we developed a MI approach that exploits the parallels between the FG model and the standard (single-event) Cox model. Due to the form of the subdistribution risk set, in the presence of random right censoring, standard software for fitting Cox models can be used to fit a FG model if, for those failing from competing events, we either a) use timedependent weights, or b) multiply impute their potential censoring times. Our approach entails multiply imputing the potential censoring times in a first step, and thereafter imputing the missing covariates analogously to the single-event setting in a second step. When the second step is done using existing SMC-FCS software, it ensures compatibility between the imputation model and the FG model. We used a simulation study to assess our approach and compare it with alternatives, including imputing compatibly with a CS model. This demonstrated that our approach is preferred when the FG model for cause 1 is correctly specified. The proposed two-step approach provides a way to impute missing covariates compatibly with a FG model, making efficient use of existing software. This work also contributes to the broader discussion on applied analysis choices in the presence of competing risks.

[1] Bartlett, J. W., & Taylor, J. M. (2016). Missing covariates in competing risks analysis. Biostatistics, 17(4), 751-763.
[2] Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., & Alzheimer's Disease Neuroimaging Initiative. (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. Statistical methods in medical research, 24(4), 462-487.*

Parallel Sessions | Monday 28 August 2023

MO8.3 Similarity of competing risks models with constant intensities in an application to healthcare pathways

Binder N.[4], Möllenhoff K.*[1], Sigle A.[3], Dette H.[2]
[1]*Mathematical Institute, Heinrich Heine University ~ Düsseldorf ~ Germany,* [2]*Department of Mathematics, Ruhr University Bochum ~ Bochum ~ Germany,* [3]*Department of Urology, Faculty of Medicine, Medical Center, University of Freiburg ~ Freiburg ~ Germany,* [4]*Institute for General Practice/Primary Care, Medical Center and Faculty of Medicine, University of Freiburg ~ Freiburg ~ Germany*

Similarity of different healthcare pathways is a critical issue when compiling treatment guidance for physicians for improving the standard of care. We understand the pathway as a temporal process with possible transitions from a single initial treatment state to hospital readmission of different types, which constitutes a competing risk setting. The development of the methodology presented in this talk is motivated by an application to hospital readmission for competing reasons after prostate cancer surgery. As the readmission intensities are low, the question is whether they are sufficiently similar for patients with prior in-house diagnostics versus without prior in-house diagnostics such that the two populations can be combined for further outcome analyses. To assess this question, in this talk, we propose a multi-state model-based approach to uncover pathway similarity between two groups of individuals. Precisely, we derive a hypothesis test based on a constrained parametric bootstrap for assessing the similarity of two competing risk models assuming constant transition intensities. Our suggested approach allows to identify thresholds for which the transition intensities are to be considered similar, meaning that patient groups can be pooled for deriving guidance. With respect to the application example, we were able to identify thresholds for which the global null hypothesis could be rejected and therefore the transition intensities are to be considered similar. Binder, N., Möllenhoff, K., Sigle, A., & Dette, H. (2022). Similarity of competing risks models with constant intensities in an application to clinical healthcare pathways involving prostate cancer surgery. Statistics in medicine, 41(19), 3804–3819. https://doi.org/10.1002/sim.9481

MO8.4 Developing & validating a competing risk joint model to characterise the prognosis of prostate cancer patients

Parr H.*[1], Porta N.[1], Tree A.[2], Dearnaley D.[1], Hall E.[1]
[1]*The Institute of Cancer Research ~ London ~ United Kingdom,* [2]*The Royal Marsden NHS Foundation Trust ~ London ~ United Kingdom*

Prostate cancer is a global health concern, with nearly 1.5 million men diagnosed worldwide, of which over half present with localised disease. Treatment options include radiotherapy with endocrine therapy, to maximise curative intent for intermediate- or high-risk patients. Over time, event-free survival rates have improved: in CHHiP [1] (N=3,216), a phase-III RCT, 10-year recurrence-free survival ~76%. As progression-free rates extend, long-term follow-up and horizon times can affect recurrence predictions, particularly in an ageing population. In the localised prostate cancer setting, deaths unrelated to the disease (when the patient is disease-free) can occur and preclude the observation of the event of interest: recurrence. Previously, we treated these as censored observations in the development of a prediction joint model (JM) for recurrence using prostate-specific antigen (PSA) measurements over time to dynamically update predictions of recurrence [2]. Competing risks (CR) methodology accounts for the occurrence of competing events that may hinder the observation of the event of interest, it provides unbiased estimates of the cumulative probability of recurrence over time. We develop and validate a CRJM to accurately predict the recurrence of cancer in the presence of the competing event of unrelated deaths. The predictive performance of CRJMs is compared to standard JMs using metrics of discrimination (AUC), and overall predictive performance, applying two frameworks: inverse probability censoring weighting (IPCW) and model-based approaches. Predictive performance is improved after several years of accrued PSA. The CRJM supersedes the standard JM AUCs in the first three years, with the IPCW CR performing best across all landmarks, with a maximum AUC of 0.89 at landmark 6 years. Despite the additional model complexity and framework used, CRJMs may give more accurate discrimination and improved predictive performance. The framework to extend JMs when considering the competing event is feasible and allows dynamic predictions to be extracted. CRJMs provide a more accurate approach to predicting recurrence, particularly in an ageing population where CRs become an increasing concern. Clinicians and patients will benefit from correctly accounting for CRs to improve their predictions of recurrence, and the framework presented here provides a valuable tool to do so. D. Dearnaley, I. Syndikus, H. Mossop, V. Khoo, A. Birtle, D. Bloomfield, J. Graham, P. Kirkbride, J. Logue, Z. Malik, J. Money-Kyrle, J.M. O'Sullivan, M. Panades, C. Parker, H. Patterson, C. Scrase, J. Staffurth, A. Stockdale, J. Tremlett, M. Bidmead, H. Mayles, O. Naismith, C. South, A. Gao, C. Cruickshank, S. Hassan, J. Pugh, C. Griffin, E. Hall, Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial, The Lancet Oncology. 17 (2016) 1047–1060. https://doi.org/10.1016/S1470-2045(16)30102-4.

[2] H. Parr, N. Porta, A.C. Tree, D. Dearnaley, E. Hall, A Personalised Clinical Dynamic Prediction Model to Characterise Prognosis for Patients with Localised Prostate Cancer. analysis of the CHHiP Phase III Trial, International Journal of Radiation Oncology, Biology, Physics. (2023). https://doi.org/10.1016/j.ijrobp.2023.02.022.

Parallel Sessions | Monday 28 August 2023

MO8.5 A novel case-cohort analytical framework for semi-competing risks within the frequentist paradigm

Zhou A.*[1], Lee K.H.[1], Stephenson B.[1], Lee S.[2], Haneuse S.[1]
[1]Harvard T.H. Chan School of Public Health ~ Boston ~ United States of America, [2]Fred Hutchinson Cancer Research Center ~ Seattle ~ United States of America

Large observational databases and cohort studies provide rich data for the investigation of complex phenomena to improve patient outcomes. Often, specific risk factors of interest may not have been collected or are difficult to ascertain due to cost constraints. The case-cohort study design is well-known as a cost-effective outcome-dependent sampling scheme [1]. Crucially, the case-cohort design has been shown to increase statistical efficiency for analysis, making it an attractive sampling strategy for studies embedded within large cohort studies. However, in clinical and public health settings, interest often lies in multiple outcomes. For example, following hematopoietic cell transplantation, interest lies in understanding simultaneous risk regarding acute graft-vs-host disease, a non-terminal event, and mortality, a terminal event. The joint model where the terminal event acts as a competing risk for the nonterminal event, but not vice-versa, is called the semi-competing risks framework[2]. Statistically, the key challenges that one faces when analyzing semi-competing risks data include respecting the competing risk role that the terminal event plays, and structuring the statistical dependence between the non-terminal and terminal events within a subject. Crucially, an appropriate semicompeting risks analysis helps avoid bias that results from naive statistical treatment of death and also permits learning about how the two events co-vary over time. The latter is particularly important to recognize because it helps align the research goals with the clinical settings they are meant to inform. Using a weighting scheme based on the selection probabilities for the case-cohort design, we have developed a method for estimation and inference with an illness-death model, the Cox model specification for semi-competing risks. Methodological results from simulations show that the results are consistent and efficient compared to analysis from the full cohort and subcohorts. We demonstrate the utility of the framework through application to transplant data from the Center for International Blood and Marrow Transplant Research. For such semi-competing risks settings, we developed a novel case-cohort analytical framework coupled with methods for estimation and inference within the frequentist paradigm.

[1] S. Haneuse, KH. Lee, Circulation Cardiovascular Quality and Outcomes, 9, 2015, 322-331.
[2] RL. Prentice, Biometrika, 73, 1986, 1-11.

PARALLEL SESSION MO9: LATENT VARIABLE MODELLING

MO9.1 Random effects models of tumour growth and interval breast cancer - a study of incident cases

Orsini L.*, Czene K., Humphreys K.
Karolinska Institutet ~ Stockholm ~ Sweden

Despite reasonably high participation in breast cancer screening programmes among women in the Nordic countries, and elsewhere in Europe, many breast cancer patients are still diagnosed symptomatically. Symptomatic cases diagnosed between two screening rounds are known as interval cancers, and have a worse prognosis than screen-detected ones. We aim to shed light on the nature of interval cancers by fitting continuous tumour growth models to data from a large study of incident breast cancer cases with detailed screening history, and tumour characteristics data. We address questions such as: how does the mean doubling time of tumours of interval cancers compare to those of screen-detected cancers? Among interval cancers, what is the tumour size distribution at the prior negative screen? Utilising previous results for the stationary distribution of tumour size in an unscreened population [1], we develop an analytical expression for the proportion of interval breast cancer cases among regularly screened women. As in Baker [2], who uses a multi-state model for lung cancer data to estimate the rate of transition from pre-clinical to clinical cancer based on modelling time to interval cancer after negative screening results, we avoid having to rely on estimated background cancer rates. Our calculations are based on parametric assumptions about tumour growth (random effects models) and detection processes (screening and symptoms). Using simulation, we show how the analytical expression can be used in a penalised likelihood approach to improve the precision of estimates of parameters in the tumour growth model when fitting to data from incident cases. We analyse data from 3493 invasive breast cancer cases diagnosed in Sweden 2001-2008. As well as using our model to characterise interval cancers in terms of doubling times and tumour size, we also show that our model-based expected incidence of interval breast cancers between screening rounds is similar to incidence patterns in a very large Nordic screening cohort. An analytical expression for the proportion of interval breast cancers in a screened population can be used to aid the fitting of tumour growth models to data from incident cases, and to shed light on the performance of population-based breast screening.

[1] G. Isheden, K. Humphreys, Modelling breast cancer tumour growth for a stable disease population, Statistical Methods in Medical Research, 28(3), 2019, 681–702.
[2] Baker, S.G., Modeling the mean time to interval cancer after negative results of periodic cancer screening, Statistics in Medicine, 40(6), 2021, 1429-1439.

Parallel Sessions | Monday 28 August 2023

MO9.2 Latent dynamic modeling with differential equations for individual disease trajectories

Hackenberg M.*[1], Pechmann A.[2], Kirschner J.[2], Binder H.[1]
[1]*Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg ~ Freiburg ~ Germany,* [2]*Department of Neuropediatrics and Muscle Disorders, Faculty of Medicine and Medical Center, University of Freiburg ~ Freiburg ~ Germany*

Using differential equations for describing dynamic processes is common practice in systems modeling, but less established in the statistics community, maybe due to difficulties in settings with larger numbers of variables to be modeled and/or a high level of noise. More specifically, ordinary differential equations (ODEs) provide a mechanistic model of the local changes of an underlying process, in contrast to regression techniques in statistics, which strive to obtain functions with a good fit on average in the course of time. From a clinical perspective, modeling of local changes with ODEs could be attractive for predicting an individual patient's disease progression in the immediate future given their current status. Yet, the shape of a function obtained from solving an ODE heavily depends on the initial value, i.e., the first observation in the course of time. A general approach for modeling with ODEs that decreases the emphasis on the initial value would thus be desirable. To address this, we propose an approach where each observation serves as the initial value to obtain multiple local ODE solutions which are used to build an estimator of the underlying dynamics. To deal with a larger number of variables, we combine this approach with dimension reduction by neural networks and use differentiable programming techniques to simultaneously optimize the dynamic model and the dimension reduction. For individual trajectories, we map baseline variables to person-specific ODE parameters. We illustrate how this enables the use of ODEs in longitudinal clinical registries. Specifically, we consider an application with spinal muscular atrophy (SMA) patients and a corresponding simulation design. We highlight how the parameters of ODEs can be used to assess the local change in health status at any point in time, in contrast to the global interpretation of functions fitted by regression modeling techniques, and discuss the challenge of interpreting such local parameters in a latent representation. We thus more generally illustrate how modeling with ODEs can be integrated in a statistical framework to provide a local perspective on underlying dynamics in a latent representation for predicting patients' individual disease course. Bandyopadhyay, S., Ganguli, B. & Chatterjee, A. (2011), 'A review of multivariate longitudinal data analysis', Statistical Methods in Medical Research 20(4), 299–330. Banga, J. R. (2008), 'Optimization in computational systems biology', BMC Systems Biology 2(1), 1–7.
Chen, T. Q., Rubanova, Y., Bettencourt, J. & Duvenaud, D. (2018), Neural ordinary differential equations, in S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi & R. Garnett, eds, 'Advances in Neural Information Processing Systems 31', 6572–658. Hackenberg, M., Harms, P., Pfaffenlehner, M., Pechmann, A., Kirschner, J., Schmidt, T. & Binder, H. (2022), 'Deep dynamic modeling with just two time points: Can we still allow for individual trajectories?', Biometrical Journal, 64, 1426– 1445.

MO9.3 Unifying probability and non-probability samples with misclassified covariate for improved inference

Yu Z.[1], Shen H.*[1], Li P.[2], Wu C.[2]
[1]*University of Calgary ~ Calgary ~ Canada,* [2]*University of Waterloo ~ Waterloo ~ Canada*

Probability samples are commonly used in clinical research as they are considered representative of the patient population of interest. However, not all variables required for analysis are precisely measured or observed. Non-probability samples can also be utilized, providing information on variables of interest, but are subject to biased sampling and measurement errors. Misclassification of categorical covariates is a common issue in both probability and non-probability samples. Ignoring misclassification and non-probability sampling can lead to biased estimates. We present a procedure for estimating the propensity score for individuals in the non-probability sample and propose a two-stage estimation process for integrating probability and nonprobability samples with misclassified covariates within a latent-variable framework. Our proposed method does not require validation data for the misclassified covariate and demonstrates improved inference over naive methods. We illustrate the performance of our proposed method using simulation studies and a real-world dataset. Our results show that the proposed method is effective in improving inference in non-probability samples with misclassified covariate and it is robust to model misspecification.

[1] Y. Chen, P. Li, C. Wu. Doubly robust inference with nonprobability survey samples. Journal of the American Statistical Association. 115(532), 2020, 2011-2021.
[2] H. Shen, R.J. Cook. Regression with incomplete multivariate surrogate responses for a latent covariate. Biostatistics & Epidemiology, 4(1), 2020, 247-264.

MO9.4 Correctly accounting for misclassification when linking latent groups With external variables

Proust--Lima C.*[3], Dussartre M.[3], Philipps V.[3], Samieri C.[3], Gustafson P.[1], Shaw P.[2]
[1]*Department of Statistics, University of British Columbia ~ Vancouver ~ Canada,* [2]*Biostatistics Division, Kaiser Permanente Washington Health Research Institute ~ Seattle ~ United States of America,* [3]*Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center ~ Bordeaux ~ France*

Latent groups constitute a convenient solution to summarize complex multidimensional exposures such as lifestyle behaviors (e.g., food intakes, physical activity) or cardiometabolic factors. Once the latent group structure is defined and each individual is assigned to a group, predictors of the latent groups or association between the latent groups and health outcomes can be assessed in subsequent regression models. Yet, the quality of inference in the subsequent analyses may be altered by the inherent error of classification when assigning each individual to a specific latent group. As part of the "measurement error and misclassification" topic group of the STRATOS initiative (TG4), our goal was to review the methods adopted in the literature to correct for this misclassification and, using simulations, compare their performance and potential biases to ultimately provide recommendations. Four methods were identified:
(i) the naive approaches which directly use the class assignment in the subsequent regression, potentially weighted by posterior class membership probabilities;
(ii) a bias-adjusted method that accounts for the assignment error in the subsequent regression using weights;
(iii) a rewriting of the subsequent regression as a latent class model with specifically determined class-membership probabilities that incorporate the assignment error (bias-adjusted LCM);
(iv) a two-stage estimation of the latent class model and the subsequent regression using their joint likelihood.
In extensive simulations exploring different levels of group separation and strengths of association with the outcome, we found that naive methods systematically showed substantial bias. The bias-adjusted weighted method showed residual bias with ambiguous classification. In contrast, bias-adjusted LCM and two-stage methods, which apply to various data (e.g., longitudinal, multivariate, survival), showed correct inference in all the scenarios provided the variance of the estimates correctly accounted for the sequential estimation procedures. The methods were further illustrated in an application assessing the association between groups of late-life lifestyle behavior and brain health outcomes (e.g., cognitive trajectory, time-todementia) in the population-based Three City cohort. Misclassification stemming from latent class models cannot be neglected when the classification is used in subsequent analyses, even when the classes are well separated.

Bakk Z & Kuha J (2021). Relating latent class membership to external variables: An overview. The British Journal of Mathematical and Statistical Psychology, 74(2), Art. 2. Elliott MR, Zhao Z, Mukherjee B, et al (2020). Methods to Account for Uncertainty in Latent
Class Assignments When Using Latent Classes as Predictors in Regression Models, with Application to Acculturation Strategy Measures. Epidemiology (Cambridge, Mass.), 31(2), 194- 204.
Proust-Lima C, Saulnier T, Philipps V, et al (2022). Describing complex disease progression using joint latent class models for multivariate longitudinal markers and clinical endpoints. ArXiv:2202.05124
Vermunt JK (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. Political Analysis, 18(4), Art. 4.

Parallel Sessions | Monday 28 August 2023

MO9.5 Estimation of the causes of fever using partial latent class analysis

Keddie S.*[1], Hopkins H.[1], Crump J.[2], Feasey N.[3], Roberts C.[1], Baerenbold O.[4], Keogh R.[1], Bradley J.[1]
[1]LSHTM ~ London ~ United Kingdom, [2]University of Otago ~ Dunedin ~ New Zealand, [3]Malawi-Liverpool-Wellcome Trust Clinical Research Programme ~ Blantyre ~ Malawi, [4]Novartis ~ Basel ~ Switzerland

Fever is a common manifestation of disease and a leading cause of healthcare seeking and hospital admission in sub-Saharan Africa and Asia. Despite its significant contribution to global morbidity and mortality, the ability to identify the cause of febrile illness that is not malaria is challenging at point-of-care facilities in many settings. This results in inadequate or inappropriate treatment. The FIEBRE study [1] was established to improve understanding of fever epidemiology in sub-Saharan and Asia and seeks to estimate the prevalence of treatable or preventable causes of fever with the goal of improving treatment and helping to guide fever case management. To achieve these goals, FIEBRE recruited paediatric and adult outpatients and inpatients as well as controls. Each recruited participant provided clinical data, pharyngeal swabs and a venous blood sample. These samples were assessed for infections using a predetermined set of diagnostic testing strategies. The case-control design of the FIEBRE study brings statistical challenges, such as the integration of; case only data (for specific infections), case and control data, multiple tests for the same infection and measurement error of these tests. The primary objective of this work is to expand the partial latent class analysis approach estimate the causes of fever in the FIEBRE dataset.
Following the analysis of The Pneumonia Etiology Research for Child Health (PERCH) study, we used Bayesian partial latent class analysis [2], allowing for the sample collection structure of FIEBRE and sensitivity and specificity of diagnostic tests of pathogens of interest. Priors for diagnostic test sensitivity and specificity imperfections were obtained from random-effect metaanalyses, which are themselves based on Bayesian latent class analyses. We will present estimates (including uncertainty) of attributable fractions by site, age, and HIV status for those infections that are treatable and/or preventable causes of fever.
The estimated prevalences will be the first of their kind for some infections and will be important additions to projects like the global burden of disease report. This work also highlights the benefits of a Bayesian approach in aetiological research.

[1] Hopkins H, Bassat Q, Chandler CI, Crump JA, Feasey NA, Ferrand RA, et al. Febrile Illness Evaluation in a Broad Range of Endemicities (FIEBRE): protocol for a multisite prospective observational study of the causes of fever in Africa and Asia. BMJ Open. 2020;10(7):e035632.
[2] Wu Z, Deloria-Knoll M, Hammitt LL, Zeger SL, Pneumonia Etiology Research for Child Health Core T. Partially latent class models for case-control studies of childhood pneumonia aetiology. J R Stat Soc Ser C Appl Stat. 2016;65(1):97-114.

PARALLEL SESSION MO10: CAUSAL INFERENCE 1

MO10.1 G-formula for causal inference via multiple imputation

Bartlett J.*[1], Olarte Parra C.[1], Daniel R.[2]
[1]London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom, [2]Cardiff University ~ Cardiff ~ United Kingdom

G-formula is a popular approach for estimating treatment or exposure effects from longitudinal data that are subject to time-varying confounding. G-formula estimation is typically performed by Monte-Carlo simulation, where potential outcomes under the treatment regimes of interest are simulated from models fitted to the original dataset. Inference for the resulting estimates is usually performed using bootstrapping. In practice the dataset to be analysed often has missing values, which are commonly handled using the method of multiple imputation (MI). Given the similarities between G-formula by Monte-Carlo simulation and multiple imputation, in this work we investigated the relationship between the two techniques, in particular to establish if a single procedure could be developed to simultaneously handle both missing data and simulation of potential outcomes. We show that G-formula can be implemented by exploiting existing methods for multiple imputation (MI) for synthetic data. This involves augmenting the original data with new rows where confounder and outcome variables are set to missing and the longitudinal treatment variables are set to the values under the treatment regime(s) of interest. We show an existing modified version of Rubin's variance estimator developed in the context of synthetic MI for surveys is valid for the G-formula via MI approach we propose. Importantly, this obviates the need to use bootstrapping. We also describe how performing G-formula using MI methods can be used to impute missing data in the original dataset as part of the procedure. Simulation results suggest the approach works well with as few as 25 imputations.
Multiple imputation seems an attractive approach to performing G-formula for causal inference, particularly when the dataset being analysed suffers from missing values. It avoids the need to use bootstrapping for inferences, which speeds up computation, instead relying on a previously proposed modification to Rubin's pooling rules. It is practically appealing, given that existing MI software can be adapted to perform G-formula via MI. However, we also provide an R package, gFormulaMI, which links with the mice MI package, to facilitate use of the approach.
Bartlett JW, Olarte Parra C, Daniel RM. G-formula for causal inference via multiple imputation, arXiv:2301.12026, 2023.

Parallel Sessions | Monday 28 August 2023

MO10.2 Implementation of g-computation in practice: a new diagnostic tool to guide outcome model specification

Shepherd D.*[1], Vansteelandt S.[2], Moreno--Betancur M.[1]
[1]*Murdoch Children's Research Institute & The University of Melbourne ~ Melbourne ~ Australia,* [2]*Ghent University ~ Ghent ~ Belgium*

Causal inference is a central goal of clinical and public health studies, investigating the effect of an exposure on an outcome of interest. For many studies, reliance on observational data is common, requiring confounding-adjustment methods to estimate causal effects. G-computation is one such method, which in the point-exposure setting extends outcome regression by allowing exposure-confounder interactions in the outcome model and predicts counterfactual outcomes across the sample under each exposure. Consistent estimation with g-computation relies on correct specification of the outcome model, which cannot be empirically verified. It is recommended that variables included in the model should be driven by expert-knowledge. However, there is no formal guidance or diagnostic tool available to aid the parametric specification of the outcome model, for example which interaction or non-linear terms to include, presenting a challenge when applying g-computation in practice. In this work we aimed to address this gap, proposing a new diagnostic tool to guide the outcome model specification in gcomputation. We propose a diagnostic tool that distinguishes between candidate outcome model specifications based on the expected bias in the resulting g-computation estimator. This bias is derived from the efficient influence curve[1] and depends on the outcome model specification and propensity scores (PS). Our method uses the expected bias for the given outcome model specification, with the PS estimated flexibly and equally across candidate outcome model specifications using SuperLearner[2]. Thus, the differences in the resulting statistic depend solely on the outcome model specification. We investigated performance of the diagnostic tool in a simulation study based on the Longitudinal Study of Australian Children (LSAC), considering a range of true outcome generation models, sample sizes, and confounding strengths. Results indicated the statistic was optimised for the correctly specified model in most settings, and appropriately discriminated the model that minimized the bias in effect estimates. The diagnostic tool was illustrated with real data in the LSAC study, with implementation of the tool available as an R function. We proposed a promising new diagnostic tool to guide outcome model specification in gcomputation, enhancing implementation of this approach in epidemiologic studies.

[1] van der Laan, MJ, Rose S. Targeted Learning. New York: Springer. 2011.
[2] van der Laan MJ, Polley EC, Hubbard AE. Super learner. Statistical applications in genetics and molecular biology. 2007. Sep 16;6(1).

MO10.3 Investigating positivity violations in marginal structural survival models: a study on estimator performance

Spreafico M.*, Fiocco M.
Mathematical Institute, Leiden University ~ Leiden ~ Netherlands

Marginal structural models (MSM) with inverse probability of treatment weighting (IPTW) are a class of statistical models for estimating the causal effect of an exposure on a survival outcome in presence of time-dependent confounders. Longitudinal observational data are increasingly used in this context, as IPTW allows to create a new pseudo-population where exposure is no longer affected by confounders and estimates are reliable if the four key causal assumptions (i.e., no interference, positivity, consistency, conditional exchangeability) hold. However, these assumptions are difficult to test and often rely on expert knowledge. The severity with which possible violations of the assumptions affect the results is not widely known nor understood, posing a statistical challenge.
In this research, we focus on the positivity assumption (i.e., each possible exposure level occurs with positive probability within each level/combination of observed confounders) and consider several scenarios where violation is present. The aim and novelty of this work is to investigate the effect of positivity violations on the performance of MSM-IPTW-estimators in a survival context where time-dependent confounding is present. Considering the algorithms for simulating longitudinal data from MSMs proposed by [1] and [2], we approached the problem using two alternative simulation settings with binary exposure. We extended both algorithms to incorporate strict violations of positivity in different scenarios, such as by varying (i) the confounder threshold below/above which no subject is exposed to treatment, (ii) the length of follow-up (to see whether violations are accentuated or attenuated over a longer period), or (iii) the sample size. We also considered non-strict cases where some exposure levels are rare within certain confounder levels. Results showed that even relatively modest violations of the positivity assumption yield to estimators that are very unstable and/or may exhibit high variability.
To assess positivity is rather delicate as the robustness of the estimators can be very poor, even when the range within which positivity violations occur is modest. To our knowledge, this is the first study that analyses violations of the positivity assumption in a survival context or, more generally, in a time-dependent context.

[1] W.G. Havercroft, V. Didelez. Simulating from marginal structural models with time-dependent confounding. Statistics in medicine, 2012; 31(30), 4190-4206.
[2] R.H. Keogh, S.R. Seaman, J.M. Gran, S. Vansteelandt. Simulating longitudinal data from marginal structural models using the additive hazard model. Biometrical Journal, 2021; 63(7), 1526-1541.

MO10.4 Estimating optimal dynamic treatment regime for survival time outcome using g-estimation

Seaman S.*[1], Vansteelandt S.[2]
[1]*University of Cambridge ~ Cambridge ~ United Kingdom,* [2]*University of Gent ~ Gent ~ Belgium*

A dynamic treatment regime specifies a rule for how treatment decisions over time should depend on observed, possibly time-dependent, characteristics of the patient. The optimal regime is the rule that optimises some outcome of interest, e.g. maximises the probability of survival to a particular time. Such regimes can be estimated from sequentially randomised trials or from observational data. We propose a g-estimation method for estimating a dynamic treatment regime that maximises the probability of survival to a particular time using observational data. This method avoids drawbacks associated with related methods based on structural nested accelerated failure time models, e.g. assumptions about the persistence of treatment effects, and it has the advantage over inverse probability weighting methods of being less prone to instability caused by highly variable weights. We shall describe this method and present a simulation study comparing its performance with that of other methods. This g-estimation method is particularly useful in settings where confounders are strongly predictive of treatment assignment. Simoneau et al. (2020) Estimating Optimal Dynamic Treatment Regimes With Survival Outcomes. Journal of the American Statistical Association, 115: 1531-1539
Hager et al. (2018) Optimal Two-Stage Dynamic Treatment Regimes from a Classification Perspective with Censored Survival Data. Biometrics 74: 1180–1192

Parallel Sessions | Monday 28 August 2023

MO10.5 Handling symptomatic treatment in alzheimer's disease trials – estimators for a hypothetical strategy

Lasch F.*[1], Loh W.W.[2], Guizzaro L.[1]
[1]*European Medicines Agency ~ Amsterdam ~ Netherlands,* [2]*University of Washington - Department of Statistics ~ Washington ~ United States of America*

In clinical trials investigating the effect of a disease-modifying treatment for Alzheimer's Disease, the initiation of symptomatic medications is a relevant Intercurrent Event following the Definition of the ICH E9(R1) addendum on Estimands and Sensitivity analyses [1]. As stated in the European Medicines Agency Alzheimer's Disease guideline [2], a hypothetical estimand strategy is of regulatory interest to understand the treatment effect in the hypothetical scenario that no symptomatic treatment occurred. However, it is unclear which estimators are best suited for estimating this estimand. In earlier work [3], we investigated a simplified scenario where the initiation of symptomatic treatment was possible at one fixed timepoint and determined demediation by g-estimation as the most promising estimator. However, in real clinical trials symptomatic treatment can be initiated at various timepoints longitudinally. Therefore, we seek to comprehensively examine the performance of de-mediation by g-estimation [4], as well as various modifications that exploit assuming a common effect across time. Building on the de-mediation by g-estimation approach [4] that removes the effect of the mediator separately at each timepoint, we developed three different modifications that utilize the assumption of a common effect of the intercurrent event. Two modifications estimate the effect of the mediator on the outcome at each timepoint and average them before de-mediation. The weights used to calculate the average are either the number of patients that started symptomatic treatment or the model-derived standard error of the mediator's effect. In a third modification, we estimate the effect of the mediator simultaneously across all timepoints by using a mixed model. The simulation study shows that all methods are unbiased empirically. The modifications empirically control the type-I-error rate at the nominal level and show a small gain in power as compared to the existing g-estimation approach. In ongoing work, we investigate the robustness of the modified approaches to model misspecifications.
The proposed modified de-mediation by g-estimation approaches are promising improvements that can enhance statistical inference assuming a common mediation effect over time. We offer detailed insights into the strengths and limitations of these modifications over existing approaches.

[1] ICH (2019), "Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1)."
[2] EMA (2018), "Guideline on the Clinical Investigation of Medicines for the Treatment of Alzheimer's Disease."
[3] Florian Lasch, Lorenzo Guizzaro, Frank Pétavy & Ciro Gallo (2022) A Simulation Study on the Estimation of the Effect in the Hypothetical Scenario of No Use of Symptomatic Treatment in Trials for Disease-Modifying Agents for Alzheimer's Disease, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2022.2055633
[4] Loh, W. W., Moerkerke, B., Loeys, T., Poppe, L., Crombez, G., and Vansteelandt, S. (2019), "Estimation of Controlled Direct Effects in Longitudinal Mediation Analyses with Latent Variables in Randomized Studies," Multivariate Behavioral Research, 55, 763–785. DOI: https://doi.org/10.1080/00273171.2019.1681251.

PARALLEL SESSION MO11: MACHINE LEARNING 1

MO11.1 Fuzzy sets in probability trees: a novel interpretable ai decision making model

Capitoli G.*[4], Ambags E.L.[3], Nobile M.S.[2], Liò P.[1]
[1]*Department of Computer Science and Technology, University of Cambridge ~ Cambridge ~ United Kingdom,* [2]*Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice ~ Venice ~ Italy,* [3]*Department of Technology, University of Eindhoven ~ Eindhoven ~ Netherlands,* [4]*Department of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy*

The need for transparency and interpretability is increasingly being recognized as a central theme to be addressed by Artificial Intelligence (AI) research, especially when it operates in healthcare [1]. Hence, the need for fully human-understandable models is increasingly being recognised as a central theme. In this work, we propose Fuzzy Sets in Probability Trees (FPT), a novel method that combines probabilistic trees and fuzzy logic. This approach is fully interpretable as it allows clinicians to generate, and verify the entire clinical decision process. FPT allows for using the existing framework of PTs, while incorporating uncertainty about the data, allowing for a flexible description of vague variables. Thus, enabling us to incorporate human expert knowledge in the form of fuzzy membership functions to probabilistic trees. While PTs require well-defined discrete concepts, this is seldom the case in real-world scenarios. Especially in the medical field, where the variables are often fuzzy, vague or ambiguous. By using the FPT approach, and by means of carefully crafted fuzzy sets, we can incorporate the inherent uncertainty in variables to balance the probabilities. The integration of PTs and fuzzy sets, leading to an FPT, will provide an AI method that is aligned with the way humans reason. Furthermore, FPTs can represent circumstances or explanations that cannot be represented with other techniques (e.g., Bayesian networks), paving the way to a novel form of interpretable AI. Constructing the FPT is an iterative process, done in collaboration with domain experts. The selection of the features and the order of the features in the tree are based on domain knowledge (deduction), however, the transition probabilities in the trees are based on the data (induction). The performance of the FPT is compared to several other interpretable decisionmaking models, namely regular PTs, Decision Trees, and Logistic Regression.
The integration of probabilistic trees and fuzzy reasoning, brings significant nuances that are lost when using the crisp thresholds set by probabilistic decision trees. This lead to a hybrid tree, which will provide an AI system that is aligned with the way humans reason and that can effectively support clinicians in the diagnosis decision process.

[1] Shortliffe, E. H., Sepulveda, M. J. Clinical decision support in the era of artificial intelligence. JAMA, 320, 2018, 2199–2200.
[2] Zadeh, L. A. Fuzzy sets. Information and control, 8, 1965, 338–353.

Parallel Sessions | Monday 28 August 2023

MO11.2 Random survival forests for analysing survival data with recurrent events

Murris J.*[1], Lavenu A.[2], Katsahian S.[3]
[1]HeKA, Inria, Inserm, Cordeliers Research Centre, Pierre Fabre, Sorbonne University, University Paris Cité ~ Paris ~ France, [2]CIC-1414, Inserm, IRMAR – CNRS 6625, University of Rennes 1 ~ Rennes ~ France, [3]CIC-1418, Georges Pompidou European Hospital, AP-HP, HeKA, Inria, Inserm, Cordeliers Research Centre, Sorbonne University, University Paris Cité ~ Paris ~ France

Random survival forests (RSF) are commonly used in medical research. Such approaches have shown their utility in modelling complex relationships between predictors and survival outcomes, overcoming for instance linearity or low dimensionality assumptions (number of individuals greater than the number of predictors). Nevertheless, such RSF have not been adapted to survival data with recurrent events. This work presents an extension of the RSF for this type of data by exploiting concepts from non-parametric survival analysis and statistical learning.
Based on the survival tree methodology introduced by Ishwaran, et al. (2008), the key construction steps were adapted to the pattern of recurrent events and using a non-parametric approach. [1] The splitting rule at each node for data partitioning is the pseudo-score test and the estimation at each terminal node is carried out by the Nelson-Aalen estimator of the mean cumulative function. As an ensemble method, the random forest is then obtained by combining several of such trees to produce one optimal model, ensuring estimates stability and addressing overfitting issues. The score is an adjustment of the Harrell's concordance index to recurrent event data and considers the event occurrence rates across individuals. Variable importance through permutation is also computed. Several trees were already built on bladder dataset from the survival package in R. Results were very promising with C-index values greater than 0.5 and up to 0.89. Further results are yet to be available based on both simulated data and readily accessible samples for the application of the proposed approach.
The proposed approach for survival analysis with recurrent events represents a novel and promising method for analysing time-to-recurrence data. It namely embodies new basics for further development and applications in statistical learning. Therefore, this work is deemed of great interest for survival analysis with recurrent events in medical research.

[1] Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests.

MO11.3 Evaluating the sample size requirements of tree-based machine learning techniques for clinical risk prediction

Kalaycioglu Oya[2] (WINNER OF ISCB44 CONFERENCE AWARD FOR SCIENTISTS), Ambler G.[1], Pavlou M.[1], Akhanli S.E.[3], Omar R.[1]
[1]Department of Statistical Science, University College London ~ London ~ United Kingdom, [2]Department of Biostatistics and Medical Informatics, Bolu Abant Izzet Baysal University ~ Bolu ~ Turkey, [3]Department of Statistics, Mugla Sitki Kocman University ~ Mugla ~ Turkey

When developing clinical risk prediction models using machine learning techniques (MLTs), it is not clear how much data are required to ensure reliable predictions using MLTs. We aimed to evaluate whether the sample size recommendations for the development of prediction models with logistic regression are applicable or can be adapted for tree-based ensemble MLTs. We considered three categories of MLTs: bagged classification trees, random forests, and gradient boosting; in addition, we used standard logistic regression. We performed simulations based on large cardiovascular datasets, one of which had non-linear associations between outcome and covariates. We determined sample size using Riley's equation based on mean absolute prediction error (MAPE) and formed training datasets by sampling covariates from the original datasets. We considered data-generation mechanisms (DGMs) where the outcome was simulated from: (i) each of the methods under investigation; (ii) a 'neutral' model that includes all main effects and two-way interactions of the categorised covariates, and (iii) logistic regression model with non-linear effects. The performance of the risk models was evaluated on validation data using MAPE, Brier score, C-statistic, and measures of calibration. Our results indicate that the MLTs could not achieve the target MAPE with the sample size calculated using Riley's equation. Across scenarios where DGM matched the model, boosting required a sample size twice that of the logistic regression to achieve the same performance while random forests and bagging did not reach the target performance even if the sample size was increased by 15 times. Boosting outperformed logistic regression in large sample sizes under the neutral DGM. In presence of non-linear relationships random forest and boosting were superior to logistic regression models that ignored non-linear effects, regardless of sample size.
The MAPE sample size equation for logistic regression can potentially be adapted for boosting, but not for random forests or bagging. The performance of boosting was compatible to that of conventional logistic regression for large sample sizes, often exceeding that of random forest and bagging.

[1] Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, Moons KGM, Collins G, van Smeden M. Calculating the sample size required for developing a clinical prediction model. BMJ 2020; 368:m441.
[2] Austin PC, Harrell FE Jr, Steyerberg EW. Predictive performance of machine and statistical learning methods: Impact of data-generating processes on external validity in the "large N, small p" setting. Stat Methods Med Res 2021; 30(6):1465-83.

Parallel Sessions | Monday 28 August 2023

MO11.4 Impact of temporal breast density changes on the prediction of breast cancer in women from screening programs

Rakez M.*[1], Guillaumin J.[2], Chick A.[2], Fillard P.[2], Amadeo B.[3], Rondeau V.[1]
[1]Biostatistics Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219 ~ Bordeaux ~ France, [2]Therapixel, 455 promenade des anglais ~ Nice ~ France, [3]EPICENE Team, Bordeaux Population Health Center, ISPED, Centre INSERM U1219 ~ Bordeaux ~ France

Breast cancer (BC) is the leading cause of cancer death in women worldwide. Mammographybased screening programs reduce BC mortality by promoting earlier detection. The frequency of screening visits is country-specific and can vary over the screening period. The mammography's sensitivity depends on breast density (BD). The latter is subject to changes over time, affecting the risk of a BC diagnosis. Women with high BD are likelier to develop BC, and their mammography's sensitivity is lessened. Thus, to better understand the impact of temporal BD changes on BC diagnosis risk in the screening setting, we propose a new methodology to predict BC risk accounting for the deep learning assessment of the sequential BD.
From the sequential and complete mammography exams of 131,455 women participating in the BC screening program, the percent density (PD), a quantitative estimation of a woman's BD at each visit, is estimated using the MammoDL[1] algorithm. This segmentation model comprises two successive modified U-Nets allowing for breast identification from the entire mammogram first and dense tissue region delineation from the breast region second. A ResNet-34 replaces the U-Net encoder to alleviate training challenges. In addition, this model is fine-tuned to extend its use to for presentation GE and Hologic vendors' images. Then, a joint model for a linear biomarker and a binary outcome[2] is implemented; First, the temporal trajectory of the PD is described using a linear mixed-effect model, adjusted on factors impacting the BD, such as age.
This sub-model is flexible in dealing with irregular intervals between screening visits and outcome-dependent drop-out. Second, the individual and dynamic prediction of BC diagnosis is estimated conditionally on the biomarker's intermediate longitudinal measurements and is defined over the screening period. This probability is derived for each woman and is dynamically updated as PD measurements accumulate.
We propose a reproducible method to estimate BD's temporal evolution and its impact on BC diagnosis. The segmentation model gives quantitative estimations of the BD at each screening visit. The joint model uses the biomarker's repeated measurements to dynamically update the individual BC diagnosis prediction throughout the screening period.

[1] R. Muthukrishnan, A. Hayler, K. Katti, et al., arXiv.2206.05575v3, 2022.
[2] R. Dandis, S. Teerenstra, L. Massuger, et al., Biometrical Journal, 62, 2020, 398-413

MO11.5 Using chatgpt for classification of pediatric injuries from emergency department records

Lorenzoni G.*[1], Berchialla P.[2], Sciannameo V.[2], Baldan G.A.[1], Bressan S.[1], Da Dalt L.[1], Gregori D.[1]
[1]University of Padova ~ Padova ~ Italy, [2]University of Turin ~ Turin ~ Italy

Timely and accurately identifying, classifying, and analyzing injury data from pediatric emergency department records are critical for injury prevention and resource allocation. However, the diagnosis reported in the emergency department records is often not coded, and the manual coding of these records is labor-intensive, time-consuming, and prone to errors [1]. In this era of rapidly evolving artificial intelligence, applying Generative Pre-trained Transformer (GPT) models offers a promising solution to streamline this process [2]. The present study aims at implementing an automatic coding system using GPT-based models to extract and classify injury data in the Italian language from pediatric emergency department records.
The study included 283,468 admission records to the pediatric Emergency Department of Padova University Hospital from 2007 to 2018. Each access is mandatorily registered in an electronic data collection system. For each emergency department access, both coded (administrative and demographic data) and free-text (diagnosis) information are reported. A random subset of 40,030 records underwent classification of free-text diagnosis (as injury or not) by an expert clinician (gold standard). ChatGPT was used for the classification task. ChatGPT is an extensive language model developed by OpenAI. It is based on the GPT architecture. Specifically, the GPT-3.5 variant was employed for the present work. ChatGPT was accessed through a public application programming interface (API) using R software. Preliminary results showed a classification accuracy of 96.2%. The tool's ability to correctly classify the injuries (sensitivity) was 95%, while the specificity was 96.5%.
The performance of the classification task was excellent. The present results demonstrate the feasibility of GPT-based models for processing unstructured free text information from medical records.

[1] Klein DO, Rennenberg R, Gans R, et al. Limited external reproducibility restricts the use of medical record review for benchmarking. BMJ open quality 2019;8:e000564.
[2] Patel SB, Lam K, Liebrenz M. ChatGPT: Friend or Foe. Lancet Digit Health 2023;5:e102.

PARALLEL SESSION MO12: REAL WORLD DATA

MO12.1 Emulating an existing trial of treatments for prostate cancer using Real-world data: methods and challenges

Chesang C.*, Sharples L., Cowling T., Keogh R.
London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom

Randomized controlled trials (RCTs) are the gold standard for establishing effects of treatments. The additional value of evidence from "real-world data" (RWD), such as electronic health records, is increasingly recognized. Emulation of existing RCTs using RWD can provide evidence of the validity of RWD for estimating treatment effects and highlight where biases might arise. We used the trial emulation framework [1] to emulate the PR07 trial [2] of treatments of prostate cancer using UK national cancer data. PR07 evaluated whether the addition of radiotherapy to hormone therapy improved survival in men with high-risk prostate cancer. Challenges for emulating this RCT include that randomization does not start at cancer diagnosis and radiotherapy could be initiated at any time within 8-weeks of randomization. We investigated methods for emulating the PR07 trial or a closely related trial.
We developed a protocol for the emulation of PR07 using UK cancer registry data, considering cancer diagnosis as the time of potential randomization. As few patients initiated radiotherapy within 8 weeks of diagnosis in the RWD, we considered target trials in which the grace period for initiating radiotherapy ranged from 4-6 months. We used the cloning-censoring-weighting (CCW) approach to estimate the average treatment effect. We also considered the sequential stratification approach, which constructs a series of trials by redefining time zero each time a participant initiates radiotherapy and matches them to contemporaneous controls. This approach targets the treatment effect in the treated. The CCW analysis using grace periods of 4, 5 and 6 months after prostate cancer diagnosis resulted in hazard ratios (HRs) of 0.57 (95% CI: 0.28-0.88), 0.53 (95% CI: 0.30-0.74) and 0.47 (95% CI: 0.29-0.65) respectively over 7 years of follow-up. Sequential stratification yielded a HR of 0.70 (95%: 0.54-0.88) over 7 years of followup. Estimates of risk and risk differences will also be presented.
Alignment of time zero with the time-varying treatment was the primary challenge in emulating PR07 trial. Our findings using RWD are broadly similar to those from the RCT, which reported a HR of 0.77 (95%CI: 0.61-0.98) over 7 years of follow-up.

[1] M.A. Hernan, J.M. Robins, Am J Epidemiol., 183, 2016, 758-764.
[2] P. Warde, M. Mason, K. Ding, et al., Lancet, 378, 2011, 2104-2111.

MO12.2 Augmenting treatment arms with data from expanded access using propensity-score weighted power priors

Polak T.[1], Labrecque J.[2], Van Rosmalen J.*[1]
[1]Department of Biostatistics, Erasmus MC ~ Rotterdam ~ Netherlands, [2]Department of Epidemiology, Erasmus MC ~ Rotterdam ~ Netherlands

The incorporation of real-world data to supplement the analysis of trials and improve decisionmaking has spurred the development of statistical techniques to account for introduced confounding. Recently, hybrid methods have been developed through which measured confounding is first attenuated via propensity scores and unmeasured confounding is addressed through (Bayesian) dynamic borrowing. Most efforts to date have focused on augmenting control arms with historical controls.
Here we consider augmenting treatment arms through expanded access, which is a pathway of non-trial access to investigational medicine for patients with seriously debilitating or lifethreatening illnesses. Motivated by a case study on expanded access, we developed a novel method (the ProPP) that provides a conceptually simple and easy-to-use combination of propensity score weighting and the modified power prior. Our weighting scheme is based on the estimation of the average treatment effect of the patients in the trial, with the constraint that external patients cannot receive higher weights than trial patients. The causal implications of the weighting scheme and propensity-score integrated approaches in general are discussed. In a simulation study our method compares favorably with existing (hybrid) borrowing methods in terms of precision and type-I error rate. We illustrate our method by jointly analysing individual patient data from the trial and expanded access program for vemurafenib to treat metastatic melanoma.
Our method provides a double safeguard against prior-data conflict and forms a straightforward addition to evidence synthesis methods of trial and real-world (expanded access) data.

Parallel Sessions | Monday 28 August 2023

MO12.3 Sequence analysis techniques to evaluate drugs-based diagnostic therapeutic paths in heart failure patients

Fontana N.*, Savaré L., Ieva F.
MOX, Department of Mathematics, Politecnico di Milano ~ Milan ~ Italy

Heart Failure (HF) is currently the most common cardiovascular reason for hospitalization in adults and its incidence is increasing globally. Pharmacotherapy is the cornerstone of its treatment, but several studies suggest high non-adherence to drugs among patients with this pathology, making therapy non-adherence a severe medical problem on a global scale [1]. For this reason, we focus on the therapeutic pathways administered to HF patients using an innovative method at the expense of more traditional ones, called state-sequence analysis (SSA). This technique provides a deeper description of drug prescriptions over time and allows extraction of drug-utilization patterns and their association with health outcomes. We apply such methods to an administrative database of hospitalized HF patients in the Lombardy region. Our first goal is to apply SSA to transform the longitudinal data about drugs purchased into simplified and workable information that can provide knowledge on the patients' behaviour to the therapies administered, focusing on the five mainly recommended drugs to HF patients. Once the sequences are constructed, a metric must be defined to measure their dissimilarity. We apply the optimal matching distance, which generates edit distances that are the minimal cost, in terms of insertions, deletions and substitutions, for transforming one sequence into another. Once the dissimilarity matrix is computed, different cluster algorithms are performed to construct partitions of the sequences into distinct groups representing different drug-utilization patterns. This information is integrated into the predictive models to evaluate its association with healthrelated outcomes, highlighting the impact of different sequences behaviours on patients' prognoses.
Compared to commonly used baseline measures, which omit some time-dependent information, this methodology yields more realistic and valuable results since it considers the evolution of each patient's clinical path and allows for the analysis of polytherapy taken by the patient using a single patient descriptor. Therefore, SSA allows a change of perspective in the analysis of the prescriptions, moving from a transversal and syntactical approach to a holistic one that exploits the information available using statistical tools, slightly more complex than traditional methods.

[1] J.Wu, D. K. Moser, T. A. Lennie, P.V. Burkhart: Medication adherence in patients who have heart failure: a review of the literature. Nursing Clinics Of North America, 43, 2008, 133-53, vii-viii.

MO12.4 Developing an algorithm to identify breast cancer recurrences using routinely collected data in england

Probert J.*[1], Dodwell D.[1], Charman J.[2], Broggio J.[2], Coleman R.[3], Darby S.[1], Mannu G.[1]
[1]Nuffield Department of Population Health, University of Oxford ~ Oxford ~ United Kingdom, [2]National Cancer Registration and Analysis Service, National Disease Registration Service ~ Birmingham ~ United Kingdom, [3]Department of Oncology and Metabolism, University of Sheffield ~ Sheffield ~ United Kingdom

Breast cancer is the commonest cancer in the UK, with over 55,000 women diagnosed each year.[1] Although a large amount of population-based information is available on death and survival from breast cancer in the UK, little information is available on breast cancer recurrence. We aimed to develop and validate a method using routinely collected data (RCD) to identify breast cancer recurrences in England. We identified all information that may indicate a recurrence within the population-level datasets curated by the National Disease Registry Service, including the Cancer Analysis System, Hospital Episode inpatient Statistics, Radiotherapy Treatment Dataset, Cancer Waiting Times dataset and Digital Imaging Dataset. We developed an algorithm to identify recurrences based on this information and compared these recurrences with the recurrence information collected, and independently verified, by a UK randomised controlled trial called the AZURE trial.[2] The AZURE trial randomised 1902 women in England with stage II-III breast cancer between 2004 and 2006. The median follow-up was 9.6 years from randomisation, and recurrences identified from RCrtD were restricted to AZURE's follow-up period. A sample of 150 women were selected for training our algorithm before it was validated on the remaining 1752 women. During validation, 543/579 (94%) women recorded in AZURE as having a recurrence in the ipsilateral or contralateral breast/regional lymph nodes or a distant metastasis were identified as having a recurrence by the algorithm using RCD. In the case of distant recurrences, the agreement between the data sources was strong (κ=0.76) with a sensitivity of 89% and sensitivity of 90%. The dates of distant recurrences recorded in AZURE and by our algorithm were within 3 months of each other in 65% of cases, and within 12 months in 85% of cases. There is good agreement between distant recurrences identified by our algorithm and those recorded in the AZURE trial. These findings demonstrate the potential of using RCD to identify breast cancer recurrences. Additional work is needed to further improve the accuracy of recurrences identified from RCD. However, it could be an important tool for future trials in breast cancer, particularly for improving follow-up surveillance and making follow-up more costeffective.

[1] Cancer Research UK. Breast cancer statistics London, UK: Cancer Research UK,; 2023 [Available from: https://www.cancerresearchuk.org/health-professional/cancerstatistics/statistics-by-cancer-type/breast-cancer#heading-Zero accessed 6/3/2023 2023.
[2] Coleman R, Cameron D, Dodwell D, Bell R, Wilson C, Rathbone E, Keane M, Gil M, Burkinshaw R, Grieve R, Barrett-Lee P, Ritchie D, Liversedge V, Hinsley S, Marshall H. Adjuvant zoledronic acid in patients with early breast cancer: final efficacy analysis of the AZURE (BIG 01/04) randomised open-label phase 3 trial. The Lancet Oncology 2014;15(9):997-1006.

Parallel Sessions | Monday 28 August 2023

MO12.5 Outlier detection in clinical performance monitoring and comparison of commonly used methods

Sui A.*, Pavlou M., Omar R., Ambler G.
Department of Statistical Science, University College London ~ London ~ United Kingdom

Monitoring of the clinical performance of health-care units is nowadays an essential aspect of national audits. Interest often lies in units whose performance (e.g. in-hospital mortality) diverges substantially from the expected performance. A commonly used method is the common mean model (CMM), which assumes a single true performance for all units and any variability in the observed performance is due to chance. This is often not a tenable assumption, e.g. due to clustering within units, because the observed variability is often higher than that assumed (i.e. overdispersion). An overdispersion correction can be applied; ideally this should be applied after the removal of potential outliers, otherwise the allowable variability will be too high. Winsorization, which shrinks some percentage of the most outlying units, is a recommended approach to minimise the effect of potential outliers [1], but the choice of the winsorization percentage can be challenging. This study aims to investigate the effect of the overdispersion correction and different winsorization percentages on the ability of the CMM to detect outliers. Simulation is used to assess the performance of the approaches under scenarios with different intra-class correlation (ICC), unit size, and percentage of outliers .
A higher ICC corresponds to a stronger violation of the CMM assumption. Performance measures include false positive rate (FPR) (proportion of non-outliers incorrectly identified as outliers) and sensitivity (proportion of outliers correctly identified).
The FPR of the uncorrected CMM is too high, indicating overdispersion correction is necessary. When there are outliers, winsorization is necessary, otherwise the FPR and sensitivity tend to be low. The FPR and sensitivity are sensitive to winsorization percentage, which is hard to choose as the simulation results show that it depends on the unknown percentage of outliers. Motivated by the simulations results, we investigate a new overdispersion correction method that is less affected by the percentage of outliers in the data than winsorization. It shows better FPR and relatively high sensitivity than the other methods.
Overdispersion correction is necessary when applying the CMM. When outliers are present, winsorization could be applied with an appropriate winsorization percentage. Alternatively, a new overdispersion correction can be used.

[1] Spiegelhalter DJ. Funnel plots for comparing institutional performance. Statistics in medicinez2005 Apr 30;24(8):1185-202.

PARALLEL SESSION MO13: CLINICAL TRIALS 3

MO13.1 Elastic priors to dynamically borrow information from historical data in clinical trials

Yuan Y.*[1], Jiang L.[2], Nie L.[3]
[1]The University of Texas MD Anderson Cancer Center ~ Houston ~ United States of America, [2]China Pharmaceutical University ~ Nanjing ~ China, [3]US FDA ~ Silver Spring ~ United States of America

Use of historical data and real-world evidence holds great potential to improve the efficiency of clinical trials. One major challenge is to effectively borrow information from historical data while maintaining a reasonable type I error and minimal bias. We propose the elastic prior approach to address this challenge. Unlike existing approaches, this approach proactively controls the behavior of information borrowing and type I errors by incorporating a well-known concept of clinically significant difference through an elastic function, defined as a monotonic function of a congruence measure between historical data and trial data. The elastic function is constructed to satisfy a set of prespecified criteria such that the resulting prior will strongly borrow information when historical and trial data are congruent, but refrain from information borrowing when historical and trial data are incongruent.
We show that the elastic prior approach has a desirable property of being information borrowing consistent, i.e. asymptotically controls type I error at the nominal value, no matter that historical data are congruent or not to the trial data. Our simulation study that evaluates the finite sample characteristic confirms that, compared to existing methods, the elastic prior has better type I error control and yields competitive or higher power. The proposed approach is applicable to binary, continuous and survival endpoints.
Jiang, L., Nie, L., Yuan, L. (2021) Elastic priors to dynamically borrow information from historical data in clinical trials, Biometrics, https://doi.org/10.1111/biom.13551

MO13.2 Dynamic borrowing of heterogeneous historical controls: how to avoid cherry picking?

Gerard F.*[1], Minini P.[1], Zhang B.[2]
[1]Sanofi ~ Chilly-Mazarin ~ France, [2]Sanofi ~ Bridgewater ~ United States of America

Standard phase 3 trials aim to evaluate the effectiveness of a new treatment against a comparator (standard of care or placebo). Several drugs may have been under development for the same disease leading to a potential large source of control arm data from different clinical trials. For example, the PRO-ACT Database contains Amyotrophic Lateral Sclerosis data from 17 clinical trials. However, when multiple historical trials are available, many factors can lead to heterogeneous data, such as trial duration or geographic location[1]. Incorporating historical control data when designing a new trial is an appealing approach, especially for rare diseases, to reduce risks, costs, and time to approval. Various statistical approaches have been proposed for data borrowing from multiple sources in the Bayesian framework and have been compared for binary outcomes[2]. In this work, we evaluate several Bayesian dynamic data borrowing approaches with various levels of heterogeneity in historical sources and provide recommendations for practical use. We evaluate methods that can account for prior/data conflict: dynamic power prior approaches with various algorithms for power parameter calculation and the robust meta-analytic predictive (rMAP) prior.
We perform simulations in the context of a 2-arm trial for normal endpoints when historical data can be borrowed for the control arm and assess various levels of historical data heterogeneity. In case of small heterogeneity, power prior approaches can increase the power while controlling type I error inflation and may be an alternative to rMAP approaches when few historical trials are available. However, in case of high heterogeneity, dynamic power prior approaches with study-specific power parameter can lead to type I error inflation over a wide range of prior-data conflict while rMAP approaches have a better control of type I error.
Power prior approaches should be used with caution in case of high historical data heterogeneity. In any case, the assumption of consistency between each historical data and current data should be checked (Pocock's criteria for example) to ensure the acceptability of each historical data and simulations must be performed when designing a new trial with historical data borrowing to find a suitable approach with desirable operating characteristics.

[1] K. T. Hall, L. Vase, D. K. Tobias, Clinical Pharmacology & Therapeutics, 109(2), 2021, 343-351.
[2] I. Gravestock, L. Held, Biometrical Journal, 61(5), 2019, 1201-1218.

MO13.3 Incorporating historical data in the design and analysis of small population clinical trials

Zheng H.*[1], Jaki T.[1], Wason J.[2]
[1]University of Cambridge ~ Cambridge ~ United Kingdom, [2]Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

Many disease conditions are so rare that the target population has only less than a hundred patients. Design and analysis of clinical trials in rare diseases are challenging, because it is clearly infeasible to enrol enough patients to achieve an adequate level of the frequentist power. Regulatory guidance (EMA, 2006) on small population clinical trials advises that alternative methods, including those in the Bayesian paradigm, may be considered to improve the analysis and interpretability of the trial results.
We develop a fully Bayesian approach to the design and analysis of small population clinical trials, where relevant historical information from multiple sources can be incorporated in a robust prior (Zheng et al., 2022). This approach takes account of pairwise (in)commensurability between parameters that underpin the historical and new clinical trials, when implementing the information borrowing. Closed-form sample size formulae are derived to ensure that the new trial has a specified chance of correctly deciding whether a new treatment is superior to or not better than the control by some clinically relevant difference. The application of our sample size formulae is illustrated by revisiting the MYPAN trial (Hampson et al., 2014; 2015), a randomised controlled trial conducted in rare and severe inflammatory blood vessel disease. The proposed methodology allows the inherent uncertainty in the estimate of model parameters and a formal incorporation of any relevant historical data. Sample size saving is possible by incorporating consistent historical data. This greatly enhances the planning of rare-disease clinical trials.

[1] EMA (2006) Guideline on clinical trials in small populations. European Medicine Agency: London UK.
[2] Zheng H, Jaki T, Wason JMS. (2022) Bayesian sample size determination using commensurate priors to leverage preexperimental data. Biometrics (Methodology), 1– 15. doi:10.1111/biom.13649
[3] Hampson LV, Whitehead J, Eleftheriou D, Brogan P. (2014) Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Statistics in Medicine, 33, 4186–4201.
[4] Hampson LV, Whitehead J, Eleftheriou D and et al. (2015) Elicitation of expert prior opinion: application to the MYPAN trial in childhood polyarteritis nodosa. PLOS ONE, 10, 1–14.

MO13.4 A bayesian sample size calculation using functional mixture weights to incorporate historical data

Whitehead L.*[1], Zheng H.[2], Wason J.[1]
[1]Newcastle University ~ Newcastle upon Tyne ~ United Kingdom, [2]MRC Biostatistics Unit, Cambridge University ~ Cambridge ~ United Kingdom

Relevant information from previous studies is often available and may sometimes be desirable to incorporate in the design of a new clinical trial. The Bayesian paradigm provides a coherent framework to quantify prior beliefs (informed by historical data) and formally incorporate them in the inference. Many established methods involve the use of a certain weight metric to allow flexible incorporation of historical data. However, it can be the case that a particular interval of the weight parameter (e.g., 0 to 0.3) incurs faster changes in the resulting sample size than its complement (e.g., 0.3 to 1). This hinders communication with subject-matter experts to elicit a sensible value of the weight for borrowing strength at the trial design stage. Focusing on a sample size formula that is nonlinear in its prior mixture weights [1], we propose a solution such that the sample size changes evenly over weights on the interval (0, 1). We consider a closed-form Bayesian sample size formula chosen to ensure that the planned trial has a specified chance of correctly deciding whether a new treatment is superior to, or not better than, control by some clinically meaningful magnitude. Pre-experimental information can be incorporated via specification of a commensurability weight between a historical source and the new study. We take inspiration from functional uniform prior methodology [2] to transform the weight such that it produces 'uniform' behaviour with respect to sample size. Essentially, we divide the sample size functional space evenly across weights using its quantiles (ranging from the 0% quantile with full incorporation of a historical data source, to the 100% quantile with no borrowing). Our approach leads to interpretable weights that represent (in)commensurability of data between a historical and planned trial, and could potentially facilitate elicitation of expert opinion on values for those weights.
The inclusion of historical data in the design of clinical trials is not yet common practice. Part of the reason might be difficulty in the interpretability of weight parameters resulting from such inclusion. We hope our work will support the uptake of innovative methods that facilitate robust incorporation of existing knowledge.

[1] H. Zheng, T. Jaki, and J. M. S. Wason, "Bayesian sample size determination using commensurate priors to leverage preexperimental data," Biometrics, vol. n/a, no. n/a, Mar. 2022, doi: 10.1111/biom.13649.
[2] B. Bornkamp, "Functional Uniform Priors for Nonlinear Modeling," Biometrics, vol. 68, no. 3, pp. 893–901, Sep. 2012, doi: 10.1111/j.1541-0420.2012.01747.x.

MO13.5 Extrapolation in pediatrics using bayesian dynamic borrowing, tipping point analysis and expert elicitation

Stock C.*, Dreher M., Erhardt E., Müller H., Sailer M.O., Voss F.
Global Biostatistics and Data Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG ~ Ingelheim am Rhein ~Germany

Pediatric drug development is commonly associated with considerable challenges regarding the creation of robust evidence on clinical efficacy. Extrapolation from existing trials in adults is an increasingly considered option. We propose a Bayesian dynamic borrowing framework to pediatric extrapolation for a continuous endpoint. It is based on a robust meta-analytic predictive (MAP) prior, tipping point analysis and expert elicitation. The tipping point analysis, as proposed by Best et al. [1] and slightly extended here, indicates, for given results of the pediatric trial and given one-sided evidence levels, how much weight on the informative component of the robust MAP prior is needed in order to conclude that the treatment is efficacious. At the planning stage, in addition to common criteria such as operating characteristics, we use the tipping point analysis as a tool to pre-specify the prior distribution. This is achieved through a formal expert elicitation exercise in which the experts are asked about inferences they would draw from the total evidence in different hypothetical scenarios and, consequently, the weights they would assign to the evidence from trials in adults. Once the data from the pediatric trial are available, the tipping point analysis serves as a sensitivity analysis to assess the impact of the chosen weight on the inferences. We illustrate the approach by an exemplary case study. Further, we discuss compatibility with new draft ICH guidance on pediatric extrapolation [2]. The publicly available R package "tipmap" is introduced which facilitates implementation of the described approach. The proposed framework may be considered in pediatric drug development, in particular when it is of interest to pre-specify a weight of the evidence from trials adults in a dynamic borrowing approach.

[1] N. Best, R.G. Price, I.J. Pouliquen, O.N. Keene. Pharm Stat, 20(3), 2021, 551-562.
[2] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Draft ICH guideline E11A on pediatric extrapolation Step 2b. 2022, EMA/CHMP/ICH/205218/2022.

Parallel Sessions | Monday 28 August 2023

PARALLEL SESSION MO14: SURVIVAL ANALYSIS 3

MO14.1 Clinical trial design based on a multistate model that jointly models progression-free and overall survival

Erdmann A.[1], Beyersmann J.[1], Rufibach K.*[2]
[1]Institute of Statistics, Ulm University ~ Ulm ~ Germany, [2]Methods, Collaboration and Outreach Group, Product

Development Data Sciences, F. Hoffmann-La Roche Ltd ~ Basel ~ Switzerland When planning a pivotal oncology clinical trial, the usual approach is to assume an exponential distribution for the time-to-event endpoints. Often, besides the gold-standard endpoint overall survival, progression-free survival is considered as a second confirmatory endpoint.
We use a survival multistate model to jointly model these two endpoints and find that neither exponential distribution nor proportional hazards will typically hold for both endpoints simultaneously. The multistate model approach allows us to consider the joint distribution of the two endpoints and to derive quantities of interest, such as the correlation between overall survival and progression-free survival. In this paper, we use the multistate model framework to simulate clinical trials with endpoints OS and PFS and show how design planning questions can be answered using this approach. In addition to the major advantage that we can model nonproportional hazards quite naturally with this approach, the correlation between the two endpoints can be exploited to determine sample size and type-I-error. We consider an oncology trial on non-small-cell lung cancer as motivating example from which we derive relevant trial design questions. We then illustrate how clinical trial design can be based on simulations from a multistate model. Key applications are co-primary endpoints and group-sequential designs. Simulations for these applications show that the standard simplifying approach often leads to underpowered or overpowered clinical trials. Our approach is quite general and can be extended to more complex trial designs, further endpoints, and other therapeutic areas. Erdmann, A., Beyersmann, J., Rufibach, K. (2023). Oncology clinical trial design planning based on a multistate model that jointly models progression-free and overall survival endpoints.

Submitted. https://arxiv.org/abs/2301.10059 Erdmann A., Rufibach K. (2023). simIDM: Simulating Clinical Trials with Endpoints Progression- Free Survival and Overall Survival using an Illness-Death Model. R package version 0.0.3, https://cran.r-project.org/package=simIDM

MO14.2 Modelling the hazard of transition into the absorbing state in the illness-death model

Tassistro E.*, Bernasconi D.P., Rebora P., Antolini L., Valsecchi M.G.
University of Milano-Bicocca ~ Monza ~ Italy

The illness-death model is the simplest multistate model where the transition from an initial state 0 to an absorbing state 2 may involve also an intermediate state 1. The impact of the transition into 1 on the subsequent transition hazard to 2 enables to increase the knowledge about the disease evolution. The standard analysis approach is modelling the transition hazards from 0 to 2 and from 1 to 2 including time to illness as a time-varying covariate and measuring time from origin even after the transition into 1. The hazard from 1 to 2 can be also modelled only on patients in state 1, measuring time from illness and including time to illness as a fixed covariate. A recently proposed approach is a model where time after the transition into 1 is measured in both scales and time to illness is included as a time-varying covariate. Another possibility is a model where time after the transition into 1 is measured only from illness and time to illness is included as a fixed covariate. This work aims to set up a strategy a statistician can follow to fit the most suitable full-sample model on the hazards of transition to state 2. Through theoretical reasoning and simulation protocols we developed sequential strategies a statistician can follow to: a) validate the properties of the illness-death process, from which the choice of the scale to measure time after illness depends, b) estimate the impact of time to illness on the hazard from 1 to 2, proposing also a novel modelling approach that ensures the interpretability of the coefficient of the time to illness. In the case of Markov data, the use of the clock forward time scale is the natural way to measure the follow-up time. The clock reset scale should be considered in case of non-Markov data, since forcing to use the clock forward scale will result in a spurious effect of the time to illness, due to the time after illness and not to a different shape of the hazard function after illness.

[1] S. Iacobelli, B. Carstensen (2013). Multiple time scales in multi-state models. Statistics in Medicine, 32, 5315-27
[2] E. Tassistro, D.P. Bernasconi, P. Rebora, M.G. Valsecchi, L. Antolini (2020). Modeling the hazard of transition into the absorbing state in the illness-death model. Biometrical Journal, 62, 836-851 MO14 .2_1 55

Parallel Sessions | Monday 28 August 2023

MO14.3 Bayesian blockwise inference for joint models of longitudinal and multistate processes

Chen S.*, Alvares D., Jackson C., Barrett J.
University of Cambridge ~ Cambridge ~ United Kingdom

Joint models for longitudinal and survival data have gained increasing interest in clinical research [1]. These models characterize the association between the outcome processes and can improve individualized predictions. In many application contexts, more complicated event processes arise, necessitating the use of joint longitudinal and multistate models that link longitudinal and transition-specific regression models. Inference of such models typically relies on the Bayesian paradigm as it allows incorporating prior knowledge, coherent uncertainty quantification and dynamic prediction, through Markov chain Monte Carlo posterior sampling. However, the computational challenges arising from increased model complexity and large sample sizes can limit the application of such models. Motivated by longitudinal multimorbidity analysis of large UK health records, we aim to develop a scalable Bayesian methodology that can handle complex event processes and large datasets, with easy implementation.
We propose two blockwise Bayesian methods for inference in the joint longitudinal and multistate models, leveraging parallel computing and the state-of-the-art sampling technique. The first approach employs competing risk decompositions of the multistate process, estimating the joint model in a blockwise manner. When focusing specifically on the association and other multistate parameters, our second block inference strategy offers further efficiency gains by utilizing transition-specific posteriors. Blockwise approaches facilitate the specification of different models for different transitions, and model/variable selection can be performed in a Bayesian framework using the Bayesian leave-one-out cross-validation. Using a simulation study, we show that the proposed approaches achieve satisfactory performance regarding posterior point and interval estimation, with notable gains in sampling efficiency over the standard estimation strategy. We illustrate our approaches using a large UK electronic health record dataset where we analysed the coevolution of routinely measured systolic blood pressure (SBP) and the progression of multimorbidity, defined as the combinations of three chronic conditions. We identified differing association structures between SBP and different disease transitions. The proposed blockwise approaches offer computational efficiency gains over the standard inference method while preserving accurate estimation and allowing easier implementation. Additionally, they inherently improve robustness against longitudinal model misspecification. Our real data application highlights their practical feasibility and scalability.

[1] Rizopoulos, Dimitris. Joint models for longitudinal and time-to-event data: With applications in R. CRC press, 2012.

MO14.4 Model selection strategies for multi-state modeling incorporating molecular data

Miah K.*[2], Goeman J.J.[1], Putter H.[1], Kopp--Schneider A.[2], Benner A.[2]
[1]*Department of Biomedical Data Sciences, Leiden University Medical Center (LUMC) ~ Leiden ~ Netherlands,* [2]*Division of Biostatistics, German Cancer Research Center (DKFZ) ~ Heidelberg ~ Germany*

In the era of precision medicine with increasing molecular information, the use of a multi-state model is essential to more accurately capture the individual disease pathway along with underlying etiologies. Especially the availability of big data with a large number of covariates presents several statistical challenges for model building. Effective data-driven model selection strategies for multi-state models are required to determine an optimal, ideally parsimonious model based on high-dimensional data. Established methods incorporate regularization in the fitting process in order to perform variable selection. In the multi-state framework, linking penalties across transitions is required to conduct joint variable selection. A useful technique to reduce model complexity is to combine homogeneous covariate effects for distinct transitions based on a reparametrized model formulation. We integrate this approach to data-driven variable selection by extended regularization methods for model selection within multi-state model building. We propose the sparse group fused lasso penalized Cox-type regression in the framework of multi-state models combining the penalization concepts of pairwise differences of covariate effects along with transition grouping. For optimization, we adapt the alternating direction method of multipliers (ADMM) algorithm to cause-specific proportional hazards regression in the multi-state setting. This raises the following challenges: First, multiple heterogeneous transitions have to be considered for consecutive treatment phases within the multi-state model. Furthermore, the number of transitions with fewer observations increases during the trajectory of the sequential event history. Finally, optimization algorithms have to be efficiently implemented in large-scale multi-state settings. As a consequence, model selection strategies for multi-state survival analysis are substantial for a more precise understanding and interpretation of individual disease pathways, specific oncological entities along with corresponding precision therapies as well as improved personalized prognoses.

PARALLEL SESSION MO15: PRECISION MEDICINE 1

MO15.1 A bayesian nonparametric approach to personalized treatment selection

Pedone M.*[1], Argiento R.[2], Stingo F.C.[1]
[1]*University of Florence ~ Florence ~ Italy,* [2]*University of Bergamo ~ Bergamo ~ Italy*

Precision medicine is an approach to disease treatment that defines treatment strategies based on the individual characteristics of the patients. We develop a predictive model that flexibly clusters patients with similar genomic profiles and identifies -via predictive inference- which one among a set of therapeutic strategies is better suited for a new untreated patient. We propose a Bayesian predictive model for personalized treatment selection that integrates prognostic and predictive biomarkers. We use prognostic biomarkers to define a baseline response probability across treatments. Since patients should not be regarded as statistically exchangeable with respect to predictive biomarkers, we leverage them to drive a clustering process among patients that received the same treatment. A cluster-specific random intercept estimates the adjustment provided by predictive biomarkers to the baseline prognostic response probability on account of groups of patients with close predictive determinants that received the same treatment. In particular, we use a product partition model with covariates (PPMx) [2], here extended to include the cohesion induced by the Normalized Generalized Gamma process. PPMx clusters observations with close values of the predictive covariates. This approach allows predicting the utility offered by each competing treatment to the new untreated patient, borrowing strength from past patients with whom the new patient shows the larger similarity in terms of predictive biomarkers. The proposed method jointly performs clustering and prediction, so that the prediction fully accounts for the uncertainty in the clustering. This represents a more rigorous and unified take on the problem with respect to alternative methods based on two-step procedures and heuristic clustering algorithms [1]. A cancer genomics case study on Lower-grade Glioma (LGG) data shows that the proposed method is well suited for predictions in scenarios with considerable heterogeneity. Moreover, our approach leads to a precise cluster characterization, identifying patients more likely to benefit from targeted treatment. The proposed Bayesian predictive model for personalized treatment selection allows for accurate prediction in heterogeneous scenarios and precise cluster characterization of patients affected by LGG tumors.

[1] J. Ma, F. Stingo, and B. Hobbs. "Bayesian personalized treatment selection strategies that integrate predictive with prognostic determinants". In: Biometrical Journal 61.4 (2019), pp. 902– 917.
[2] P. Müller, F. Quintana, and G. Rosner. "A product partition model with regression on covariates". In: Journal of Computational and Graphical Statistics 20.1 (2011), pp. 260–278.

Parallel Sessions | Monday 28 August 2023

MO15.2 Precision medicine in type 2 diabetes: bayesian non-parametric modelling of glucose-lowering therapy efficacy

Cardoso P.*[1], Young K.[1], Hopkins R.[1], Jones A.[1], Pearson E.[2], Hattersley A.[1], Shields B.[1], Bowden J.[1], Mckinley T.[1], Dennis J.[1]
[1]University of Exeter ~ Exeter ~ United Kingdom, [2]University of Dundee ~ Dundee ~ United Kingdom

Choosing between SGLT2-inhibitors (SGLT2i) and GLP1-receptor agonists (GLP1-RA) requires understanding each drug class's relative benefits and risks for individual patients. We aim to establish whether routine clinical characteristics alter HbA1c response to these agents and could help inform optimal therapy choices for individual patients.
We analysed UK primary care data (CPRD Aurum) from 31,346 patients with type 2 diabetes initiating SGLT2i or GLP1-RA therapies. State-of-the-art, non-parametric Bayesian Causal Forests (BCF) [1] were used to develop a model to predict individual-level differences in 12- month HbA1c response with the two therapies based on routine clinical characteristics. BCF is a decision tree-based modelling technique that splits into two components: the prognostic effect on the control treatment response and the moderating effect corresponding to the treatment effect heterogeneity of the competing treatment. The model was externally validated in hold- back CPRD data (20,865 patients). There is a significant challenge when developing models with observational data where the patient only initiated one of the treatments. We approach this problem by splitting the population into subgroups based on their predicted conditional average treatment effect (CATE) estimates and comparing them to the average treatment effect (ATE) within each subgroup. Routine clinical characteristics, in particular sex, showed marked heterogeneities in glycaemic response to both therapies, with 2,193 (7%) individuals predicted a >5mmol/mol benefit on SGLT2i over GLP1-RA (denoted the SGLT2-optimal subgroup). In contrast, 2,056 (7%) individuals had a predicted >5mmol/mol benefit on GLP1-RA over SGLT2i (denoted the GLP1-optimal subgroup). Compared to GLP1-optimal patients, SGLT2-optimal patients were more likely to be male (73% versus 28%) and younger, with higher baseline HbA1c and eGFR measurements. The model validated well, with average treatment benefits (versus alternative treatment) of 8.1 mmol/mol with SGLT2i in the SGLT2-optimal subgroup and 8.2 mmol/mol with GLP1-RA in the GLP1-optimal subgroup. BCF offers a flexible method to build treatment selection models that can predict individualised treatment effects. Our novel framework provides a principled approach to evaluating the calibration of predicted treatment effects at the subgroup level. The model shows clear potential for targeting type 2 diabetes therapies based on differences in important clinical outcomes.

Chung WK, Erion K, Florez JC, et al.. Precision medicine in diabetes: a Consensus Report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2020; 63(9): 1671-1693.
Dennis JM, Young KG, McGovern AP. Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes: a retrospective cohort study. The Lancet. 2022; 4(12): 873-883.
Hahn PR, Murray JS, Carvalho CM. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion). Bayesian Analysis. 2020; 15(3): 965-1056.

MO15.3 Estimating optimal rules for personalized treatment decisions through functional survival analysis

Gregorio C.*[1], Barbati G.[2], Ieva F.[1]
[1]Politecnico di Milano ~ Milano ~ Italy, [2]University of Trieste ~ Trieste ~ Italy

There has been growing research interest regarding personalized treatments. However, treatment effect heterogeneity and possibly time-varying nature of treatment effects are two aspects still often overlooked in clinical studies. The main challenge is providing statistical tools to estimate optimal rules able to identify for which subjects different treatment decisions should be made. We aim at providing an innovative method to obtain personalized treatment recommendations in a time-to-event framework, taking into account a set of relevant covariates. The proposed method does not require the assumption of proportional hazards for the treatment effect, which is rarely realistic. The approach is based on the novel idea of exploiting the functional nature of spline-based survival models [1] to 1) estimate time-varying conditional treatment effects 2) use functional clustering [2] of the treatment-effect curves to identify decision rules. The application that motivated this work is the discontinuation of treatment with MRAs in Heart Failure patients. In this setting, there is no clear evidence regarding the patient's profile-specific trade-off between side effects and therapy benefits on the risk of hospitalization and death. Data comes from an observational retrospective study involving 1244 patients. Customized recommendations for the discontinuation of MRAs were obtained according to age, hyperkalemia, diabetes, chronic kidney disease, New York Heart Association class and left- ventricular ejection fraction. Furthermore, the analysis was adjusted for possible other confounding variables through inverse probability of treatment weights. Finally, a simulation study performed to assess the ability of the proposed method in identifying the decision rules under different scenarios, achieved good accuracy. We propose a method to identify treatment decision rules based on a set of covariates of interest in order to minimize the risk of adverse events. The developed methodology has proven successful in simulated data. Moreover, it provides novel insights regarding a real-world application concerning the pharmacological treatment of Heart Failure.

[1] Royston, P., & Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine, 21(15), 2175–2197.
[2] Tarpey, T. and Kinateder, K. K. J. (2003) Clustering functional data. Journal of Classification, 20, 93–114.

MO15.4 Causal effects of salvage therapy using joint models for longitudinal and time-to-event data

Dimitris R.*[1], Jeremy T.[2]
[1]Erasmus University Medical Center ~ Rottedam ~ Netherlands, [2]University of Michigan ~ Ann Arbor ~ United States of America

Prostate cancer patients who undergo prostatectomy are closely monitored for recurrence and metastasis using routine prostate-specific antigen (PSA) measurements. When PSA levels rise, salvage therapies are recommended in order to decrease the risk of metastasis. However, due to the side effects of these therapies and to avoid over-treatment, it is important to understand which patients and when to initiate these salvage therapies.
In this work, we use the University of Michigan Prostatectomy registry Data to tackle this question. Due to the observational nature of this data, we face the challenge that PSA is simultaneously a time-varying confounder and an intermediate variable for salvage therapy. We define different causal salvage therapy effects defined conditionally on different specifications of the longitudinal PSA history. We then illustrate how these effects can be estimated using the framework of joint models for longitudinal and time-to-event data. All proposed methodology is implemented in the freely-available R package JMbayes2.
Joint models account for time-varying confounding without requiring an explicit specification of a model for the probability of receiving treatment conditional on the history of longitudinal confounders and past treatments. The derived causal effects are in the flavor of the parametric G-formula and, by conditioning on different specifications of the longitudinal PSA history correspond to different targets of inference.
Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Chapman & Hall/CRC, Boca Raton

Parallel Sessions | Monday 28 August 2023

PARALLEL SESSION MO16: LONGITUDINAL ANALYSIS 1

MO16.1 Combined shrinkage of fixed and random effects in linear mixed models using empirical bayes

Amestoy M.*, Van De Wiel M., Van Wieringen W.
Amsterdam Umc ~ Amsterdam ~ Netherlands

Linear mixed models are widely employed by researchers as a fundamental tool for analyzing longitudinal data, when dealing with continuous outcomes. Nonetheless, practitioners often encounter limitations when attempting to fit models of considerable complexity. This is because conventional solvers are incapable of processing high-dimensional data and/or complex random effect structures, which require regularization. Bayesian solvers, on the other hand, are able to handle such challenges, but necessitate informed choices for the prior parameters. This can prove problematic, particularly in situations where adequate information is not readily available, such as with the covariance of the random effects. In light of this, we propose a novel data-driven method for the joint estimation of the fixed and random effect priors' parameters. Our proposed empirical Bayes method maximizes the model's marginal likelihood, which is computed efficiently by a Laplace approximation. To evaluate our method, we conducted extensive simulations and compared it against standard solvers. Our results indicate that our method significantly improves the accuracy of parameter estimates and enhances the prediction power of the model. Furthermore, we applied our method to a real-world air pollution data set, where we found that employing a more complex model, combined with regularization, led to improvement in prediction accuracy. In summary, data-driven regularization by our empirical Bayes methods is a valuable approach for estimating complex linear mixed models. This methodology enables the use of more comprehensive models while simultaneously improving parameter inference and predictive performance of the model.

MO16.2 A bayesian functional principal component analysis framework for longitudinal genome-wide association studies

Temko D.*, Nolan T., Richardson S., Ruffieux H.
MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom

When interrogated along with genetic data, large clinical longitudinal datasets have potential to yield new scientific insights on the genetic contribution to temporal patterns of disease progression and response to treatments. However, existing approaches for longitudinal genome-wide association studies (GWAS) are not suited to unlock the full power of large datasets, as they typically rely on strong distributional assumptions on the longitudinal outcome.
We propose to reframe the longitudinal GWAS problem as a joint latent variable estimation and regression problem in which genetic variants influence longitudinal trajectories via effects on functional latent variables. We leverage existing Bayesian inference frameworks for functional principal component analysis (FPCA)[1] and sparse spike-and-slab regression[2] to develop a two-stage variational inference scheme for this model that conveys uncertainty from the principal component estimation into the second-stage regression. Using simulations, we show that our approach is both scalable and accurate, and that our modelling approach can recover SNPs that influence latent dynamics underlying longitudinal trajectories. Moreover, the uncertainty estimates provided by our approach are better calibrated than those from a naïve implementation that does not take account of uncertainty in the first- stage inference. We further demonstrate the usefulness and applicability of our framework in a study of genetic effects on longitudinal outcomes in the UK biobank. We present a modelling scheme that has the computational scalability and flexibility to take advantage of large datasets for longitudinal modelling, while rigorously handling estimation uncertainty.

[1] T. Nolan, J. Goldsmith, D. Ruppert, arXiv, 2021, doi:10.48550/ARXIV.2104.00645, https://arxiv.org/abs/2104.00645
[2] H. Ruffieux, J. Carayol, R. Popescu, M. Harper, R. Dent, W. Saris, A. Astrup, J. Hager, A. Davison, A. Valsesia, PloS Comput Biol., 2020, 16(6): e1007882

MO16.3 A bayesian model to study the genetic risks driving alzheimer's disease progression patterns

Fournier N.*, Durrleman S.
Paris Brain Institute ~ Paris ~ France

Alzheimer's Disease (AD) has been subject to multiple Genome-Wide Association Studies (GWAS) to identify its risk factors driven by genetics. Multiple loci strongly associated with AD diagnosis have been identified. Yet for most of these loci, it is unclear how they relate to the disease evolution over its full course. We propose a method to characterize the effect of selected genetic mutations on disease development by integrating and leveraging genetic information into a Disease Progression Model (DPM).
We developed and validated a Bayesian non-linear mixed-effect DPM to model multimodal clinical trajectories and capture the influence of mutations on the evolution of the disease, which we adapted from a state-of-the-art DPM: the Disease Course Mapping [1]. Interpretable effects of these genetic mutations on the progression (such as their influence speed of progression of each monitored feature, influence on the onset time of the disease, etc) can be estimated. The Bayesian formulation of the model allows for estimating posterior credible intervals for each of the aforementioned effects.
We applied this model to the Alzheimer's Disease Neuroimaging Cohort (ADNI), on a combination of both longitudinal structural and PET imaging biomarkers, as well as cognitive scores of subjects. We selected Single Nucleotide Polymorphisms (SNP) whose association with AD diagnosis has been reliably established in recently published GWAS [2] and included them in our modelling.
Our results suggest that multiple profiles of risk-increasing or protecting SNP coexist. First, we show that we can provide novel strong evidence supporting the impact of commonly known risk factors for AD diagnosis, such as mutations on the APOE gene, onto the set of features examined, and disentangle its effect on feature subsets (protein loads / structural atrophy / cognitive functioning). Secondly, we examine a wide array of SNPs selected from a GWAS and show that, despite all being associated with AD diagnosis, they have very diverse involvement in the disease progression patterns.
We propose a model that can shine a new light on AD genetic risk factors through their influence on the disease dynamics. This refines GWAS performed on case-control cohorts and could provide a better understanding of the disease.

[1] Schiratti, J. B., Allassonnière, S., Colliot, O., & Durrleman, S. (2017). A Bayesian mixed- effects model to learn trajectories of changes from repeated manifold-valued observations. The Journal of Machine Learning Research, 18(1), 4840-4872.
[2] Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., ... & Rotter, J. I. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature genetics, 51(3), 414-430.

Parallel Sessions | Monday 28 August 2023

MO16.4 Distributional models for the quantification of within-individual lung function variability in cystic fibrosis

Palma M.*[2], Keogh R.H.[3], Wood A.[4], Muniz--Terrera G.[1], Barrett J.K.[2]
[1]University of Edinburgh ~ Edinburgh ~ United Kingdom, [2]MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom, [3]Department of Medical Statistics, London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom, [4]Cardiovascular Epidemiology Unit, University of Cambridge ~ Cambridge ~ United Kingdom

In multiple biomedical fields, there is an increasing interest in quantifying within-individual variability of health indicators measured over time, such as for example blood pressure and lung function, to inform about disease progression. Simple summary statistics (such as the within- individual standard deviation) are often used but they do not account for different sample sizes and number of measurements per individual, and do not exploit the longitudinal nature of the data. We present a broad class of models to estimate variability at the individual level and show an application on cystic fibrosis data. Mixed-effects location-scale models (MELSM, [1]) extend the linear mixed model framework by specifying a new submodel for the standard deviation of the repeated measures for each individual using covariates and random effects. We draw a connection between MELSM and distributional regression models, in particular Bayesian additive models for location, scale and shape (BAMLSS, [2]), based on the specification of random effects in the submodel for the scale parameter. We explore the performance of Bayesian MELSM and bamlss in terms of posterior estimates and computational efficiency, using a cohort of adult patients from the UK cystic fibrosis registry. The evolution of FEV1 (a measure of lung capacity) mean and within-individual variability over time is modelled as a function of sex, age at each visit and age at diagnosis, genotype and birth cohort. Mean lung function decreased with age, while lung function variability showed a quadratic trend by age. FEV1 variability was also associated with sex, age at diagnosis and genotype. We also show the average trajectory of lung function variability for different subgroups in the UK CF adult population. In cystic fibrosis, key variables known to be linked with mean lung function in cystic fibrosis patients are also associated with a patient's lung function within-individual variability. Distributional regression models could bring new insights about the analysis of longitudinal data in other biomedical settings and the use of within-individual variability could help to predict disease progression as well as key outcomes such as mortality.

[1] Hedeker, D., Mermelstein, R. J., and Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics, 64(2):627–634
[2] Umlauf, N., Klein, N., and Zeileis, A. (2018). BAMLSS: Bayesian additive models for location, scale and shape (and beyond). Journal of Computational and Graphical Statistics, 27(3):612–627

MO16.5 A hierarchical modelling approach for principal components analysis on multiple longitudinal variables

Nolan T.*, Ruffieux H., Richardson S.
MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom

The increasing availability of longitudinal measurements presents opportunities to unlock scientific insights, benefitting biomedical research (data from wearables, electronic health records, etc.). In order to be fully exploited, this data paradigm calls for principled modelling approaches that can flexibly infer the latent dynamics underlying related longitudinal measurements. Functional principal components analysis (FPCA) is an important tool for identifying the major modes of variability in longitudinal measurements on a single variable, which normally provide the analyst with an understanding of the scientific properties captured in the dataset. Extending this approach to longitudinal measurements on multiple and correlated measurements is an important concept in biomedical research.
Here we present a Bayesian hierarchical modelling approach for performing joint FPCA on p > 1 longitudinal variables. Our approach treats all quantities from the Karhunen–Loeve expansion as unknown and estimates them jointly, using variational Bayesian inference. It models subject- specific scores that are shared across all p variables, which conveniently results in a parsimonious representation of the data, based on the main modes of joint variation of the variables. We construct a variational message passing algorithm that allows us to conveniently extend a Bayesian FPCA algorithm to the multivariate setting. Our numerical experiments show that our approach (i) is fast and accurate; (ii) improves estimation of related longitudinal variables; and (iii) is robust to model misspecification, with essentially no loss of performance when the variables are only weakly related. We use it to characterise recovery from SARS-CoV-2 infection, exploiting immune, metabolic and inflammatory measurements, collected longitudinally from infected individuals over one year post disease onset.
Our approach effectively captures the temporal covariation of related longitudinal variables and reconstructs their trajectories at the subject level; in our COVID-19 study, this helped clarify how different biological pathways coordinate the organismal response to infection over time and drive systemic recovery.

[1] J. O. Ramsay, B. W. Silverman, Functional Data Analysis, 2005, Springer, New York.
[2] C. Happ, S. Greven, Journal of the American Statistical Association, 113, 2018, 649-659.

Parallel Sessions | Monday 28 August 2023

PARALLEL SESSION MO17: HIGH DIMENSIONAL DATA 2

MO17.1 A bartlett-type correction for likelihood ratio tests for testing equality of gaussian graphical models

Banzato E.*[1], Chiogna M.[2], Djordjilovic V.[3], Risso D.[1]
[1]University of Padova ~ Padova ~ Italy, [2]University of Bologna ~ Bologna ~ Italy, [3]Ca' Foscari University of Venice ~ Venice ~ Italy

This work defines a new correction for the likelihood ratio test for a two-sample problem within the multivariate normal context. The corrected statistic finds a natural application in the context of decomposable Gaussian graphical models, where tests with different dimensions, p, need to be simultaneously handled. We proved that the phase transition boundary [1] is improved over some alternatives. This boundary characterizes the approximation accuracy by establishing the necessary and sufficient condition for the chi-square approximation to hold when p increases with the sample size n. In particular, we showed that the chi-square approximation of the adjusted statistic holds for $p/n \to 0$. Simulations confirmed that the adjusted test statistic is well approximated by a chi-square distribution both for small and large values of p. In the context of decomposable Gaussian graphical models, the problem of testing the equality of two models can be broken down into a sequence of problems defined on smaller sets of variables (cliques). Using the information on the graphical structure allows us both to improve the power of detecting a difference between the two distributions under study, and to localize that difference. Simulations showed that the size of the test is reached for different configurations of p and n and, in the presence of a difference in two conditions, the adjusted statistic is able to detect it, still controlling the type I error in the other cliques. The adjusted statistic leads to valid inference at different dimensionality regimes and overcomes some weaknesses that occur at small sample sizes and in particular when the dimension p is close to the sample size n. The use of this correction in the context of graphical models renders tractable inference in the setting of large-scale graphical models, where the dimension p is higher than the available sample size n. In real data analysis, this statistical tool is useful in detecting differences between networks of two conditions, such as biological or social networks.

[1] He, Y., Meng, B., Zeng, Z., Xu, G., 2021. On the phase transition of Wilks' phenomenon. Biometrika 108, 741–748.

MO17.2 Pearson's chi-squared meets distance and kernel tests: an application to complex disease genetics

Castro--Prado F.*[1], Gonzalez--Manteiga W.[2], Costas J.[3], Edelmann D.[4]
[1]University and Health Research Insitute ~ Santiago de Compostela ~ Spain, [2]University of Santiago de Compostela ~ Santiago de Compostela ~ Spain, [3]Health Research Institute ~ Santiago de Compostela ~ Spain, [4]German Cancer Research Centre ~ Heidelberg ~ Germany

Understanding epistasis (genetic interaction) may shed some light on the genomic basis of common diseases. It is standard practice among practitioners to assume additive models and then try to test for association between genetic variants. Many approaches have been attempted for this task, all with limited success [1]. Our goal is to present a more realistic statistical modelling of the effect of point mutations, by allowing them to take values on (semi)metric spaces. This will involve adapting general association measures, developed on the 21st century, to our geometry. We will then explore the relationships of the resulting tests to classical statistical methodology, from the beginning of the 20th century. We study single-nucleotide polymorphisms, for which study subject can carry 0, 1 or 2 copies of the "mutated" allele. Instead of treating this data as continuous (or even as ordinal), we allow it to dwell an arbitrary (semi)metric space. One can test for independence among such variables with distances (yielding an association measure called "distance covariance") or with kernels (via the Hilbert–Schmidt independence criterion), with both approaches being equivalent [2]. By applying such techniques to our discrete setting, we derive closed-form expressions for the asymptotic distribution of our (U-)statistics. We then apply this methodology to data from a genome-wide association study of schizophrenia collected by our group, finding an over- representation of putative interactions between genes that are expressed in brain tissue. Now we go back to our statistic and rewrite it in a form that is very similar (but not equal) to Pearson's chi-squared statistic. We use this expression, as well as simulations and the real data example, to show when and how the performance of our statistic and that of the classical one compare. General association measures provide meaningful insight into genetic architecture. These novel statistical tests perform well in practice and yield biologically sound results (at least in our data). This methodology gives similar results to Pearson's chi-squared test in some cases, but outperforms it in other settings.

[1] F. Castro-Prado, J. Costas, W. González-Manteiga, D. R. Penas, arXiv, 2012.05285.
[2] D. Edelmann, J. J. Goeman, Statistical Science, 37, 2022, 562–579.

MO17.3 Exploring between-subject consistency in fmri signals through partial conjunction null hypotheses

Vesely A.*[1], Heller R.[2], Dickhaus T.[1]
[1]University of Bremen ~ Bremen ~ Germany, [2]Tel Aviv University ~ Tel Aviv ~ Israel

Functional magnetic resonance imaging (fMRI) measures changes in blood flow and oxygenation levels in the brain under a sequence of stimuli. A typical study examines multiple subjects with the goal of inferring and mapping neural activity. However, results can be challenging to interpret due to functional misalignment, i.e., individual differences between participants that may partially linger even after suitable pre-processing. Such misalignment can be mainly attributed to the considerable variability in brain structure and organization, but it may also be exacerbated by differences in data acquisition (scanner settings, positioning, motion artifacts etc.). The aim of our study is identifying regions that exhibit consistent neural activation across participants.
To assess signal consistency, we exploit the partial conjunction (PC) null hypotheses framework. From standard pre-processing and first-level analysis, for each subject we obtain a map of p-values at different locations (voxels). Then we test, for each voxel and at any granularity γ, the PC null hypothesis that the voxel is active in less than γ subjects over the total. These hypotheses are tested simultaneously not only over all voxels but also for all possible granularities, with control of the false discovery rate (FDR). First, we apply the method introduced in [1] and [2]. Subsequently, we discuss a two-stage approach that alleviates the conservativeness of the previous one by eliminating the least promising PC null hypotheses. We explore the behavior of the methods on parametric and permutation-based p-values obtained from both real and simulated data.
Preliminary analyses on simulated data confirm that the considered methods control the FDR and suitably identify regions where signal is shared between subjects; moreover, the second approach is generally more powerful. We expect that results on real data will be in line with existing studies, finding activation in regions known to be related to the stimuli of interest. Ultimately, we argue that assessing the coherence of imaging signals across different individuals can enhance our understanding of neural activation.

[1] Y. Benjamini, R. Heller. Screening for Partial Conjunction Hypotheses. Biometrics 64, 2008, pp. 1215-22.
[2] R. Heller, Y. Golland, R. Malach, Y. Benjamini. Conjunction group analysis: An alternative to mixed/random effect analysis. NeuroImage 37, 2007, pp. 1178-85.

MO17.4 Towards a power analysis and sample size estimation for pls-based methods

Andreella A.*[1], Stocchero M.[2]
[1]Ca' Foscari University of Venice ~ Venice ~ Italy, [2]University of Padua ~ Padua ~ Italy

In recent years, power analysis has become widely used in several fields due to the increasing importance of replicability issues in the literature and scientific research in general. However, in the case of methods not based on statistical models, like the partial least squares (PLS) based approaches, the development of power analysis turns out to be challenging. Since these nonparametric methods are generally used to analyze correlated high-dimensional data, several factors must be considered when implementing a power analysis: the responses are correlated, redundant, and noisy, and the number of observations is generally smaller than the number of responses. We propose a potential power analysis in the context of PLS for classification [1]. The method simulates data under the alternative hypothesis considering the complex correlation structure typical of analysis with this distribution-free approach. The idea is to model the score components from the PLS to obtain the multivariate distribution of the scores under the alternative hypothesis. First, the optimal number of latent components is founded by permutation-based statistical tests under the null hypothesis of no dependence between predictors and class membership. The statistical tests proposed are based on the eigenvalues computing during the PLS model estimation, which, if significant, must be larger than the ones under random permutations of the residuals. The inference process proposed controls for family-wise error rate exploiting the sequential structure of the null hypothesis. Then, for each sample size analyzed, the power is calculated using a nonparametric statistical test based on the Matthews Correlation Coefficient. We also suggested a definition of the effect sizes based on the Matthews Correlation Coefficient and the predictive scores [2]. The users can then fix the values of the power and Type I error, and the proposed method returns the optimal sample size by controlling for several aspects, such as the complex multidimensional structure of the pilot data. Prospective and retrospective power analysis is evaluated by simulations considering different scenarios characterized by different values of effect sizes, within-cluster correlations, and within- cluster kurtosis. Finally, a retrospective power analysis was shown, analyzing metabolic data.

[1] Stocchero, M., De Nardi, M., & Scarpa, B. (2021). PLS for classification. Chemometrics and Intelligent Laboratory Systems, 216, 104374.
[2] Stocchero, M., & Paris, D. (2016). Post-transformation of PLS2 (ptPLS2) by orthogonal matrix: a new approach for generating predictive and orthogonal latent variables. Journal of Chemometrics, 30(5), 242-251.

Parallel Sessions | Monday 28 August 2023

MO17.5 Selective inference in factorial designs with high-dimensional response

Finos L.*
University of Padova ~ Padova ~ Italy

Factorial designs with high-dimensional responses pose the challenge of selective inference twice: the factorial design requires post-hoc correction, while high dimensionality demands multiple testing methods to accurately select relevant responses. While both problems are well- known and explored in statistical literature, there is a lack of solutions that address both problems jointly. Existing solutions either avoid estimating dependence among test statistics, resulting in underpowered methods, or rely on asymptotic results that are rarely met in practice. This problem is not purely theoretical and is commonly encountered in experimental fields such as genomics and neuroscience, where the response dimension can easily reach hundreds of thousands or even millions with only a few dozen observations. In this paper, we address this problem by extending the rotation-based approach [1] to the random restricted rotations The resulting test is exact under multivariate normality and guarantees asymptotic exactness in other cases while maintaining excellent control of the type I error. Additionally, this approach is computationally efficient and easily parallelizable. Furthermore, it provides an exact solution to the (multivariate) Behrens-Fisher problem even under heteroscedasticity among samples. Through a simulation study, our method demonstrates excellent control of the type I error, even under non-normal distributions, such as counts. The joint distribution of test statistics is derived through resampling, allowing for the adoption of any multiplicity control method. Our results show enhanced power compared to the most commonly used method.

[1] Solari, A., Finos, L., & Goeman, J. J. (2014). Rotation-based multiple testing in the multivariate linear model. Biometrics, 70(4), 954-961.

PARALLEL SESSION MO18: META-ANALYSIS

MO18.1 Bias corrections for study weights in meta-analyses with binary outcomes

Walter S.*, Balakrinshnan N.
McMaster University ~ Hamilton, Ontario ~ Canada

In the standard methodology of meta-analyses, the component studies are typically weighted by the inverse variance of their estimated treatment effect. This approach assigns greater weight to larger and/or more precise studies, and it gives maximum precision of the overall treatment effect. The weights are usually obtained by simply taking the reciprocal of the estimated variance. However, because reciprocation is a non-linear transformation, this naïve approach leads to positive bias for the weights, and hence all the component studies are over-weighted, especially small ones. This also then affects the overall estimated treatment effect and its standard error. Our objective was to develop bias corrections for the estimated study weights when the outcome is binary, extending our earlier work with continuous outcomes. We developed several analytic approximations for the inverse variances, according to which effect measure is adopted (risk difference, log relative risk, or log odds ratio). This leads to correction factors that permit a meta-analysis to be conducted more correctly. We found that the study weights using the standard reciprocal method are always over- estimated, particularly for small studies. Correspondingly, the variance of the estimated treatment effect is thereby under-estimated, possibly substantially. We evaluate the bias and its correction for typical scenarios, and show examples of how it affects actual meta-analyses. The standard practice of simply taking the reciprocal of a study variance to establish weight in a meta-analysis can be modified to correct for the over-weighting of component studies.
Walter SD, Balakrishnan N (2022). A method was developed for correcting the bias in the usual study weights in meta-analyses. J Clinical Epidemiology 152, 23-29.

MO18.2 Bayesian nonparametric approaches and the bias-corrected meta- analysis model for combining disparate studies

Verde P.E.*
Coordination Center for Clinical Trials, University Hospital, Heinrich Heine University Düsseldorf ~ Düsseldorf ~ Germany

Bayesian nonparametric (BNP) approaches for meta-analysis have been developed to flexibly handle the heterogeneity of random effects distributions. These types of models account for possible clustering and multimodality of the random-effects distribution. However, when we combine studies of varying quality or different types, the posterior mean of a location parameter (e.g. pool odds ratio) is not only a combination of results of interest but also multiple biases coming from low-quality studies.
The Bias Corrected meta-analysis model (Verde 2021) is a parametric random effects model designed to adjust for internal validity bias. This model is based on a mixture of two random effects distributions, where the first component corresponds to the model of interest and the second component to the hidden bias structure. In this way, the resulting model of interest is adjusted by the internal validity bias of the studies included in a systematic review. The aim of this work is to bring together BNP approaches and the BC meta-analysis model. In this way, a more flexible nonparametric BC meta-analysis model can be developed.
We illustrate the connections between BNP and BC models with simulated examples and with a real meta-analysis, which investigates the influence of baseline risk factors in complications and mortality in patients with positive COVID-19. In addition, we show that the use of simple random-effects meta-analysis models exaggerates the risk of comorbidities in those patients. We implement the nonparametric BC model in the R package jarbes, which facilitates its practical application.

Verde, PE. (2021) A bias-corrected meta-analysis model for combining, studies of different types and quality. Biometrical Journal. 2021; 63: 406– 422. https://doi.org/10.1002/bimj.201900376
Verde, PE (2022). jarbes: Just a Rather Bayesian Evidence Synthesis. R package version 2.1.0. https://CRAN.R-project.org/package=jarbes

MO18.3 Prospective and retrospective sequential meta-analysis using trial sequential analysis

Soerensen A.L.*[1], Soerensen A.L.[2], Soerensen A.L.[3], Harboe Olsen M.[3], Lange T.[1], Gluud C.[3], Gluud C.[4]
[1]Section of Biostatistics, Department of Public Health, University of Copenhagen ~ Copenhagen ~ Denmark, [2]Department of Mathematical and Physical Sciences, Macquarie University ~ Sydney ~ Australia, [3]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital - Rigshospitalet ~ Copenhag ~ Denmark, [4]Department of Regional Health Research, The Faculty of Health Sciences, University of Southern Denmark ~ Odense ~ Denmark

Many meta-analyses are updated in their lifetime, producing a sequence of point estimates, confidence intervals and p-values. The repeated testing of the same hypothesis comes with a cost of an increased risk of type-I-errors and a less obvious interpretation of the updated meta- analysis. We have investigated when and how to achieve valid statistical inference for a sequential meta-analysis depending on it being prospective or retrospective. Recall that, Trial Sequential Analysis (TSA) aims at controlling the type-I-error and produce valid inference when a meta-analysis has been sequentially updated [1]. This can be achieved when the meta-analysis is prospective. When a meta-analysis is retrospective, control of type-I-error cannot be guaranteed using TSA. However, TSA still outperforms naive testing. We have recently implemented an updated version of TSA in R, which we use to make extensive simulations to study the behavior of various types of updated meta-analyses. The results show where caution needs to be taken with the interpretation of the results of both naive testing and using TSA depending on whether the meta-analysis is prospective, retrospective, or sometimes a mixture of the two. We illustrate our main findings using case studies.. This also includes, how we would handle a sequentially updated meta-analysis and how results can be (or not be) interpreted. It is well known that TSA controls type-I-errors in sequentially updated meta-analyses. We found it difficult to achieve complete control of the type-I-error of sequentially updated meta-analyses especially when the meta-analyses are retrospective. Guidelines on validity and interpretation of results are provided.

[1] Wetterslev J, Thorlund K, Brok J, Gluud C. Journal of Clinical Epidemiology, 61, 2008, 64-75.

MO18.4 Pseudo-values approach for quantile analysis in individual patient data meta-analysis

Meddis A.*[1], Bouaziz O.[2], Paoletti X.[3], Latouche A.[4]
[1]Department of Public Health, University of Copenhagen ~ Copenhagen ~ Denmark, [2]Université Paris Cité, CNRS, MAP5 ~ F-75006 Paris ~ France, [3]Institut Curie & University of Versailles St Quentin / Paris Scaly & Inserm U900 ~ Paris ~ France, [4]Institut Curie & Conservatoire National des Arts et Métiers ~ Paris ~ France

Individual Patient Data (IPD) meta-analysis synthesizes evidence based on multiple clinical trials. For survival endpoints, several approaches have been previously proposed, that mostly involves hazard ratios and proportional hazard models. Here, we extend quantile regression for survival data to IPD meta-analysis. Quantiles are more flexible and robust quantitative tools for characterizing event times than mean-based regression models. Moreover, the proportional hazard assumption is not needed and it allows us to detect potential late treatment effects. We introduce a one-step quantile regression method for IPD meta-analysis to assess the treatment benefit, for fixed quantiles, in terms of difference in survival times. For every individual, we calculate the pseudo-values of the survival function, which is estimated by a weighted sum of trial-specific Kaplan-Meier estimators. At each quantile of interest we fit a GEE on the pseudo-values to account for the intra-trial correlation and a sandwich estimate for the variance-covariance matrix is defined. Further adjustment on prognostic factor can be implemented. Employing the functional delta method we provide the gain at the specific quantile on the difference in survival times with the respective confidence interval. Asymptotic results are obtained for a fixed number of trials, where the trials sample sizes goes to infinity. An illustration of the method is provided for an IPD meta-analysis of 11029 women collected from 17 randomized clinical trials of first-line therapy in advanced ovarian cancers where the effect on overall survival of novel therapies is compared to the standard strategy. Despite many methods have been introduced for correlated survival data, they all consider asymptotic results for infinite number of groups. This is not verified for a meta-analysis where few clinical trials are pooled together. We provide more insights on this topic and details on the definition of pseudo-values and quantile regression in meta-analysis. Ahn, Kwang Woo, and Brent R. Logan. "Pseudo-value approach for conditional quantile residual lifetime analysis for clustered survival and competing risks data with applications to bone marrow transplant data." The annals of applied statistics 10.2 (2016): 618.

TIN1 INVITED SESSION

EVALUATION OF PREDICTIVE ALGORITHMS AND MODELS: UNCERTAINTY AND IMPACT ON MEDICAL CARE

ORGANIZER | CHAIR: EWOUT STEYERBERG

TIN1.1 Sources of uncertainty in clinical risk prediction modeling

Van Calster B.*
KU Leuven ~ Leuven ~ Belgium

Clinical risk prediction models using classical or modern algorithms is everywhere in medicine. The estimated risk can be used to counsel the patient, or decide whether or not to offer a specific intervention. The risk is an estimate, and hence uncertain. I provide an overview of sources that contribute to the overall uncertainty in the estimate, and provide illustrations.
I start with discussing the key distinction between aleatoric (the outcome cannot be predicted with certainty) and epistemic uncertainty (uncertainty in the amount of aleatoric uncertainty). Further, epistemic uncertainty is divided into approximation uncertainty (sample size) and modeler (modeling choices) uncertainty. Measures to capture approximation uncertainty are described. Beyond the modeling effort itself, uncertainty is further increased by geographical and temporal heterogeneity affecting the data generating mechanism for model development and validation data. All abovementioned issues affect model performance as well as individual risk estimates, and are illustrated with a case study on ovarian cancer diagnosis. Risk prediction models and the resulting risk estimates have to be interpreted in terms of all sources of uncertainty. A simple consequence is that achieving correct individual risk estimates is impossible. The value of prediction models for individual counseling, including how risks should be presented, as well as for clinical decision making needs further attention.

[1]Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine [2]Learning 2021;110:457-506. Van Calster B, Steyerberg E, Wynants, van Smeden M. There is no such thing as a validated prediction model. BMC Medicine 2023;21:70.

Invited Sessions | Tuesday 29 August 2023

TIN1.2 Reporting and methodological quality of machine learning prediction model studies: an overview of results

Dhiman P.*[1], Ma J.[1], Andaur Navarro C.[2], Speich B.[3], Bullock G.[1], Damen J.[2], Hooft L.[2], Kirtley S.[1], Riley R.[4], Van Calster B.[5], Moons K.[2], Collins G.[1]
[1]University of Oxford ~ Oxford ~ United Kingdom, [2]Utrecht University ~ Utrecht ~ Netherlands, [3]University of Basel ~ Basel ~ Switzerland, [4]University of Birmingham ~ Birmingham ~ United Kingdom, [5]KU Leuven ~ Leuven ~ Belgium

Prediction models in cancer help inform cancer diagnoses, prognoses once diagnosed with cancer, and risk of developing cancer in the future. Use of machine learning to develop clinical prediction models in cancer has increased sharply in the last 10 years with a view to improve their predictive performance. However, there are growing concerns about the reporting and methodological quality, the risk of bias and potential overinflation of results from these studies. The objectives were to describe and critically appraise the reporting and methodology of prediction models developed using machine learning methods in oncology. We conducted a systematic review, searching MEDLINE and EMBASE for oncology-related studies that developed and validated a prognostic model using machine learning published between 01/01/2019 and 05/09/2019. We reviewed 62 publications (48 development-only; 14 development with validation). 152 models were developed across all publications and 37 models were validated. We assessed reporting adherence using the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) reporting guideline[1] and found that the median adherence to TRIPOD reporting items was 41% [range: 10%-67%]. We assessed risk of bias using the Prediction model Risk Of Bias ASsessment Tool (PROBAST)[2] and found that 84% (95% CI: 77 to 89) of developed models and 51% (95% CI: 35 to 67) of validated models were at overall high risk of bias; mostly due to shortcomings in the analysis including insufficient sample size, split-sample internal validation and poor handling of missing data. We used existing spin frameworks to describe areas of highly suggestive spin practices and found inconsistent reporting between methods and the results in 27% of studies due to additional analysis and selective reporting. Thirty-five studies (56%) used an overly strong or leading word in their title, abstract, results, discussion or conclusion. Quality of prediction models using machine learning in oncology is poor. We found poor reporting and low adherence to TRIPOD, high risk of bias from using insufficient sample sizes, poorly handling missing data and continuous predictors. We also found areas highly suggestive of spin in oncology prediction model research.

[1] Collins GS, Reitsma JB, Altman DG, Ann Intern Med,162, 2015, 55–63.
[2] Wolff RF, Moons KGM, Riley RD, Ann Intern Med; 170,2019 51–58.

TIN1.3 Measuring clinical utility: uncertainty in net benefit

Wynants L.*
Maastricht University ~ Maastricht ~ Netherlands

The impact of introducing a prediction model in clinical practice to inform clinical decisions on interventions (eg. treat patient vs. do not treat patient) can be quantified by Net Benefit (NB). NB is calculated as TP/N – FP/N * w, where TP is the number of true positives, FP is the number of false positives, and w is a weight reflecting the benefit of a TP and the harm of a FP. NB and decision curves (where NB is plotted for a range of w) are population-level quantities that can tell policymakers whether using a prediction model is better than using alternative strategies (such as treat all or treat none). Nonetheless, the NB estimate itself is uncertain. The objective of this talk is to investigate the origins and measures of NB uncertainty. Sampling variability and heterogeneity between populations are sources of uncertainty about NB. We will show that despite wide confidence and prediction intervals around NB, the choice of optimal strategy may be unaffected. A first measure of uncertainty is the probability of usefulness. It is the probability that the model is the optimal strategy among competing strategies and can be calculated through a random effects meta-analysis. The probability of usefulness has conceptual links with a second measure, the Net Benefit Value of Information (NB VOI). VOI is a concept borrowed from decision theory that quantifies the expected loss due to not confidently knowing which of competing strategies is the best. The methods will be illustrated with case studies in ovarian cancer diagnosis and prognosis after myocardial infarction. Uncertainty in NB can be large. The probability of usefulness from a random-effects meta-analysis reflects heterogeneity in clinical utility across populations, while the NB VOI can be used to determine whether more validation data from a certain population is needed before the model can safely be implemented.

[1] Wynants L, Riley RD, Timmerman D, Van Calster B. Random-effects meta-analysis of the clinical utility of tests and prediction models. Stat. Med. 2018;37(12):2034-52. doi:10.1002/sim.7653
[2] Sadatsafavi M, Lee TY, Wynants L, Vickers A, Gustafson P. Value of Information Analysis for External Validation of Risk Prediction Models. 2022 doi:10.48550/arXiv.2208.03343.

TIN2 INVITED SESSION

HIGH-DIMENSIONAL INFERENCE IN BIOSTATISTICS

ORGANIZER | CHAIR: FRANCESCO C. STINGO

TIN2.1 Spacex: gene co-expression network estimation for spatial transcriptomics

Acharyya S., Baladandayuthapani V.*, Zhou X.
University of Michigan ~ Ann Arbor ~ United States of America

The analysis of spatially resolved transcriptome enables the understanding of the spatial interactions between the cellular environment and transcriptional regulation. In particular, the characterization of the gene-gene co-expression at distinct spatial locations or cell types in the tissue enables delineation of spatial co-regulatory patterns as opposed to standard differential single gene analyses. To enhance the ability and potential of spatial transcriptomics technologies to drive biological discovery, we develop a statistical framework to detect gene co-expression patterns in a spatially structured tissue consisting of different clusters in the form of cell classes or tissue domains. We develop SpaceX (spatially dependent gene co-expression network), a Bayesian methodology to identify both shared and cluster-specific co-expression network across genes. SpaceX uses an over-dispersed spatial Poisson model coupled with a high-dimensional factor model which is based on a dimension reduction technique for computational efficiency. We show via simulations, accuracy gains in co-expression network estimation and structure by accounting for (increasing) spatial correlation and appropriate noise distributions. In-depth analysis of two spatial transcriptomics datasets in mouse hypothalamus and human breast cancer using SpaceX, detected multiple hub genes which are related to cognitive abilities for the hypothalamus data and multiple cancer genes (e.g. collagen family) from the tumor region for the breast cancer data. The SpaceX R-package is available at github.com/bayesrx/SpaceX.

TIN2.2 Bayesian hierarchical models for large-scale pharmacogenomic screens of Drug combinations

Zucknick M.*
University of Oslo ~ Oslo ~ Norway

With high-throughput drug sensitivity screens we can quickly test compounds on cancer cell lines to determine treatment efficacy. Since molecular characterisation of the cell lines by various omics data sets is frequently available, we can link molecular features to treatment efficacy. The estimation of drug synergy is important when testing multiple compounds in combination, but in-vitro cell viability measurements can be imprecise due to measurement errors, especially for drug combination experiments. To address this, we propose a modelling setup that uses our recently developed Bayesian model for synergy estimation with full uncertainty quantification.
We first introduce the motivation and challenges faced in drug combination screens, which include the large variance in drug synergy estimates mentioned above as well as the sample size being even smaller than in single-drug screens and the large number of drug combinations to test. We then present possible approaches for identification of biomarkers for drug synergy, starting with individual models for each drug combination separately using the regularised horseshoe prior. Because of the typically low sample size, there is often not enough signal to decisively escape the global shrinkage in the horseshoe prior. To address this issue, we use a variable selection approach called Signal Adaptive Variable Selector to separate the selection probability of a molecular feature from its (conditional) effect size. This approach allows for sparse estimates of individual main and interaction effects, but it does not borrow information across drug-combos. To address this limitation and thereby increase power and improve accuracy in the identification of potential molecular biomarkers for synergistic effects, we will sketch some avenues of how we can borrow information across drug combinations using joint Bayesian hierarchical models. Overall, our work demonstrates the potential of Bayesian hierarchical models for improving biomarker discovery in the challenging setting of large-scale drug combination screens. This is joint work with Leiv Rønneberg and Paul Kirk. Rønneberg L, Cremaschi A, Hanes R, Enserink JM, Zucknick M (2021). bayesynergy: flexible Bayesian modelling of synergistic interaction effects in in vitro drug combination experiments, Briefings in Bioinformatics 22(6), bbab251.

TIN2.3 Outcome-guided multi-view bayesian clustering for integrative omic data Analysis

Paul K.*
University of Cambridge ~ Cambridge ~ United Kingdom

Although the challenges presented by high dimensional data in the context of regression are well- known and the subject of much current research, comparatively little work has been done on this in the context of clustering. In this setting, the key challenge is that often only a small subset of the covariates provides a relevant stratification of the population. Identifying relevant strata can be particularly challenging when dealing with high-dimensional datasets, in which there may be many variables that provide no information whatsoever about population structure, or - perhaps worse - in which there may be (potentially large) variable subsets that define irrelevant stratifications. For example, when dealing with genetic data, there may be some genetic variants that allow us to group patients in terms of disease risk, but others that would provide completely irrelevant stratifications (e.g. which would group patients together on the basis of eye or hair colour).
Bayesian profile regression is an outcome-guided model-based clustering approach that makes use of a response in order to guide the clustering toward relevant stratifications. Here we show how this approach can be extended to the "multiview" setting, in which different groups of variables ("views") define different stratifications. We present some results in the context of breast cancer subtyping to illustrate how the approach can be used to perform integrative clustering of multiple 'omics datasets.

When there are multiple clustering structures present in data, existing (single view) clustering approaches can fail to recover the most relevant clustering structure, even when guided by an appropriate response. Moreover, traditional variable selection approaches for clustering do not necessarily improve matters, since they tend to select variables that define the dominant clustering structure, regardless of whether or not it is associated with a response of interest. Real molecular datasets can and do possess multiple clustering structures, and our outcome-guided multi-view model can allow both relevant and irrelevant structures to be identified.

Molitor, et al. Bayesian profile regression with an application to the National Survey of Children's Health. Biostatistics. 2010.
Kirk, Pagani, Richardson. Bayesian outcome-guided multi-view mixture models with applications in molecular precision medicine. arXiv 2023

PARALLEL SESSION TO1: CLINICAL TRIALS 4

TO1.1 How (not) to conduct a simulation study for a trial design: a case of dose-finding clinical trials

Mozgunov P.*[1], Paoletti X.[2], Jaki T.[1]
[1]University of Cambridge ~ Cambridge ~ United Kingdom, [2]Institute Curie ~ Paris ~ France

It is increasingly common to evaluate a newly proposed clinical trial design via simulations. This can be particularly useful if there are no closed-form solutions for the proposed design, to study the small sample size behavior, and to understand the robustness of the novel approach. However, the scenarios for these simulations are chosen by the developers of the method themselves. This can add subjectivity to the assessment as one can (almost) always find scenarios in which one design would outperform another. Early on, this problem was recognized for Phase I dose-escalation trials, for which numerous design alternatives are proposed in the literature. To tackle it in a monotherapy trial with binary outcomes, a tool called non-parametric optimal benchmark was developed. This tool provides the upper bound on the accuracy of a dose-finding design in a given simulation scenario. However, since then, the trial settings have greatly evolved, and non-binary outcomes and combination trials became of major interest. In this talk, using a motivating example of an actual Phase I dose-escalation design that we were asked to evaluate, we will introduce a generalization of the non-parametric optimal benchmark to more advanced settings such as multiple endpoints with arbitrary discrete/continuous endpoint and the combination setting. We will show how the proposed tool could point out the flaws of designs proposed in the literature that were originally presented with no competing approaches. We will also talk about generalizations of the non-parametric optimal benchmark to other trial settings. The developed tool can provide a better context for the evaluation of dose-finding designs in many settings. Despite being originally developed for Phase I trials, the proposal can provide a basis for further developments of the unified framework of the evaluation tools for trial designs.

TO1.2 Using ctdna as a novel biomarker of efficacy for dose-finding trials in oncology

Chen X.*, Mozgunov P., Jaki T.
University of Cambridge ~ Cambridge ~ United Kingdom

Dose-finding trials aim at coming up with a safe and effective drug administration during the early phase of clinical trial designs. There is a vast literature on Bayesian adaptive dose- escalation methods in cancer, which have shown limited use in clinical practice. One of the impediments is that efficacy or activity outcomes might not be available soon enough to apply decision rules to choose treatments for next patients. In the last decade, `liquid biopsy' technologies have been developed, which allow for the incorporation of novel early response endpoints. This project considers the use of a rapidly available biomarker, circulating tumor DNA (ctDNA), when using the bayesian adaptive EffTox approach in dose-finding studies. We conduct simulations to compare the original EffTox and the Biomarker-Informed EffTox (BMI- EffTox) with additional information based on weekly ctDNA under various scenarios and a list of settings of the trajectory of ctDNA over time. Simulation results show that the proposed BMI-EffTox approach yields significantly shorter trial duration, does not expose patients to additional risk of toxicity, and possesses some desirable properties in terms of the proportion of correct decisions in dose-finding studies.
Orloff, John, et al. "The future of drug development: advancing clinical trial design." Nature reviews Drug discovery 8.12 (2009): 949-957. Burnett, Thomas, et al. "Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs." BMC medicine 18.1 (2020): 1-21.

Parallel Sessions | Tuesday 29 August 2023

TO1.3 Estimating the similarity between adult and pediatric dose-toxicity curves to inform pediatric dose-finding

Zocholl D.*[1], Götte H.[2], Wiesenfarth M.[4], Schüler A.[2], Habermehl C.[2], Kopp--Schneider A.[3], Günhan B.[2]
[1]Charité - Universitätsmedizin Berlin ~ Berlin ~ Germany, [2]Merck Healthcare KGaA ~ Darmstadt ~ Germany, [3]Division of Biostatistics, German Cancer Research Center (DKFZ) ~ Heidelberg ~ Germany, [4]Cogitars ~ Heidelberg ~ Germany

We consider Bayesian model-based approaches to dose-escalation trials in pediatric patients, for which information from a trial in adults with the same compound can be borrowed in form of an informative prior (Takeda and Morita, 2018, and Zocholl et al. 2022). The degree of the borrowing should ideally be controlled by the similarity between adult and pediatric dose-toxicity profiles. This similarity can be informed by other compounds that are similar in their structure and mode of action, if available (e.g. PD-L1-inhibitors in oncology). If these other compounds have previously been tested in adult and pediatric patients, it may be possible to estimate the (average) degree of similarity between adult and pediatric toxicity profiles, and to use this information to inform the degree of borrowing for the upcoming pediatric trial. We design a pediatric dose-finding trial, where adult data from the compound under investigation is available as well as adult and pediatric data from several other similar compounds. The adult data from the same compound is borrowed by applying robust priors as shown previously (Takeda and Morita, 2018, and Zocholl et al. 2022). To adjust the degree of borrowing based on available evidence, we use a similarity parameter which represents the degree of similarity between adult and pediatric dose-toxicity curves estimated from the other compounds. We propose two methods to estimate the similarity parameter and demonstrate how it can be applied to some of the existing borrowing methods. The performance of the methods is investigated in a simulation study. Depending on the available information, it may be possible to inform the borrowing process for pediatric dose-finding trials, which can benefit the pediatric dose-finding trial. However, the amount of available information is critical and in real-word applications there might often not be enough information available to arrive at a meaningful estimation of the degree of similarity. Takeda, K. and Morita, S. (2018). Bayesian dose-finding phase i trial design incorporating historical data from a preceding trial. Pharmaceutical statistics, 17(4):372–382. Zocholl, D., Wiesenfarth, M., Rauch, G., and Kopp-Schneider, A. (2022). On the feasibility of pediatric dose-finding trials in small samples with information from a preceding trial in adults. Journal of Biopharmaceutical Statistics, pages 1–19.

TO1.4 Incorporating patient-reported outcomes in dose-finding clinical trials with continuous patient enrollment

Anais A.*[1], Biard L.[2], M Lee S.[3]
[1]Saryga ~ Tournus ~ France, [2]INSERM U1153 Team ECSTRRA, Université Paris Cité ~ Paris ~ France, [3]Mailman School of Public Health, Columbia University ~ New York ~ United States of America

Dose-finding clinical trials in oncology aim to estimate the maximum tolerated dose (MTD), based on safety traditionally obtained from the clinician's perspective. While the collection of patient-reported outcomes (PROs) has been advocated to better inform treatment tolerability, there is a lack of guidance and methods on how to use PROs for dose assignments and recommendations. The PRO continual reassessment method (PRO-CRM) has been proposed to formally incorporate PROs to estimate the MTD, requiring complete follow-up of both clinician and patient toxicity information per dose cohort to assign the next cohort of patients We propose two extensions of the PRO-CRM, allowing continuous enrollment of patients and handling longer toxicity observation windows to capture late-onset or cumulative toxicities. We use a weighted likelihood to include the partial follow-up information from PRO and clinician- reported toxicity outcome. The first method, the TITE-PRO-CRM, uses both the PRO and the clinician's information to estimate the MTD during and at the end of the trial. The second method, the TITE-CRM + PRO, uses clinician's information solely to inform dose assignments during the trial and incorporates PRO at the end of the trial for dose recommendation. Simulation studies show that the TITE-PRO-CRM performs similarly to the PRO-CRM in terms of dose recommendation and assignments during the trial while almost halving trial duration in case of an accrual of 2 patients per observation window. The TITE-CRM + PRO slightly underperforms compared to the TITE-PRO-CRM, but similar performance can be attained by requiring larger sample sizes. We also show that the proposed methods have similar performance under higher accrual rates, different toxicity hazards, and correlated time-to- clinician toxicity and time-to-patient toxicity data. Basch E., L. J. Rogak, A. C. Dueck. 2016. Methods for Implementing and Reporting Patient- reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clinical therapeutics 38(4):821-830. doi:10.1016/j.clinthera.2016.03.011. Retzer, A., O. L. Aiyegbusi, A. Rowe, et al. 2022. The Value of Patient-Reported Outcomes in Early-Phase Clinical Trials. Nature Medicine 18-20. doi: 10.1038/s41591-021-01648-4.
U.S. Food and Drug Administration. FDA's Project Optimus. https://www.fda.gov/about- fda/oncology-center-excellence/project-optimus Lee, S. M., X. Lu, B. Cheng. 2020. Incorporating patient-reported outcomes in dose-finding clinical trials. Statistics in medicine 39(3):310-325. doi: 10.1002/sim.8402.

TO1.5 Designing patient-centred dose-finding trials with patient-reported outcomes: opportunities and challenges

Alger E.*, Yap C.
The Institute of Cancer Research ~ London ~ United Kingdom

Within early phase dose-finding trials, the recommended phase 2 dose is often determined by dose limiting toxicities (DLTs) graded by clinicians during the DLT assessment period (usually up to cycle 1 or 2). As new treatments emerge, patients may receive more cycles of treatment, thus lower grade toxicities often under-reported by clinicians may become intolerable for patients. Echoing the FDA's Project Optimus initiative, we now look to develop new methodology to assess drug tolerability beyond initial treatment cycles and to incorporate patient-reported outcomes (PROs). This work aims to explore different ways to integrate PROs within dose-finding designs to evaluate patients' health-related quality of life alongside a clinicians' assessment of toxicity. Currently, the only published phase I model design using a PRO component is the PRO-CRM design[1]. This model, integrating a binary PRO-DLT within a continual reassessment method (CRM) model-based design, has so-far only been implemented in one phase I study of endometrial cancer[2]. There is obvious scope to develop additional trial designs incorporating PROs. Building a new model requires three considerations: (1) developing a new statistic to summarise PRO data, (2) developing a model to predict PRO tolerability and (3) selecting a trial design which combines the PRO endpoint with clinician assessed DLT. Within this work, we explore opportunities and challenges associated with the prediction of tolerability when using PROs as an ordinal variable, binary variable, and repeated measure. Of particular note, we consider the effect of missing at random PRO data, implementation considerations and the interpretability of the fitted model. Once we have utilised PROs to predict tolerability, we look to combine this with a clinician's assessment of tolerability using an adapted, contemporary clinical trial design assessing joint outcomes. To achieve patient-centred drug development, there is an increasing interest in the development of models which incorporate PROs within dose-finding designs and to account for tolerability assessments beyond initial treatment cycles. Following considerations presented here, we look to originate a new trial design with appropriate simulation studies and comparisons to the PRO- CRM design.

[1] Lee SM, Lu X, Cheng B. Incorporating patient-reported outcomes in dose-finding clinical trials. Stat Med. Feb 10 2020;39(3):310-325. doi:10.1002/sim.8402
[2] Wages NA, Nelson B, Kharofa J, Meier T. Application of the patient-reported outcomes continual reassessment method to a phase I study of radiotherapy in endometrial cancer. Int J Biostat. Nov 17 2022;doi:10.1515/ijb-2022-0023

PARALLEL SESSION TO2: SURVIVAL ANALYSIS 4

TO2.1 Sign-flip test for coefficients in the cox regression model

De Santis R.*[1], Finos L.[1], Goeman J.[2]
[1]University of Padova ~ Padova ~ Italy, [2]Leiden University Medical Center ~ Leiden ~ Netherlands

The Cox regression model [1] is a milestone in survival analysis, whose aim is to quantify the measure of some covariates on the hazard ratio of a certain event of interest. The impact of these covariates is checked through a statistical test on the regression coefficient, which is derived through the proven asymptotic normality of the coefficients. The most common choice -- a default choice in the most known statistical packages -- consists in the use of the Wald test. However, it can show a slow convergence to the nominal level when some of the coefficients are nuisance parameters, especially when they are correlated. We propose to use the sign-flip approach in order to improve the quality of inference for small sample sizes. We derive a semi-parametric score test, which is carried out by means of an appropriate modification of score contributions, while it does not require the estimation of the Fisher information. Indeed, the test is based on firstly define a test statistic, based on the effective score function, and then defining null-invariant transformations of this test statistic in order to get an asymptotically exact test. This approach has already been successfully applied in generalized linear models [2]. Through some simulations we aim to compare the new method with the existing parametric approach in the setting of a correctly specified model, with different sample sizes and correlation between covariates, and also the behavior in presence of possible misspecifications of the model. The simulations show a substantial improvement on the rate of convergence of the new test proposed to the nominal level for small sample size when the model is correctly specified. Further investigation will reveal potential robustness against misspecification of the model, e.g. when the proportionality assumption does not hold.

[1] D.R. Cox, "Regression models and time tables", Journal of the Royal Statistical Society, Ser. B, 34, 1972, 187-220.
[2] J. Hemerik, J.J. Goeman, L.Finos, "Robust testing in generalized linear models by sign flipping score contributions", Journal of the Royal Statistical Society, Ser. B, 82, 2020, 841-864.

Parallel Sessions | Tuesday 29 August 2023

TO2.2 Penalized likelihood estimation of cox models with doubly truncated and interval censored survival times

Webb A.*, Ma J.
Macquarie University ~ Sydney ~ Australia

Truncation is a common phenomenon in studies where a time-to-event outcome is of interest. Double truncation arises when participants included in the study are restricted to those with an event time falling within some random interval which may be subject-specific. Left truncation and right truncation are special cases of double truncation which may arise, respectively, when study design delays entry for some or all participants, or when data is sampled retrospectively. In addition, the event times of interest in a study affected by truncation may also be right, left, or interval censored or some combination thereof, which we refer to as partly-interval censoring. Estimation methods for some specific combinations of truncation and censoring types are well established; for example, the use of partial likelihood estimation for data that is left truncated and subject to right censoring is common. However, existing methods for incorporating general truncation alongside general partly-interval censoring are limited. To address this, we propose a penalised likelihood method for fitting a Cox proportional hazards model for possibly doubly truncated survival data subject to a general partly-interval censoring scheme. Our method includes estimation of the regression parameters and a smooth basis function approximation of the non-parametric baseline hazard function. We present asymptotic variance results allowing for inferences on the regression parameters and survival quantities. We compare the performance of our proposed method to existing methods for truncated data via simulation studies, including partial likelihood estimation for left truncation and inverse probability weighting methods for right and double truncation. Simulation results indicate that the performance of our MPL approach is comparable to partial likelihood and inverse probability weighting methods and offers greater flexibility for scenarios with complex truncation and censoring schemes. We illustrate our method via an application to real data from a study of Parkinson's disease.

TO2.3 Impact of non-informative censoring on propensity score based estimates of marginal hazard ratios

Wang De Faria Barros G.*, Häggström J.
Umeå University ~ Umeå ~ Sweden

In medical and epidemiological studies, one of the most common settings is studying the effect of a treatment on a time-to-event outcome, where the time-to-event might be censored before the end of study. A common parameter of interest in such a setting is the marginal hazard ratio (MHR). When a study is based on observational data, propensity score (PS) based methods are often used, in an attempt to make the treatment groups comparable despite having a non- randomized treatment [1]. Previous studies have shown censoring to be a factor that induces bias when using PS based estimators [2]. In this paper we study the magnitude of the bias under different rates of non-informative censoring when estimating the MHR using PS weighting or PS matching. A bias correction involving the probability of event is suggested and compared to conventional PS based methods.
A monte carlo simulation was created to study the impact of non-informative censoring in MHR estimation. We considered settings equivalent to observational and randomized studies, censoring proportions ranging from 10% to 90%, population sizes from 2000 to 10000, uniform and Weibull censoring distributions, true MHR ranging from 0.5 to 2 for a total of 270 scenarios. For each scenario, we simulated 1000 datasets and estimated the MHR with PS matching and PS weighting with and without our suggested probability of event correction. We found MHR estimations without any corrections to have an increasing bias according to the strength of the treatment effect and the censoring proportion, even in settings equivalent to randomized studies. Our suggested correction reduces bias and results in good estimations for most scenarios, with the exception being the most extreme censoring scenarios. The problem of biased PS based MHR estimation under high non-informative censoring is not yet solved, so further studies in this area are needed and, although a procedure to reduce this bias was presented, it has not been fully eliminated.

[1] Austin, P. C. and E. Stuart. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. Statistical Methods in Medical Research 26 (2015), 1654–1670.
[2] Wyss, R. et al,. Use of Time-Dependent Propensity Scores to Adjust Hazard Ratio Estimates in Cohort Studies with Differential Depletion of Susceptibles. Epidemiology 31.1 (2020), 82–89.

TO2.4 Impact of omitted covariates on treatment estimates in propensity score matched studies

Strobel A.*[1], Wienke A.[1], Kuß O.[2]
[1]Martin-Luther-University Halle-Wittenberg ~ Halle (Saale) ~ Germany, [2]German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich-Heine-University ~ Düsseldorf ~ Germany

Propensity score (PS) matching has become a popular method for estimating causal treatment effects in nonrandomized studies. However, for time-to-event outcomes, the estimation of hazard ratios based on propensity scores can be challenging if omitted covariates are present, but disregarded. In general, they will induce "unobserved" heterogeneity and not accounting for such covariates could lead to heavily biased treatment estimates [1]. Researchers often do not know whether (and, if so, which) covariates will cause this bias. To address this issue, we extended a previously described method, "Dynamic Landmarking", which was originally developed for randomized trials. "Dynamic Landmarking" was developed to identify biased treatment estimates in randomized trials. The method is based on successively deletion of (sorted) observations and gradually fitting Cox models until no sufficient number of events is contained in the data. In addition, the balance of observed, but omitted covariates can be measured by the z-differences [2]. If an omitted covariate will induce heterogeneity, a systematic selection of patients is expected, which results in biased treatment estimates. By simulation, we show here that "Dynamic Landmarking" provides a good visual tool for detecting biased treatment estimates also in propensity score matched data. We found, that biased treatment estimates are particularly present if the omitted covariate is not associated with the PS and has a high impact on the survival outcome. Finally, we illustrate "Dynamic Landmarking" also in a real data set from cardiac surgery. In summary, "Dynamic Landmarking" can indeed identify biased treatment estimates in propensity score matched studies. However, this only becomes an issue, if the omitted covariate is independent from the PS, has a high impact on survival, and if an effective treatment is given. Our empirical example shows that it could be important to check whether omitted covariates could induce heterogeneity, to avoid biased treatment estimates due to applying only the standard Cox model.

[1] W. Pan, H. Bai, Propensity score analysis: Fundamentals and developments, Guilford Press, 2015
[2] O. Kuss, The z-difference can be used to measure covariate balance in matched propensity score analyses, Journal of Clinical Epidemiology. 66(11): 2013, 1302-1307.

TO2.5 Comparing overall benefit/risk of treatments by weighted cox model on ordering scores for relevant events

Pap Á.F.*
Clinical Data Sciences & Analytics, Bayer AG ~ Wuppertal ~ Germany

It is often of interest to compare efficacy and safety of a new treatment with a control with a single composite measure estimated from time to event endpoints of different severity1. A modified win ratio measure is proposed which is estimated using weights derived from the severity of the events.
Follmann et al2 introduced estimation of 'win ratio' for ranked events by constructing ordering scores. For each patient the worst event they experienced is selected and the ordering score is derived as the sum of the time to the worst event or censoring and the multiple of the maximal observation time over the entire study where the multiplication factor is based on the rank of the event (0 for the worst event). Cox model is fitted on ordering scores with multiple interval censoring data structure where each patient has at least 1 record or more depending on the severity of the event. The exponentiated parameter estimate for treatment can be interpreted as 'loss ratio' and its reciprocal as 'win ratio'. An extension of the method is presented which considers the severity of the event by weighted Cox regression with weight applied to all records for the same event: for the most severe event the weight is 1, for less severe event it is less than 1. Time to efficacy and safety event data are simulated based on different incidence rates and treatment effects for each event type. Distribution and power characteristics of win ratios are compared from equally weighted and event-severity-weighted Cox models on ordering scores and contrasted to corresponding estimates obtained by the usual Cox model on time to first event. The weighted win ratio estimate can be more favorable than the other estimates if the weight for the adverse event are smaller than those for the efficacy events. For the assessment of the overall treatment effect on several efficacy and safety evens, the 'win ratio' obtained by weighted Cox regression on the ordering scores can give more relevant insight than the simple 'win ratio' estimation or the corresponding estimate from the time to first event Cox model.

[1] B. Redfors, J. Gregson, A. Crowley, T. McAndrew, O. Ben-Yehuda, G.W. Stone, S.J. Pocock. European Heart Journal, 41, 2020, 4391–4399
[2] D. Follmann, M. P. Fay, T. Hamasaki, S. Evans, Statistics in Medicine, 39, 2020, 602-616.

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO3: STUDENTS AWARD

TO3.1 Sensitivity analysis for missingness assumptions in causal inference: accommodating the substantive analysis

Zhang J.*, Dashti S.G., Carlin J.B., Lee K.J., Moreno--Betancur M.
the University of Melbourne; Murdoch Children's Research Insititute ~ Melbourne ~ Australia

When conducting multiple imputation (MI), maintaining compatibility between the imputation model and substantive analysis is important for avoiding bias. Recently, two compatible MI approaches have been developed: the "Substantive Model Compatible Fully Conditional Specification (SMCFCS)" which accommodates the substantive outcome model in the FCS procedure to ensure compatibility [1]; and a stacked-imputation-based approach (SMC-stack), which multiply imputes non-outcome variables ignoring the outcome and stacks them into a single dataset that is analysed using weights proportional to the substantive outcome model density [2]. Both methods are guaranteed to be unbiased under the "missing at random" assumption. In practice, however, it is common for an incomplete outcome to be a cause of its own missingness or to be associated with its missingness given all other analysis variables, which implies a violation of this assumption. Although methods such as "not-at-random (NAR) FCS" provide an appealing approach for sensitivity analysis with multivariable missingness, compatible approaches are lacking, which is a key gap as using incompatible imputation in sensitivity analysis may induce bias. This is particularly pertinent when estimating the average causal effect (ACE) using g-computation, which uses an outcome model including exposure- confounder interactions.
To address this gap, we propose two approaches for compatible sensitivity analysis for the missingness assumptions when an incomplete outcome causes or is associated with its own missingness. The proposed approaches, NAR-SMCFCS and NAR-SMC-stack, extend SMCFCS and SMC-stack, respectively, by incorporating the outcome missingness indicator with an associated sensitivity parameter ("delta") when imputing the outcome. We evaluated the performance of our proposed methods through a simulation study motivated by a real case study, which we also analysed. We considered a range of outcome models and multivariable missingness mechanisms, where the substantive analysis aimed to estimate the ACE using correctly specified g-computation. The simulation results showed that both approaches reduced the bias in ACE estimation compared to the default NARFCS approach. However, the variance estimator for the NAR-SMC-stack approach was downward biased. We conclude that NAR-SMCFCS is preferred over NAR-SMC-stack and NARFCS to conduct sensitivity analysis for missingness assumptions in causal inference.

[1] Bartlett, Jonathan W., Shaun R. Seaman, Ian R. White, James R. Carpenter, and Alzheimer's Disease Neuroimaging Initiative. "Multiple imputation of covariates by fully conditional specification: accommodating the substantive model." Statistical methods in medical research 24, no. 4 (2015): 462-487.*
[2] Beesley, Lauren J., and Jeremy MG Taylor. "A stacked approach for chained equations multiple imputation incorporating the substantive model." Biometrics 77, no. 4 (2021): 1342-1354.

TO3.2 A location-scale joint model with a time-dependent subject-specific variance of the marker and competing event

Courcoul L.*, Barbieri A., Tzourio C., Jacqmin--Gadda H.
Univ. Bordeaux, INSERM, Bordeaux Population Health, U1219, France ~ Bordeaux ~ France

A high level of blood pressure is a well-known risk factor for several major cardio- and cerebrovascular diseases (stroke, myocardial infarction, etc) but an increasing number of studies suggests that individual blood pressure variability may also be an independent risk factor for these events. However, these studies suffer from significant methodological weaknesses and often consider a time-independent variability. The objective of this work was to develop and apply a location-scale joint model with a time-dependent subject-specific variance for the longitudinal marker and competing events to study the association between blood pressure variability and health events. The proposed joint model combines a mixed model for longitudinal data and a cause-specific model for competing events with proportional intensities. The residual variance of the marker is modelled according to subject-specific random intercept and random slope and possibly covariates. The risk of events may depend simultaneously on the current value of the residual variance, as well as, the current value and the current slope of the marker. The model is estimated by maximizing the likelihood function, using the Marquardt-Levenberg algorithm. The estimation procedure is implemented in an R-package and is validated through a simulation study. The model was applied to the data of the PROGRESS clinical trial for the prevention of the recurrence of stroke that includes 6105 subjects followed over 5 years with 12 measurement times for blood pressure. We demonstrated that the risk of major cardio- or cerebrovascular events and the risk of competing death increased with the variability of blood pressure. The results highlighted the significant association between the variability of blood pressure measurements and the risk of cardiovascular or cerebrovascular events. Our method can also improve the prediction of an event of interest compared with a model that does not take into account this individual variability. Finally, it may also be relevant for investigating the association between the variability of markers and the risk of health events in various fields of medical research (e.g., emotional instability and psychiatric events, glucose variability and prognosis of diabetes).

[1] H. de Courson, K. Leffondre, and C. Tzourio. Blood pressure variability and risk of cardio- vascular event : is it appropriate to use the future for predicting the present ? European heart journal., 39(47) :4220–4220, 2018.
[2] J.K. Barrett, R. Huille, R. Parker, Y. Yano, and M. Griswold. Estimating the association between blood pressure variability and cardiovascular disease : An application using thearic study. Statistics in medicine, 38(10) :1855–1868, 2019.

TO3.3 A bayesian joint modelling for misclassified interval-censoring and competing risks

Yang Z.*[1], Rizopoulos D.[1], Heijnsdijk E.[1], Newcomb L.[2], Erler N.[1]
[1]Erasmus Medical Center ~ Rotterdam ~ Netherlands, [2]Fred Hutchinson Cancer Center ~ Seattle ~ United States of America

Our work is motivated by the Canary Prostate Active Surveillance Study (PASS) [1] that closely follows prostate cancer patients with low-grade tumors. Longitudinal prostate-specific antigen (PSA) measurements are collected for these patients to monitor their progression and schedule biopsies. Our interest is to study the association between serial PSA and the risk of progression. The challenges are threefold: (1) the primary event, cancer progression, detected by periodic biopsies, is interval-censored, (2) biopsies with imperfect sensitivity lead to misclassification , and (3) patients may decide to initiate treatment without progression.
The framework of joint models (JMs) for longitudinal and time-to-event data has been previously used in this context to study the association between PSA and progression. However, these previous attempts have not accounted for misclassification and competing risks. This work proposes a novel extension of joint models to account for these issues. Our contribution is two- fold, first, in the model specification, and second, in the formulation of the likelihood. In particular, the likelihood of the model constists of three parts. For patients with detected progression, it includes the probabilities that progression may have been detected immediately or that one or more prior biopsies missed it. For censored or treated patients, it models the possibility that progression occurred in any of the intervals between biopsies but was missed by all biopsies. A further contribution of our work is accounting for the uncertainty in the sensitivity of biopsies. Under the Bayesian framework that we follow, the uncertainty about the sensitivity may be taken into account via a prior distribution, which, however, may raise an identifiability issue. We explore options to resolve this as well as alternative approaches, such as combining MCMC samples from models assuming different fixed biopsy sensitivities. Considering biopsy sensitivities of 80% or 60% in the JMs fitted on the Canary PASS data led to a 4% - 25% increase in the hazard ratio of the biomarker of interest. Lower biopsy sensitivities resulted in higher baseline hazards. This connection indicates that restrictions on the baseline hazard specification can help to obtain identifiable models when using a prior distribution for the sensitivity parameter.

[1] Newcomb LF, Thompson IM, Boyer HD, et al. Outcomes of active surveillance for clinically localized prostate cancer in the prospective, multi-institutional Canary PASS cohort. Journal of Urology 2016; 195(2): 313-320. http://dx.doi.org/10.1016/j.juro.2015.08.087 doi:10.1016/j.juro.2015.08.087

Parallel Sessions | Tuesday 29 August 2023

TO3.4 A joint model for (un)bounded longitudinal markers, competing risks, and recurrent events using registry data

Miranda Afonso P.*[3], Rizopoulos D.[3], Palipana A.[2], Clancy J.P.[1], Szczesniak R.D.[2], Andrinopoulou E.[3]
[1]Cystic Fibrosis Foundation ~ Bethesda ~ United States of America, [2]Department of Pediatrics, Cincinnati Children's Hospital Medical Center ~ Cincinnati ~ United States of America, [3]Department of Biostatistics & Epidemiology, Erasmus University Medical Center ~ Rotterdam ~ Netherlands

Cystic fibrosis (CF) is a genetic disease that affects the lungs and digestive system. Pulmonary exacerbations (PEx) are recurrent complications that increase the need for a lung transplant and the risk of death. The body mass index (BMI) and the percentage of predicted forced expiratory volume in 1 second (ppFEV1) are commonly measured in CF patients to monitor their nutritional status and lung function, respectively. Our primary goal is to simultaneously investigate the associations between ppFEV1, BMI, and the risks of PEx, transplantation, and death, using all available US CF Foundation Patient Registry data. Due to the size and complexity of this dataset, previous analyses have been limited to the first PEx, disregarding subsequent occurrences and neglecting informative censoring due to transplantation and death.[1] Furthermore, despite ppFEV1 being bounded, it has previously been modeled using a Gaussian distribution, leading to predictions outside the feasible range. We develop a Bayesian shared-parameter joint model for recurrent events, competing events, and multiple longitudinal markers following different distributions. For our application, we model ppFEV1 and BMI assuming beta and Gaussian distributions, respectively. We allow the specification of various functional forms to link the longitudinal and time-to-event processes. Our model accommodates discontinuous intervals of risk, as well as the gap and calendar timescales. The model is available in the R package JMbayes2.[2] The posterior sampling algorithms are implemented purely in C++, allowing fast model fitting. Our results show that ppFEV1 and BMI are associated with the risk of experiencing PEx. For example, a ten-unit increase in the rate of ppFEV1 decline increases the hazard of PEx by 14.69% (95%CI 13.09-14.69). The incidence of PEx is positively associated with transplantation and death, with a one-SD increase in the frailty term increasing the hazard by 290.74% (95%CI 264.96-317.43) and 229.95% (95%CI 211.98-247.93), respectively. The proposed joint model allows for more accurate estimates of the risks posed by PEx and can improve monitoring strategies to reduce the frequency and severity of the symptoms. By making our model available in JMbayes2, we hope to assist others in performing joint analyses of longitudinal and time-to-event data in complex settings.

[1] E.-R. Andrinopoulou, J.P. Clancy, R.D. Szczesniak, Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. BMC Pulmonary Medicine, 20, 2020, 1–11.
[2] D. Rizopoulos, G. Papageorgiou, P.M. Afonso, JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data, https://drizopoulos.github.io/JMbayes2/, 2023.

TO3.5 Parametric estimation of the mean number of events in the presence of competing risks

Entrop J.P.*[3], Jakobsen L.H.[1], Crowther M.J.[2], Clements M.[4], Eloranta S.[3], Weibull C.E.[3]
[1]Department of Mathematical Science, Aalborg University ~ Aalborg ~ Denmark, [2]Red Door Analytics AB ~ Stockholm ~ Sweden, [3]Clinical Epidemiology Division, Department of Medicine, Solna, Karolinska Institutet ~ Stockholm ~ Sweden, [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet ~ Stockholm ~ Sweden

Recurrent events are common in time-to-event analyses. Examples include CVD events, hospital admissions and childbirths. When a recurrent event is the only possible outcome, a useful summary measure is the mean number of events for which estimation methods exist and have been implemented. However, estimation is more challenging in the competing risk setting, and results are commonly limited to presenting the cause-specific cumulative incidence function for the first occurring event. A variety of methods are available for jointly modelling both recurrent and competing events, e.g. multistate models and joint frailty models. Nevertheless, methods for estimating the mean number of events are so far limited to non-parametric models with scare implementations. To this end we have developed and implemented a parametric estimator of the mean number of events, as well as transformations of it. Cook and Lawless [1] suggested a non-parametric estimator of the mean number of events that was further developed by Ghosh and Lin [2] and based on the Kaplan-Meier estimator (for the competing events), and the Aalen-Johansen estimator (for the intensity function of the recurrent event). We propose a flexible parametric model to jointly estimate the survival function of the competing event and the intensity function of the recurrent event as an alternative to the non- parametric estimators suggested by Cook and Lawless. Confidence intervals for the parametric estimator are obtained using the delta method. The proposed method is implemented in the R- package JointFPM which is available on GitHub (https://github.com/entjos/JointFPM). Simulations demonstrated low bias and good coverage of the proposed estimator. We applied our method to data from the Swedish lymphoma register, comparing the mean number of childbirths between non-Hodgkin lymphoma survivors with different subtypes, in the presence of death as competing risk. Providing estimates of the mean number of events can be used to augment time-to-event analyses where recurrent and competing events exist. The proposed parametric estimator offers estimation of a smooth function across time as well as estimation of different contrasts which is not available using the non-parametric model. A natural extension of the proposed work includes implementations in other model frameworks, particularly for joint frailty models.

[1] Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. Stat Med. 1997;16(8):911-24
[2] Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. Biometrics. 2000;56(2):554-62

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO4: CAUSAL INFERENCE 2

TO4.1 Is inverse probability of censoring weighting a safe alternative to per- protocol analysis?

Xuan J.*[1], Mt--Isa S.[2], Latimer N.[3], Yorke--Edwards V.[1], White I.[1]
[1]MRC Clinical Trials Unit, University College London ~ London ~ United Kingdom, [2]BARDS, MSD ~ Zurich ~ Switzerland, [3]School of Health and Related Research, University of Sheffield ~ Sheffield ~ United Kingdom

Introduction: Intervention deviation (deviation from the assigned treatment) occurs in many clinical trials. When targeting a hypothetical estimand without the occurrence of some or all forms of intervention deviation, per-protocol (PP) analyses are widely adopted. In non-inferiority trials, even without estimands explicitly defined, the common belief that intention to treat (ITT) is anti-conservative leads to the frequent use of PP. However, PP suffers from selection bias when intervention deviation is associated with time-varying confounders that also predict counterfactual outcomes. Inverse probability of censoring weighting (IPCW) is an extended version of PP which removes selection bias by accounting for these confounders. However, the performance of IPCW heavily relies on its "No unmeasured confounders" (NUC) assumption whose plausibility is not statistically testable. In a simulation study, we evaluated the performances of IPCW and PP to explore whether IPCW is a safe alternative to PP. Methods: We simulated datasets with sustained treatment and time-varying covariates informed by a trial in paediatric HIV infection. We varied the prevalence of intervention deviation, whether it occurred in one or both arms, and whether two confounders caused selection bias in the same or opposite direction. The estimand was the risk difference between arms at week 96 if intervention deviation had not occurred. A range of potential scenarios where IPCW omits a confounder, includes the correct confounders or includes unnecessary covariates were designed to emulate potential IPCW implementation in practice. Results: In the presence of selection bias, PP provides biased estimates and IPCW, with its NUC assumption satisfied, provides unbiased estimates. IPCW with an omitted confounder is biased, but less biased than PP in most scenarios, except for an unusual case where the selection bias caused by two confounders is in opposite direction, and 'cancels' out. Including unnecessary covariates in IPCW increases standard errors but to an acceptable extent. Discussion: IPCW with different combinations of covariates outperforms PP in most scenarios. Therefore, when targeting a hypothetical estimand in clinical trials with intervention deviation when selection bias is anticipated, IPCW is a safe alternative to PP to obtain a less biased estimate.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 550-560. Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. American journal of epidemiology, 168(6), 656-664.

TO4.2 Combining sequential stratification and iptw weights to estimate the survival benefit of liver transplantation

Prosepe I.*, Van Geloven N., Putter H.
Leiden University Medical Center ~ Leiden ~ Netherlands

Due to scarcity of donors, patients in need of liver transplant are placed on a wait-list. Recently, organ allocation organisations have expressed interest in allocating donors based on survival benefit, defined as the contrast between the expected survival of wait-listed patients with and without transplant. Estimating survival benefit from observational registry data presents two main challenges: time-dependent confounding and multiple time scales. The time-dependent confounding is results from a combination of biomarkers, summarized through the so-called MELD score, which is highly predictive of wait-list survival as well as the likelihood of receiving a donor-liver. The time axis relevant for creating comparability between treatment groups is time- since-entry on the wait-list while the time axis at which treatment decisions are made is calendar time (when donors become available). Estimates of survival benefit should be applicable to a dynamic population of patients, with some just recently joining the wait-list and others listed for a longer time. In this work, we propose an estimation strategy for dynamically estimating individualized survival benefit for all patients on the wait-list at any time a donor becomes available.
Previous work has proposed sequential stratification to account for the two time axes. In the current work, sequential stratification is combined with adjustments for time-dependent confounding using inverse probability of treatment weighting (IPTW). Differently from previous work, the proposed approach estimates survival benefit via only one marginal structural model, where treatment is a time-dependent covariate. The combination of sequential stratification with IPTW allows for predictions that can be applied repeatedly in calendar time to all patients on the wait-list. We first present simulation results showing that our method can estimate survival benefit on the dynamic population unbiasedly, as opposed to simpler methods. The proposed method is then applied to liver transplant data from the Eurotransplant region. In this real-data application, we show the potential effect of allocating according to survival benefit. Combining sequential stratification and IPTW allows for the dynamic estimation of survival benefit for currently waitlisted patients. This type of predictions can help supporting decisions on which wait-list patients should be prioritized. Q. Gong, Douglas E. Schaubel, Biometrics, 69(2), 2013, 338-347.

TO4.3 Reducing time-lag bias when comparing treated patients to controls with a different start of follow-up

Van Eekelen R.*[1], Bossuyt P.[2], Van Wely M.[2], Van Geloven N.[3]
[1]Amsterdam UMC, location VUmc ~ Amsterdam ~ Netherlands, [2]Amsterdam UMC, location AMC ~ Amsterdam ~ Netherlands, [3]Leiden University Medical Center ~ Leiden ~ Netherlands

In target trial emulation, treated and control patients might have a different time origin i.e. the start of follow-up. This occurs when treatment is started later and it is especially apparent when data on the two conditions is obtained from two separate sources, where the start might be defined implicitly. When ignored, this difference in follow-up start leads to biased estimates for the treatment effect due to time-lag bias of which the core reason is unobserved heterogeneity between patients. Here we study time-lag bias in more depth, discuss terminology, its differentiation from immortal time bias and depletion of susceptibles, formulate estimands following several scenarios and explore methodological solutions. We first conducted a simulation study in which we purposefully introduced time-lag bias, then attempted to remove this bias using five methodological approaches: follow-up as a prognostic covariate, matching (targeting ATE, ATC or ATT) and left truncation. We illustrate that even a small discrepancy in follow-up start can bias estimated treatment effects considerably but also show that the methods adequately remove the time-lag bias. We then apply our methods to an example target trial emulation using data on intrauterine insemination treatment compared to usual care for low-fertility couples, which is a treatment that is commonly initiated much later than the initial diagnosis. As even small discrepancies in the start of follow-up can have a lot of impact on estimates, time- lag bias cannot be ignored in target trial emulation. If the data is available, proper alignment of the time axes solves the issue and allows for valid inference. It is crucial to collect all data necessary to reconstruct the timeline of patients or, when this is not possible, assume a certain time-lag for the treated as a sensitivity analysis.

[1] Aalen OO, Cook RJ, Roysland K. Does Cox analysis of a randomized survival study yield a causal treatment effect? Lifetime Data Anal 2015;21:579-93.
[2] Hernan MA. The hazards of hazard ratios. Epidemiology 2010;21:13-5.

T04.4 Basing discrete event simulators for organ allocation on counterfactual mortality risks

De Ferrante H.*[1], Smeulders B.[1], De Rosner--Van Rosmalen M.[2]
[1]Eindhoven University of Technology ~ Eindhoven ~ Netherlands, [2]Eurotransplant ~ Leiden ~ Netherlands

Discrete event simulators are routinely used for policy evaluation in organ allocation. Such simulators require complete information on future events happening to transplant candidates until waitlist death/removal. In reality, transplantation prevents observation of a complete event set for most candidates. Existing organ allocation simulators such as KPSAM and LSAM therefore match candidates at transplantation to patients at risk with comparable expected mortality risks. We propose to match instead on expected counterfactual mortality risks, which we estimate by correcting for selection bias by transplantation with inverse probability censoring weighting (IPCW).
The procedure we propose is closely related to a multiple imputation procedure for modelling restricted mean log-survival times (RMLST) from pre-determined landmark times [1]. However, we model the RMLST from any moment after listing such that patients without future status updates can always be matched to comparable at-risk, transplant candidates. Two factors complicate directly modelling the RMLST. Firstly, transplantation prevents observation of time-until-event for most transplant candidates. We resolve this by constructing pseudo-observations for the remaining log survival time, where IPCW is used to correct for dependent censoring by transplantation (as in [1]). Secondly, most candidates have multiple status updates, and time-until-events for such status updates are correlated within the individual. To account for such correlations, we estimate a model for the RMLST with Quasi-Least Squares with a Markov correlation structure [2]. For each candidate lacking future status updates, we then construct a risk set of candidates with similar predicted RMLST. Heterogeneity of these risk sets is constrained by matching too on other patient characteristics (e.g., urgency status, disease group, and transplantation history). We use Kaplan-Meier with IPCW to estimate risk-set specific counterfactual survival curves. For discrete event simulations, the candidate is then matched to a candidate in the risk set by inverse transform sampling from these survival curves. Our results show that IPCW increases risk-set specific estimates of 90-day mortality risks on average by 10%. This leads to an additional 10-15 waitlist deaths per year in discrete event simulations of Eurotransplant's liver allocation system. We proposed a procedure to base imputation of future status updates in discrete event simulators on counterfactual mortality risks. For liver transplantation, this meaningfully increases the total number of simulated waitlist deaths.

[1] Nabihah Tayob and Susan Murray. Statistical consequences of a successful lung allocation system - recovering information and reducing bias in models for urgency. Statistics in Medicine, 36(15):2435–2451, July 2017
[2] Jichun Xie, Justine Shults, Jon Peet, Dwight Stambolian, and Mary Frances Cotch. Quasi-least squares with mixed linear correlation structures. Statistics and its interface, 3(2):223–234, 2010.

Parallel Sessions | Tuesday 29 August 2023

T04.5 Continuous-time mediation analysis for repeated mediators and outcomes

Le Bourdonnec K.*[1], Samieri C.[1], Valeri L.[2], Proust--Lima C.[1]
[1]Université de Bordeaux ~ Bordeaux ~ France, [2]Columbia University ~ New-York ~ United States of America

Mediation analysis consists in retrieving the underlying causal mechanisms between an exposure and an outcome, through an intermediate variable called mediator. Initially developed for cross-sectional studies, it has been extended to longitudinal data by discretizing the assessment times of mediator/outcome [1]. Yet, processes in play in longitudinal studies are often defined in continuous time and measured at irregular and subject-specific visits so that discrete-time techniques are not appropriate. This is the case for instance in dementia research when interested in causal mechanisms involving brain lesions and cognitive functioning measured at follow-up visits. Our objective was thus to propose a novel methodology to estimate the causal mechanisms between a time-fixed exposure, a mediator process and an outcome process both measured repeatedly over time.
In the absence of time-dependent confounders, natural direct and indirect effects can be estimated under assumptions including consistency, sequential ignorability and cross-world independence. However in the presence of time-dependent confounders, the latter does not hold anymore and natural effects cannot be identified. As an alternative, we propose to estimate path-specific effects which decompose paths through confounder and mediator or through mediator only. We defined the identifiability assumptions required to get the path-specific effect estimable with continuous-time processes. Then, we used a dynamic multivariate model based on differential equations and mixed effects as a working model to estimate them from repeated data. The method was validated in simulations and applied in a population-based cohort of cerebral aging to investigate the causal pathways between a genetic factor (APOE4) and cognitive functioning through vascular brain lesions and brain atrophy.
By handling the continuous-time nature of phenomena in play, this methodology extends mediation analyses to the longitudinal data usually encountered in health cohort studies. It accounts for both the history of the processes and their measures collected at irregular times.

[1] M.-A. C. Bind, T. J. Vanderweele, B. A. Coull, J. D. Schwartz, Causal mediation analysis for longitudinal data with exogenous exposure, Biostatistics, Volume 17, Issue 1, January 2016, Pages 122–134

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO5: MACHINE LEARNING 2

TO5.1 Ensemble algorithm based on shapley values beyond binary classification: simulations and clinical application

Capitoli G., Bernasconi D.*, Valsecchi M.G., Galimberti S.
Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy ~ Monza ~ Italy

The task of distinguishing between benign and malignant thyroid nodules is of utmost importance to drive appropriate clinical indication (e.g. surgery or follow-up). Machine learning and deep learning classifiers (penalized regression, artificial neural network, decision trees, and support vector machine), based on proteomic mass spectrometry features, were developed. Each algorithm has its own strength and weaknesses and the performance of each single classifier lead to promising but non entirely satisfactory results. To minimize the error rate that can be caused by using a single classifier, the voting ensemble technique can be used, combining the classification results of different classifiers to improve the final classification performance. This work aims to compare the existing voting ensemble techniques (class predicted by the majority of classifiers, averaging the predicted class probability over the classifiers, averaging weighted by accuracy of each classifiers) with a new cooperative-game-theory derived approach based on "shapley values". This method consists in assigning a weight to the prediction of each classifier to any data point accounting for both the classifier performance but also the redundancy of the prediction with respect to other algorithms. Moreover, we extend this methodology to multiclass-classification. We compared the performance of each single classifier and of the different ensemble techniques in a Monte Carlo simulation study, considering different scenarios based on the number of observations, the total number of features and the percentage of relevant features. We observed an improvement of the accuracy for ensemble system based on Shapley values for the binary and multinomial problem respectively, with respect to performances obtained from the best single classifier and from the standard voting systems. On the clinical application, however, when we considered in the ensemble also the algorithms with the lowest accuracy, the performance of the approach based on shapley values was not better than the single best classifier. The ensemble voting algorithm based on shapley values showed good performance compared to more standard ensemble approaches, both for binary and multiclass classification. Selecting reliable and complementary classifiers to be included in the ensemble is needed to improve the final performance.

[1] Capitoli G, Piga I, Clerici F, Brambilla V, Mahajneh A, Leni D, Garancini M, Pincelli AI, L'Imperio V, Galimberti S, Magni F, Pagni F. Analysis of Hashimoto's thyroiditis on fine needle aspiration samples by MALDI-Imaging. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2020, 1868(11):140481.
[2] Rozemberczki B, Watson L, Bayer P, Yang HT, Kiss O, Nilsson S, Sarkar R. The Shapley Value in Machine Learning. arXiv preprint, 2022, arXiv:2202.05594.

TO5.2 Comparison of classification methods for multiplex digital pcr data

Chen Y.*[1], De Spiegelaere W.[1], Trypsteen W.[1], Vynck M.[1], Gleerup D.[1], Thas O.[2]
[1]Ghent University ~ Ghent ~ Belgium, [2]Hasselt University ~ Hasselt ~ Belgium

Digital PCR (dPCR) is a highly accurate technique for the quantification of target nucleic acid(s). dPCR has shown great potential in clinical operations, like tumour liquid biopsy, non-invasive prenatal diagnosis, and validation of biomarkers. The technology proceeds by massive partitioning of the sample into individual partitions, in each of which PCR reactions result in an end-point fluorescence intensities that are subsequently used to classify partitions as positive or negative. We focus on the quantification of multiple target molecules in a single run. The classification of partitions, which is based on clustering methods, is crucial to avoid bias in the estimation of the concentration of the target molecules. We have evaluated many clustering methods, from kmeans, dbscan to specific methods for dPCR and flowcytometry, on both simulated and real-life data. The simulations are based on a mixture distribution of a Poisson point process and a skew-t distribution. This model succeeds well in capturing the irregularities of the cluster shapes and the randomness of partitions between groups ("rain"). The goodness-of-fit of the model was checked by density and depth- depth plots. Various scenarios can be easily generated from the model, including 1) very small groups; 2) non-convex patterns of groups; 3) heavy rain; and 4) bad separation. An adjusted rand index and sensitivity analysis were used for performance evaluation of those clustering methods. Our results show that most of the methods have difficulty in identifying small groups, especially the model-based ones. In general, the presence of "rain" can impede the clustering. Prior knowledge about the cluster centroids really helps in some cases when the group size is not big enough. Based on our extensive comparison of clustering methods for a large variety of scenarios, we describe the limits of these methods, and formulate guidelines to dPCR users to choose methods that suit their data. In addition, we have derived a model that generates realistic dPCR data. The database of experimental dPCR data augmented with the labeled simulated data can serve as training and testing data for new clustering methods.

[1] Baddeley, Adrian, and Rolf Turner, Journal of statistical software, 12, 2005, 1-42.

TO5.3 Comparative analysis of supervised integrative methods for multi-omics data

Broc C.*[1], Novoloaca A.[1], Beloeil L.[1], Yu W.[2], Becker J.[1]
[1]BIOASTER ~ Lyon ~ France, [2]Gates MRI ~ Boston ~ United States of America

Recent advances in sequencing, mass spectrometry and cytometry technologies have enabled researchers to collect multiple omics data from a single sample. These large datasets have led to a growing consensus that a holistic approach was needed to identify new candidate biomarkers and unveil mechanisms underlying disease aetiology, key to precision medicine. While many reviews and benchmarks have been conducted on unsupervised approaches (Bersanelli et al. 2016), their supervised counterpart have received less attention in the literature, no gold standard has emerged yet (Krassowski et al. 2020). In this work, we present a thorough comparison of a selection of five methods, representative of the main families of integrative approaches (matrix factorization, multiple kernel methods, ensemble learning and graph-based methods). As non-integrative controls, random forest was performed alongside, on each data type separately. Methods were evaluated on both simulated and real-world datasets, the latter being carefully selected to cover different medical applications (infectious diseases, oncology and vaccine) and data modalities. A set of nineteen simulations were designed from the real-world datasets to explore a large and realistic parameter space (e.g. dimensionality, confounding effects, effect size). Overall, Integrative approaches showed comparable or higher performances on simulations and outperformed non-integrative methods on real-world data. More specifically, multiple kernel and matrix factorization demonstrated a strong ability to uncover modest effects in high dimensional settings. The strengths and limits of those methods will be discussed into details as well as guidelines for future applications.

Matteo Bersanelli et al. Methods for the integration of multi-omics data:mathematical aspects. BMC Bioinformatics, 17(S2):S15,December 2016.
Michal Krassowski et al. State of the Field in Multi-OmicsResearch: From Computational Needs to Data Mining andSharing. Frontiers in Genetics, 11:610798, December 2020.

Parallel Sessions | Tuesday 29 August 2023

TO5.4 Optimal transport for automatic alignment of non-targeted metabolomic data

Breeur M.*[1], Stepaniants G.[2], Keski--Rahkonen P.[1], Rigollet P.[2], Viallon V.[1]
[1]International Agency for Research on Cancer (IARC) ~ Lyon ~ France, [2]Massachusetts Institute of Technology, Department of Mathematics ~ Cambridge MA ~ United States of America

Untargeted metabolomic profiling through liquid chromatography-mass spectrometry (LC-MS) allows the measurement of a wide range of metabolites in a biospecimen. However, untargeted features measured in untargeted metabolomics studies are only defined through their mass-to- charge ratio (m/z) and retention time (RT) and are therefore not immediately identifiable. Furthermore, m/z and RT measured under different conditions are subject to variations, and features common to two different studies cannot be directly identified. This limitation hampers the external validation of results and more generally the comparison of results across different studies. It also prevents the pooling or meta-analysis of untargeted metabolomics data, thus limiting the statistical power of untargeted metabolomics studies. We developed GromovMatcher, an unsupervised method to automatically match features from two LC-MS untargeted datasets by combining information on their m/z, RTs and signal intensities. GromovMatcher primarily pairs features with compatible signal intensities by making use of the Gromov-Wasserstein distance [1] (an extension of optimal transport designed to couple sets by taking advantage of their structure) between the features within each dataset. An additional constraint allows us to restrict this coupling to pairs of features sharing similar m/z. Finally, the deviation of the RTs between the two studies is estimated to retain only those pairs with compatible RTs in our final matching. GromovMatcher was tested through an extensive simulation study and performed better than similar methods in terms of precision and recall. GromovMatcher also achieved very good performance when applied to real untargeted metabolomics data acquired in sub-studies nested within a large European cohort, where a small subset of features had previously been matched manually by an expert biochemist. Unlike other existing methods, GromovMatcher requires the setting of only a few parameters, and is implemented using an open-source programming language to facilitate its use and possible future developments. Our work could have multiple applications in metabolomics, from the comparison of acquisition protocols to the pooling or meta-analysis of data from different studies. This would allow a better use of the increasingly available non-targeted metabolomic data, for example in cancer epidemiology.

[1] Mémoli, F. Gromov–Wasserstein Distances and the Metric Approach to Object Matching. Found Comput Math 11, 417–487 (2011).

TO5.5 Artificial intelligence for the prediction of weaning readiness outcome in mechanically ventilated patients

Lanera C.*[1], Andrea P.[1], Annalisa B.B.[2], Paolo N.[2], Dario G.[1]
[1]Unit of Biostatistics, Epidemiology, and Public Health - Dep. Cardiac, Thoracic, Vascular Sciences, and Public Health - University of Padova ~ Padova ~ Italy, [2]Department of Medicine - University of Padova ~ Padova ~ Italy

Mechanical ventilation (MV) is a critical intervention for patients experiencing acute respiratory failure. Physicians assess the readiness for MV withdrawal daily through a two-phase process: Readiness Testing (RT) and a 30-minute spontaneous breathing trial (SBT). The outcomes are mutually exclusive: no SBT attempt, SBT failure, or SBT success.[1] Artificial Intelligence models have been shown to predict weaning readiness.[2] On the other hand, there is a lack of methodology and experimentation involving continuous patient signals continuously, and on the side of daily and baseline information. We developed an artificial intelligence (AI) model to predict the day's likely outcome early in the morning, utilizing patient clinical data, previous day's clinical diary information, and minute-by- minute mechanical ventilator parameter recordings. These data were sourced from a retrospective observational multi-center study conducted in Italy over 27 months. The AI model employs a deep learning approach using multi-source neural network topology and multiple recurrent architectures. Hyperparameter optimization and cross-validation were used to select the model, with 36 out of 182 patients reserved for testing final model performance across various metrics, including a custom score emphasizing clinical impact. Our AI model achieved an accuracy of 79% [74, 83%]. This performance surpassed comparison models, such as the XG Boost model trained on the previous day's daily and baseline clinical data, which achieved an accuracy of 61% [56%, 66%]. Our AI model effectively approximates current clinical management by providing early morning predictions for MV withdrawal outcomes. Furthermore, the model's clinical utility has the potential for improvement through the incorporation of additional tailored training data.

[1] Girard TD, Alhazzani W, Kress JP, Ouellette DR, Schmidt GA, Truwit JD, et al. An Official American Thoracic Society/American College of Chest Physicians Clinical Practice Guideline: Liberation from Mechanical Ventilation in Critically Ill Adults. Rehabilitation Protocols, Ventilator Liberation Protocols, and Cuff Leak Tests. Am J Respir Crit Care Med. 2017 Jan 1;195(1):120– 33.
[2] Gallifant J, Zhang J, del Pilar Arias Lopez M, Zhu T, Camporota L, Celi LA, et al. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. Br J Anaesth. 2022 Feb;128(2):343–51.

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO6: CLINICAL TRIALS 5

TO6.1 Sample size estimation for clinical trials using complex responder endpoints

Wason J.*
Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

In some clinical areas it is common for trials to use composite responder endpoints that classify participants as responders or non-responders based on several variables, some of which are continuous. Traditionally these endpoints are analysed as binary, which means a large amount of information is discarded as the continuous variables are dichotomised. Various methods, referred to as augmented binary approaches (e.g. (1–3)), have been proposed to analyse these endpoints more efficiently by utilising the continuous variables to improve the precision. The efficiency gained from doing this varies hugely, depending on the number of continuous components and which of them divide responders from non-responders. There is a need to consider how best to choose the sample size when designing a trial that will use an augmented binary approach as the primary analysis to avoid unnecessarily high sample sizes. I will show how a latent variable model can be used to estimate the probability of response and demonstrate how to conduct sample size estimation using pilot data. The sample size required is reduced by at least 30% compared to using a binary analysis whilst maintaining the type I error rate and power. This reduction can be much higher in some circumstances. As a case study I will discuss how this approach was used in the design of a trial of primary biliary cholangitis to reduce the sample size required and the cost of conducting the trial by £600k. Using composite responder endpoints is common but the traditional methods of analysis mean the sample sizes needed are much higher than they should be. A more efficient analysis can be taken into account at the sample size calculation stage and lead to considerably reduced sample size.

1 Wason JMS, Seaman SR. Using continuous data on tumour measurements to improve inference in phase II cancer studies. Stat Med [Internet]. 2013; Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84885431341&partnerID=MN8TOARS
2 Wason J, McMenamin M, Dodd S. Analysis of responder-based endpoints: improving power through utilising continuous components. Trials. 2020 Dec;21(1):427. 3McMenamin M, Barrett JK, Berglind A, Wason JM. Employing a latent variable framework to improve efficiency in composite endpoint analysis. Stat Methods Med Res. 2021 Mar;30(3):702– 16.

TO6.2 Sample size adaptations in clinical trials comparing restricted mean survival times – advantages and drawbacks

Herrmann C.*[1], Blanche P.[2]
[1]*Charité – University Medicine Berlin, Institute of Biometry and Clinical Epidemiology ~ Berlin ~ Germany,*
[2]*University of Copenhagen, Section of Biostatistics ~ Copenhagen ~ Denmark*

The time until an event is of primary interest in many clinical trials. The frequently applied hazard ratio is not straightforward to interpret under non-proportional hazards. Another challenging point can be the sample size determination for a clinical trial. However, the sample size is of great importance. Imprecise parameter assumptions can lead to under- or overpowering of the trial. Hence, the aim of our research is to consider mid-trial sample size adaptations in clinical trials that compare restricted mean survival times (RMST) in the non-proportional hazards setting, where the RMST describes the mean survival time within a pre-specified interval of interest. More precisely, we apply an adaptive group sequential study design. We assume one interim analysis at a fixed calendar time. The information of the two stages is combined by applying the inverse normal combination test with appropriate adjustment of the local significance levels. We address the pipeline data at the interim analysis by left-truncation at the second stage [1]. The sample size update is based on the conditional power: Under updated distributional parameters, the probability of correctly rejecting the null hypothesis at the end of the trial is calculated. The smallest sample size that fulfills a specific conditional power requirement or a pre-defined maximally feasible sample size is then chosen for the second stage. The performance is evaluated in terms of power and sample size. The approach comes along with advantages and disadvantages. One gains more flexibility in regards to planning a clinical trial that compares RMSTs. The methodology was missing several months ago at the planning stage of a specific cardiologic trial. Moreover, the design offers the potential of saving patients and costs. Nevertheless, it also involves more complicated statistical methods. Addressing sample size adaptations in the time-to-event data setting is not straightforward. However, it comes along with the potential of reducing an often long trial duration. We close a gap in the literature by presenting a sample size adaptation strategy based on the conditional power for trials with the RMST.

[1] N. Keiding, T. Bayer, S. Watt-Boolsen, Confirmatory analysis of survival data using left truncation of the life times of primary survivors, Statistics in Medicine, 6(8), 1987, 939-944.

TO6.3 A hybrid approach to sample size reestimation in cluster randomized trials with continuous outcomes

Sarkodie S.*, Wason J., Grayling M.
Newcastle University ~ Newcastle Upon Tyne ~ United Kingdom

To develop a hybrid (Bayesian-frequentist) approach to sample size reestimation (SSRE) for cluster randomized trials with continuous outcome data, allowing for uncertainty in the intra- cluster correlation (ICC).
We present a general framework for performing SSRE in both frequentist and hybrid settings, assuming normally distributed outcome data and a conventional linear mixed model analysis. Uncertainty in the hybrid framework is captured by placing a truncated normal prior on the ICC and controlling the expected power. A simulation study is used to assess when a hybrid approach may help overcome known issues with the frequentist approach.
On average, both the hybrid and frequentist approaches mitigate against the implications of misspecifying the ICC at the trial's design stage. In addition, both frameworks lead to SSRE designs with approximate control of the type I error-rate at the desired level. An evident problem with the frequentist approach, noted in previous works, is highlighted again in terms of high variability in the reestimated sample size. It is clearly demonstrated how the hybrid approach is able to reduce such variability, with the level of reduction a clear function of the informativeness of the prior assumed for the ICC. The drawback of the hybrid approach comes when there is prior misspecification; it is shown how a highly informative prior quickly results in substantial power loss if the prior is misspecified. In conclusion, a hybrid approach could offer potential advantages in the design and analysis of cluster randomized trials using SSRE. Specifically, when there is available data or expert opinion to help guide the choice of prior for the ICC in the hybrid framework, the hybrid approach can reduce the variance of the reestimated required sample size compared to a frequentist approach. As SSRE is unlikely to be employed when there is substantial such available data (i.e., when a constructed prior is highly informative), the greatest utility of a hybrid approach to SSRE likely lies when there is low-quality evidence available to guide the choice of prior.

1. Friede T, Kieser M. Blinded sample size re-estimation in superiority and noninferiority trials: Bias versus variance in variance estimation. Pharm Stat. 2013;12(3):141-146. doi:10.1002/pst.1564
2. Campbell MK, Grimshaw JM, Elbourne DR. Intracluster correlation coefficients in cluster randomized trials: Empirical insights into how should they be reported. BMC Med Res Methodol. 2004;4:1-5. doi:10.1186/1471-2288-4-9
3. Ip EH, Wasserman R, Barkin S. Comparison of intraclass correlation coefficient estimates and standard errors between using cross-sectional and repeated measurement data: The Safety Check cluster randomized trial. Contemp Clin Trials. 2011;32(2):225-232. doi:10.1016/j.cct.2010.11.001
4. Wu S, Crespi CM, Wong WK. Comparison of Methods for Estimating the Intraclass Correlation Coefficient for Binary Responses in Cancer Prevention Cluster Randomized Trials. Contemp Clin Trials. 2012;33(5):869-880. doi:10.1016/j.cct.2012.05.004.
5. Murray DM, Varnell SP, Blitstein JL. Design and Analysis of Group-Randomized Trials: A Review of Recent Methodological Developments. Am J Public Health. 2004;94(3):423-432. doi:10.2105/AJPH.94.3.423
6. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and health care: A systematic review. Health Technol Assess (Rockv). 1999;3(5). doi:10.3310/hta3050
7. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: Extension to cluster randomised trials. BMJ. 2012;345(7881):1-21. doi:10.1136/bmj.e5661

See the conference APP for the complete list of references

TO6.4 Hybrid sample size calculations for cluster randomised trials using assurance

Williamson S.F.*[1], Wilson K.[2], Tishkovskaya S.[3]
[1]*Biostatistics Research Group, Population Health Sciences Institute, Newcastle University ~ Newcastle ~ United Kingdom,* [2]*School of Mathematics, Statistics & Physics, Newcastle University ~ Newcastle ~ United Kingdom,* [3]*Faculty of Health and Care, Lancashire Clinical Trials Unit, University of Central Lancashire ~ Preston ~ United Kingdom*

Sample size determination for cluster randomised trials (CRTs) is challenging because it requires robust estimation of the intra-cluster correlation coefficient (ICC). Typically, the sample size is chosen to provide a certain level of power to reject the null hypothesis in a two-sample hypothesis test. This relies on the minimal clinically important difference (MCID) and estimates of the overall standard deviation, ICC and, if cluster sizes are assumed to be unequal, the coefficient of variation of cluster size. Varying any of these parameters can have a strong effect on the required sample size. The ICC can be particularly challenging to estimate and, if the value used in the power calculation is far from the true value, can lead to trials which are over- or under-powered.
In this talk, we present a hybrid approach which uses Bayesian assurance (or expected power) to determine the sample size for a CRT in combination with a frequentist analysis. Assurance is a robust alternative to traditional power which incorporates uncertainty on key parameters through prior distributions. We suggest specifying prior distributions for the overall standard deviation, ICC and coefficient of variation of cluster size, while still utilising the MCID. We consider using a joint prior for the ICC and standard deviation to model their dependency. This approach is motivated by a parallel-group CRT in post-stroke incontinence. Although a pilot study was conducted for this trial, the resulting ICC estimate was of low precision and could not be used as a reliable source for the sample size calculation. Instead, we illustrate the effects of redesigning this trial using the hybrid approach, with a prior distribution for the ICC elicited from expert opinion and previous studies. The impacts of misspecifying this prior are considered and the results compared to those obtained from a standard power calculation. The hybrid approach can be applied to prevent incorrectly powered studies resulting from ill- estimated model parameters, to mitigate the impact of uncertainty in the ICC and other design parameters, and to incorporate expert opinion or historical data when designing a CRT. This is particularly important when there is difficulty obtaining reliable ICC estimates. Williamson SF, Wilson KJ, Tishkovskaya SV. Hybrid sample size calculations for cluster randomised trials using assurance. In submission; Available on request.

TO6.5 The anytime-valid logrank test for flexible collaborative meta-analysis and platform trials

Ter Schure J.*[1], Pérez--Ortiz M.[2], Ly A.[2], Grünwald P.[3]
[1]*Department of Epidemiology & Data Science, Amsterdam UMC ~ Amsterdam ~ Netherlands,* [2]*Machine Learning Group, CWI ~ Amsterdam ~ Netherlands,* [3]*Mathematical Institute, Leiden University ~ Leiden ~ Netherlands*

Platform trials and prospective meta-analysis require extensive collaboration. While platform trials can enforce top-down stopping rules, this is difficult in systematic reviews and meta- analysis. Most common approaches to the latter lack sequential analysis and thus do not control type-I errors, even if implemented in living systematic reviews and adaptive approaches to prospective meta-analysis. The remedy can be found in the field of anytime-valid statistics. These are methods based on e-values that provide the flexibility of analysis required for bottom- up collaboration, and – thanks to a rich pallet of multiple testing methods for e-values – the possibility to simplify top-down stopping rules in platform trials. E-value methods were developed for survival analysis following the Cox proportional hazards model – the anytime-valid logrank test – adding to already existing ones for e.g. t-tests[1]. These methods guarantee type-I error control under continuous monitoring with an unlimited horizon; no need for a rigid stopping rule, alpha-spending function or maximum sample size. When setting a stopping rule at a given significance level, the test can be shown to have a rejection region similar to O'Brien-Fleming alpha spending, while adding the potential for 100% power. The anytime-valid logrank test was applied in a collaborative live (ALL-IN) prospective meta-analysis of seven COVID-19 clinical trials[2], which inspired further development for the setting of large and long-term clinical trials in oncology. E-values and anytime-valid confidence intervals proved to be intuitive to monitor in a dashboard by a multitude of collaborators. This encourages adaptation of sample size in planned and ongoing trials. By enabling bottom-up analysis, the anytime-valid logrank test increases flexibility and reduces research waste.

[1] Grünwald P, de Heide R, Koolen W. Safe Testing. Journal of the Royal Statistical Society, Series B (Accepted for publication as a read paper), 2023; Available on arXiv since 2019:1906.07801.
[2] ter Schure J, Ly A, Belin L, Benn C, Bonten M, Cirillo J, et al. Bacillus Calmette-Guérin vaccine to reduce COVID-19 infections and hospitalisations in healthcare workers-a living systematic review and prospective ALL-IN meta-analysis of individual participant data from randomised controlled trials. medRxiv 2022;12.15.22283474.

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO7: SURVIVAL ANALYSIS 5

TO7.1 Flexible parametric accelerated failure time models with cure

Akynkozhayev B.*, Christoffersen B., Clements M.
Karolinska Institutet ~ Stockholm ~ Sweden

Accelerated failure time (AFT) models offer an attractive alternative to Cox proportional hazards models. These models are collapsible and have a pleasant interpretation on the time scale [1]. From practical applications of Crowther et al. [1], we have found several issues. First, flexible parametric models based on splines on the log cumulative hazard scale may not be constrained to be monotone across their support. Second, we represented time-varying acceleration factors using a cumulative formulation, which is difficult to interpret for more than one time-varying effect. Third, the flexible parametric AFT could have poor properties when the data generating mechanism exhibited cure. Fourth, it is unclear whether AFT models can be used with left-truncated data. To address these issues, we have extended the flexible parametric AFT models in three ways: (i) the addition of monotone natural splines; (ii) developed mixture/non-mixture cure models; and (iii) allowed for time-varying acceleration factors, possibly with cure. If the spline expansion is on the cumulative hazard scale, it is strongly preferred that the cumulative hazard is monotone increasing across its support. Monotone natural splines can be implemented using a linear inequality constraint based on the null-space projection matrix. Non- mixture cure models are implemented by a further null-space constraint with zero slope on the right boundary. Compared with non-mixture, the mixture cure models allow for greater flexibility in allowing the cure fraction to vary by covariates. The extension to time-varying acceleration factors can be implemented using Gaussian quadrature. These extensions have been implemented in the rstpm2 package on CRAN. Simulations demonstrated that these extensions had low bias and reasonable coverage for sufficient degrees of freedom. We have extended the flexible parametric AFT models to allow for monotone splines, cure and time-varying acceleration factors. The AFT models may be preferred over proportional hazards when we are concerned with non-collapsibility or estimands that are proportional on the time scale. For left-truncated data, AFT models require untestable assumptions about exposure in the unobserved time period. Further extensions could include random effects for correlated events.

[1] Crowther, M. J., Royston, P. and Clements, M. (2022). A flexible parametric accelerated failure time model and the extension to time-dependent acceleration factors, Biostatistics (Oxford, England) p. kxac009.

Parallel Sessions | Tuesday 29 August 2023

T07.2 Penalized likelihood approach for mixture cure model with interval censoring - an application to thin melanoma

Lo S.N.*[2], Webb A.[1], Ma J.[1]
[1]Department of Mathematics and Statistics, Macquarie University ~ Sydney ~ Australia, [2]Melanoma Institute Australia, The university of Sydney ~ Sydney ~ Australia

Advancements in cancer treatment have resulted in a significant proportion of patients achieving favorable prognosis and even being classified as cured. The standard approach to investigate covariates effect on survival in these cancer survivors is the mixture cure model that divides and models separately patients in two groups: cured (incidence model) and non-cured (latency model). Current methods are limited to right-censored data, and when interval censoring occurs, they may generate biased estimates that lead to erroneous conclusions. Hence, there is a need for alternative methods that can account for interval censoring in mixture cure models. We proposed a new computational approach that can deal with both the cured fraction issues and the interval censoring challenge. To do so, we extended the traditional mixture cure Cox model to accommodate data with partly interval censoring for the observed event times. We developed a new algorithm that directly optimizes the log-likelihood function as opposed to the expectation-maximization (EM) algorithm used in traditional methods. Extensive Monte Carlo simulations demonstrated that the new method outperformed the EM algorithm-based method in terms of bias, variance and coverage probability. When applied to a cohort of thin melanoma (Breslow thickness less or equal 1.0mm) to investigate factors associated with risk of recurrence (either local, regional or distant), the new method showed Breslow thickness was significant in the incidence model but not in the latency model. Ulcerated tumor and location on the head & neck, the leg or the trunk instead of the arm significantly increased the risk of recurrence. Sex was also significantly associated with risk of recurrence, with males in the non-cured population having a significantly lower risk of recurrence than females. Age was not significant in either the incidence or the latency model. The new algorithm estimates the clinically relevant quantities including survival and hazard function plots with point-wise confidence intervals. The asymptotic variance matrices for all the estimated parameters were derived for inference purposes. An R package is now available at GitHub and will be uploaded to R CRAN.

Webb A, Ma J, LO* S.N., Penalized likelihood estimation of a mixture cure Cox model with partly interval censoring-An application to thin melanoma. Stat Med. 2022 Jul 30;41(17):3260-3280. doi: 10.1002/sim.9415

T07.3 Mixture cure semi-parametric accelerated failure time models with partly interval-censored data

Li I., Ma J., Liquet B.*
School of Mathematical and Physical Sciences/ Macquarie University ~ Sydney ~ Australia

In practical survival analysis, the situation of rare events can arise, which means a portion of the population may never experience the event of interest. Under this circumstance, one remedy is to adopt a mixture cure Cox model to analyse the survival data. However, if there clearly exhibits an acceleration (or deceleration) factor among their survival times, then an accelerated failure time (AFT) model will be preferred, leading to a mixture cure AFT model. In this work we consider a penalised likelihood method to estimate the mixture cure semi- parametric AFT models, where the unknown baseline hazard is approximated using Gaussian basis functions. We allow partly interval-censored survival data which can include event times and left-, right-, and interval-censoring times. The penalty function helps to achieve a smooth estimate of the baseline hazard function. We will also provide asymptotic properties to the estimates so that inferences can be made on regression parameters and hazard related quantities. Simulation studies are conducted to evaluate the model performance, which includes a comparative study with an existing method from the smcure R package. The results show that our proposed penalised likelihood method generally outperforms the competing method. To illustrates the application of our method, a real case study involving melanoma recurrence is conducted and presented.

Li, J. and Ma, J. (2020). On hazard-based penalized likelihood estimation of accelerated failure time model with partly interval censoring. Statistical Methods in Medical Research. 29, 3804 – 3817.
Webb, A., Ma, J., and Lo. S. (2022). Penalized likelihood estimation of a mixture cure cox model with partly interval censoring – An application to thin melanoma. Statistics in Medicine pages 1–21.

T07.4 A flexible bayesian prevalence-incidence mixture model for screening Data

Klausch T.*, Coupé V.
Amsterdam University Medical Centers ~ Amsterdam ~ Netherlands

To estimate disease incidence rates in populations using data from screening programs, flexible models are needed that can account for a variety of factors including (a) a wide class of sojourn time distributions, (b) heterogeneity in transition rates by including covariates, (c) unknown disease prevalence at baseline, and (d) imperfect test accuracy. Our proposed Bayesian prevalence-incidence mixture model addresses these needs and extends upon existing models (e.g., [1,2]) by allowing arbitrary parametric sojourn time distributions (including non-proportional and non-constant hazards), co-estimation of test sensitivity, and inclusion of prior information. We demonstrate the method's efficacy with a case study of non-advanced adenoma (nAA) incidence in an international high-risk cohort screened for colorectal cancer (CRC). Each subject has a series of interval-censored screening moments which end when the individual is observed in the disease state or lost to follow-up. The disease status at baseline is unobserved (prevalence). We model the unobserved sojourn time with an accelerated failure time (AFT) model with covariates, which allows, e.g., Weibull, loglogistic, or lognormal sojourn distributions. We estimate the model by a Metropolis-within-Gibbs algorithm that iterates over augmenting the unobserved variables, drawing the model parameters by Metropolis-Hastings, and drawing the sensitivity parameter from its full conditional distribution. We evaluate the method in a simulation study. As CRC screening was done by colonoscopy with a known sensitivity of approximately 0.75, an informative prior, centred at 0.75, was used for this parameter. For all other parameters, weakly informative priors were chosen. Using the Widely Applicable Information Criterion we determined the lognormal model to fit the data better than a Weibull or loglogistic model. The posterior median sensitivity was 0.80 (95% CI: [0.65, 0.92]) and prevalence of nAA was 13% ([6%, 20%]). For a healthy subject at baseline, five- and ten-year posterior median probabilities of progression were 23% ([16%, 30%]) and 43% ([36%, 49%]), respectively, with higher baseline age associated with faster progression. Our AFT model for interval-censored screening data is a flexible method for simultaneous estimation of incidence, baseline prevalence, and test sensitivity. An implementation in R is available in the package BayesTSM.

[1] Cheung LC, Pan Q, Hyun N, et al. Mixture models for undiagnosed prevalent disease and interval-censored incident disease: applications to a cohort assembled from electronic health records. Statistics in Medicine. 2017;36(22):3583-3595. doi:10.1002/sim.7380
[2] Hyun N, Cheung LC, Pan Q, Schiffman M, Katki HA. Flexible risk prediction models for left or interval-censored data from electronic health records. The Annals of Applied Statistics. 2017;11(2):1063-1084. doi:10.1214/17-AOAS1036

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO8: PREDICTION MODELS 2

TO8.1 Comparing uncertainty in individual probability predictions with various models and model average

Wang H.*[1], Wang J.[2], Coupé V.[1]
[1]*Amsterdam University Medical Center ~ Amsterdam ~ Netherlands,* [2]*Utrecht University ~ Utrecht ~ Netherlands*

Clinical prediction models (CPMs) are frequently developed to predict risks of outcomes for patients, while the predicted outcome probability will vary depending on different sample sizes or different model development methods. We refer to this problem as model uncertainty and assess it by the standard deviation of predicted outcome probabilities of CPMs. The objectives of this study: 1) to establish whether increasing sample size of a dataset used for model development reduces prediction uncertainty; 2) to explore the relation between model types and prediction uncertainty; 3) to establish whether prediction uncertainty can be decreased by averaging over different models. Random samples with sample sizes of 500, 1000, and 5000 were drawn from a clinical dataset. Each dataset was bootstrapped 100-fold, with each different model type developed on each of the 100 bootstrapped datasets and subsequently applied to the original dataset to obtain individual risk predictions. The models under investigation included logistic regression (LR), random forest (RF), support vector machine (SVM), XGBoost and the average of all models. We calculated the standard deviation of the predicted probabilities from 100 bootstrap models for each individual and take the average of standard deviations of all individuals as the measure of prediction uncertainty. At a sample size of 500, mean standard deviations for LR, RF, SVM, XGBoost and the average of these models were 0.0234, 0.0417, 0.0050, 0.0384and 0.0223, respectively; at sample size 1000, were 0.0151, 0.0368, 0.0028, 0.0347 and 0.0185, respectively; at sample size of 5000, were 0.0076, 0.0430, 0.0012, 0.0355and 0.0182, respectively. With increasing samples size, the prediction uncertainty decreases for LR, SVM and the average of all models, but not for RF nor XGBoost. Based on the same sample size, SVM has the lowest prediction uncertainty than all other methods. We did not find evidence that prediction uncertainty is related to model type. The model average also provided average level of prediction uncertainty. Fedesoriano 2021, accessed 2 March 2023, < https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

TO8.2 Evaluating the uncertainty of the risk predicted from the two-stage landmarking model

Gu Z.*[1], Wood A.[2], Paige E.[3], Barrett J.[1]
[1]*Medical Research Council Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom,* [2]*British Heart Foundation Cardiovascular Epidemiology Unit, University of Cambridge ~ Cambridge ~ United Kingdom,* [3]*Australian National University ~ Canberra ~ Australia*

Systematically identifying people at high risk of developing cardiovascular disease (CVD) using electronic health records (EHRs) has the potential to target limited healthcare resources more efficiently to those in need. Typically risk prediction models only make point estimates of risk, and the validity of risk prediction models is assessed using population-level statistics that measure calibration and discrimination. Little attention has been paid to the uncertainty of the predicted risk, which is important for evaluating the reliability of prediction at individual level. In this study, we aim to: (1) quantify the uncertainty of the predicted risk in the two-stage landmarking model, which has been recently used for risk prediction utilizing historical data in EHRs; and (2) explore the extent of uncertainty around the predicted risk in the primary care EHRs from the Clinical Practice Research Datalink (CPRD) in England. The two-stage landmarking model predicts CVD risk using a Cox proportional hazards model utilizing the predicted random effects from a multivariate linear mixed-effects model in the first stage. The uncertainty of the predicted risk is aggregated from two sources. The first is the variability of model estimation, including the parametric parameter estimation of the mixed- effects model and the Cox model, and the estimation of the non-parametric cumulative baseline hazard. The second is the variance of the predicted random effects of each individual, which can be the main contributor of uncertainty when the repeated measurements are sparse. We propose a procedure that sequentially samples parameter estimates for the mixed model, predicted random effects, and parameter estimates for the Cox model to take into account the uncertainty associated to the parametric model parameters and the prediction of random effects, combining with a perturbation-resampling method to take into account the variability associated with the cumulative baseline hazard. We conduct simulation studies to validate the evaluated uncertainty and apply the method to the CPRD data. Our proposed procedure evaluates the uncertainty of the predicted risk, which should be considered alongside the point estimates when identifying high-risk individuals.

Paige, E., Barrett, J., Stevens, D., Keogh, R. H., Sweeting, M. J., Nazareth, I., Petersen, I., & Wood, A. M. (2018). Landmark Models for Optimizing the Use of Repeated Measurements of Risk Factors in Electronic Health Records to Predict Future Disease Risk. American Journal of Epidemiology, 187(7), 1530–1538. https://doi.org/10.1093/aje/kwy018
Ferrer, L., Putter, H., & Proust-Lima, C. (2019). Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. Statistical Methods in Medical Research, 28(12), 3649–3666. https://doi.org/10.1177/0962280218811837

Parallel Sessions | Tuesday 29 August 2023

TO8.3 A simulation approach to calculating minimum sample sizes for Prediction modelling: the pmsims package for r

Carr E.*[1], Forbes G.[1], Shamsutdinova D.[1], Stahl D.[1], Zimmer F.[2]
[1]*Department of Biostatistics & Health Informatics, King's College London ~ London ~ United Kingdom,* [2]*Department of Psychology, University of Zurich ~ Zurich ~ Switzerland*

Developing prediction models with small samples leads to poor performance and imprecise estimation of model parameters. Ensuring sample sizes are sufficient before model development would reduce research waste and improve patient outcomes by avoiding models developed with inadequate samples. Guidance for minimum sample sizes has been developed using analytical methods for linear, logistic, or time-to-event outcomes [1,2]. However, these approaches cannot be easily adapted to more complex scenarios. Several studies have used simulation to estimate minimum sample sizes, for example, using penalised logistic regression [3] or clustered data [4]. However, no generalised, simulation-based, framework or software package currently exists. Here, we introduce an efficient and flexible framework to derive minimum sample sizes for prediction modelling, implemented as a user-friendly R package. Our approach is generalisable to any data type (e.g., longitudinal or clustered) or modelling strategy (e.g., gradient boosted trees). Our workflow starts with specification of (i) a data-generating function, (ii) a modelling function, (iii) the expected model performance, and (iv) the range of sample sizes to evaluate. We then use simulation to identify the minimum sample size that replicates the large-sample model performance to within a specified precision threshold (e.g., within 0.05 of the expected AUC). We adopt a surrogate modelling approach proposed by Zimmer [5] that uses Gaussian process regression to efficiently search for the minimum sample size, greatly reducing the number of simulations required.
Our workflow is implemented as a user-friendly R package. The package offers convenient defaults for common scenarios (e.g., linear or logistic models) as well as the ability to specify user-written data-generating functions and models, capturing varied data and model types (e.g., longitudinal or clustered data). This presentation will demonstrate our package and draw comparisons with existing approaches (e.g., pmsampsize [6]). Our package provides user-friendly tools to efficiently derive minimum sample sizes for prediction modelling. By offering a flexible simulation framework that can accommodate various data and model types, our approach is more widely applicable than existing analytical tools. This will contribute to improved research design and reduced research waste.

[1] Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. Stat Med. 2019;38(7):1262–75.
[2] Riley RD, Snell KI, Ensor J, Burke DL, Jr FEH, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38(7):1276–96.
[3] van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. Stat Methods Med Res. 2019 Aug;28(8):2455–74.
[4] Wynants L, Bouwmeester W, Moons KGM, Moerbeek M, Timmerman D, Van Huffel S, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. J Clin Epidemiol. 2015 Dec;68(12):1406–14.
[5] Zimmer F, Debelak R. Simulation-based Design Optimization for Statistical Power. Utilizing Machine Learning [Internet]. PsyArXiv; 2022 [cited 2023 Mar 31]. Available from: https://psyarxiv.com/tnhb2/
[6] Ensor J, Martin EC, Riley RD. pmsampsize: Calculates the minimum sample size required for developing a multivariable prediction model [Internet]. 2022. Available from: https://CRAN.R-project.org/package=pmsampsize

TO8.4 Synthesis calibration curves

Munoz J.*, Debray T., De Jong V.
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht ~ Utrecht ~ Netherlands

The performance of diagnostic or prognostic models can be quantified in terms of discrimination, i.e., the ability of the model to discriminate subjects into groups, and calibration, the agreement between the risk estimated by the model and the observed proportion of an outcome. Incorrect calibration could adversely affect clinical decision making, e.g., in relation to treatment decisions. To communicate the calibration of a model over the range of estimated risks, calibration curves can be used, as this visualization tool allows to identify not only the direction of decalibration, but also the regions of decalibration, which is useful for recalibration adjustments. When multiple data sources are used, e.g., meta-analyses of individual participant data or multiple imputed data sets, the setup of calibration curves and corresponding confidence intervals becomes complicated.
We describe possible methods for constructing calibration plots when multiple data sources are used. These include methods that ignore clustering, those that pool curves from each cluster where subjects are divided into strata based on predicted risk, and others where calibration curves are estimated in each cluster using regression models and then the information from the models is pooled. These methods were compared in terms of bias, coverage, and length of the 95% confidence interval at marginal level and conditional on the predicted risk. Methods that ignored clustering had little bias but did not provide nominal confidence interval coverage. Some, but not all, methods that accounted for clustering had little bias. Pointwise meta-analysis method which pool cluster-specific predictions of the observed proportion and their standard errors provided less bias with a coverage close to the nominal level. We have identified and described multiple methods for constructing calibration plots when using multiple data sources. The point-wise meta-analysis method was the most favourable in terms of both bias and coverage in scenarios with heterogeneous calibration curves across clusters. Austin, P. C., & Steyerberg, E. W. (2019). The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Statistics in medicine, 38(21), 4051-4065.
Belias, M., Rovers, M. M., Hoogland, J., Reitsma, J. B., Debray, T. P., & IntHout, J. (2022). Predicting personalised absolute treatment effects in individual participant data meta-analysis: An introduction to splines. Research Synthesis Methods, 13(2), 255.

TO8.5 Accounting for missing values in the calibration and application of prediction models

Bart M.*
Leiden University Medical Centre ~ Leiden ~ Netherlands

We consider the calibration of prediction models when missing values are present in both the calibration data as well as in the future patient records to which the predictor must be applied. We study the application of multiple imputation in this context to account for the presence of missing values, both in the calibration of the predictor itself, as well as for the imputation of missing data in future patient predictor records. We compare with complete-case only based prediction calibration.
We contrast two distinct prediction calibration approaches with imputation, (1) the first of which applies Rubin's rules to estimate a single pooled model for prediction, (2) while the second averages the distinct predictions obtained from direct application of the prediction models which are fit on the multiply-imputed datasets. The theoretical foundational background to both approaches is reviewed and discussed. Substantive reductions in variability of predictions are achieved using the prediction-averaging method (2) as compared to Rubin's rules based implementations, with little difference in bias between both methods. Simulations are used to demonstrate how, in addition to higher variability, complete-case-based predictive calibration may often also suffer from substantial biases, especially with moderate-to-small sample sizes and higher levels of missing data. Two real datasets are used to illustrate and compare method performance, in addition to a new test set we recently obtained for additional verification. We focus on the prediction of binary outcome using logistic regression, but our study has shown results carry over generally, including for censored survival outcome.
In the presence of missing values, predictive-averaging should preferentially be applied with multiple imputation as compared to Rubin's rules or complete-case-only calibration. Complete- case-based calibration should be treated with caution in prediction applications. Our results show that substantial numbers of imputations – typically closer to 1000s - may be needed in prediction applications. Single-imputation implementations perform especially poorly in the prediction context.
Mertens, B.J.A., Banzato, E. and de Wreede, L.C. (2020) Construction and assessment of prediction rules for binary outcome in the presence of missing predictor data using multiple imputation and cross-validation: Methodological approach and data-based evaluation. Biometrical Journal, 62,3, 724-741

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO9: LONGITUDINAL ANALYSIS 2

TO9.1 A weighted quantile sum regression with penalized weights and twoIndices

Renzetti S.*[1], Gennings C.[2], Calza S.[1]
[1]Università degli Studi di Brescia ~ Brescia ~ Italy, [2]Icahn School of Medicine at Mount Sinai ~ New York ~ United States of America

New statistical methodologies were developed in the last decade to face the challenges of estimating the effects of exposure to multiple chemicals. Weighted Quantile Sum (WQS) regression is a recent statistical method that allows estimating a mixture effect associated with a specific health effect and identifying the components that characterize the mixture effect. In this study, we propose an extension of WQS regression [1] that estimates two mixture effects of chemicals on a health outcome in the same model through the inclusion of two indices with the introduction of a penalization term. To evaluate the performance of this new model we performed both a simulation study and a real case study where we assessed the effects of nutrients on obesity among adults using the National Health and Nutrition Examination Survey (NHANES) data. The method showed good performance in estimating both the regression parameter and the weights associated with the single elements when the penalized term was set equal to the magnitude of the Akaike information criterion of the unpenalized WQS regression. The two indices further helped to give a better estimate of the parameters (Positive direction Median Error (PME): 0.017; Negative direction Median Error (NME): -0.023) compared to the standard WQS (PME: -0.141; NME: 0.078). In the case study, WQS with two indices was able to find a significant effect of nutrients on obesity in both directions identifying caffeine and magnesium as the main actors in the positive and negative association respectively. Through this work, we introduced an extension of the WQS regression that showed the possibility to improve the accuracy of the parameter estimates when considering a mixture of elements that can have both a protective and a harmful effect on the outcome; and the advantage of adding a penalization term when estimating the weights.

[1] Carrico, C., et al., 2015. Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. J. Agric. Biol. Environ. Stat. 20, 100–120.

TO9.2 Flexible parametric regression for correlated data with transformation models

Balint T.*
University of Zurich ~ Zurich ~ Switzerland

Regression models are widely used in the analysis of experimental and observational data. However, traditional parametric approaches often assume a predefined distribution type, which can be difficult to choose in practice and lead to misspecification issues. Additionally, correlated observations and non-linear predictor-outcome relationships are common in clinical research, and censored and truncated data can further complicate analysis. While regression models should address all these issues for accurate inference and prediction, most available implementations only partially handle them. Transformation models offer a general framework for regression modeling that approximates the conditional response distribution in a data-driven way. They allow for likelihood contributions of various types of observations, including continuous, discrete, censored and truncated data, making them applicable for modeling any at least ordered outcome types. Transformation models have a modular structure, with many variants and well-known special cases available. This talk presents a fully parametric, likelihood-based approach to extend transformation models with random effects and penalized additive terms. The accompanying software [1] is built on well-known and efficient packages of mixed-effects and additive models, resulting in a fast and versatile implementation for a wide set of regression problems. An example application is provided using one-stage individual participant data meta-analysis of a collection of longitudinal studies on burn patient recovery with bounded quality-of-life responses. Mixed-effects additive transformation models provide a flexible, distribution-free framework for the regression analysis of complex, dependent data. The proposed approach and accompanying software offer a valuable tool for addressing common issues encountered in regression modeling, such as correlated observations and censored data. The example application demonstrates the potential benefits of using these models in clinical research, and their wide applicability suggests they may be useful in many other fields as well.

[1] Tamási, Bálint, and Torsten Hothorn. "tramME: Mixed-Effects Transformation Models Using Template Model Builder." The R Journal 13, no. 2, 2021: 398–418.

TO9.3 Comparison of conditional and marginal means in distribution based marginalized multilevel models

Lang Z.*[1], Ózsvári L.[2], Veres K.[1]
[1]Department of Biostatistics, University of Veterinary Medicine Budapest ~ Budapest ~ Hungary, [2]Department of Veterinary Forensics and Economics, University of Veterinary Medicine Budapest ~ Budapest ~ Hungary

We introduce distribution based marginalized multilevel models (DMMM), which are a special class of marginalized multilevel models[1] (MMM) for analyzing hierarchical clustered data. MMMs assess marginal, population level and conditional, cluster specific means simultaneously. In DMMMs the probability distributions of the conditional means belong to a given parametric family. This approach allows for simple and flexible control over the marginal and conditional means. We provide methods and conditions for determining the increase or attenuation of conditional coefficients relative to marginal coefficients in DMMMs, extending the relationship known for the probit-normal and logit-normal models[2]. We introduce a transformation of conditional means in DMMMs, which separates marginal and conditional effects in a locally additive manner on the same scale. Based on this decomposition we form sufficient conditions of increase or attenuation of conditional coefficients relative to marginal coefficients in the linear predictors. A factor of proportionality of the ratio of conditional and marginal coefficients is defined. Several example models are discussed, including the well- known attenuation of the marginal coefficients relative to conditional coefficients in probit-normal and logit-normal models. We describe a DMMM where the conditional means follow Pareto distributions of the second kind with a location parameter, and where the coefficients of the conditional model are attenuated relative to the marginal model. An illustrative application to records of milk ELISA diagnostic tests of dairy cows exposed to paratuberculosis in intensive cattle farms is presented. Marginally, the prevalence of infection is modelled with the age at first calving using logit link. Here, we fit the beta distribution to herd-specific prevalence and show the attenuation of marginal, population-level coefficients compared to conditional, herd-level coefficients. The transformations of the conditional means eliminate the influence of the random effects from the difference of the transformed conditional means. Consequently, the conditional and marginal coefficients of the DMMM can be compared without respect to the actual magnitude of the random effect. The relative increase or attenuation of the conditional and marginal coefficients can add to the comparative analysis of the conditional and marginal components of the DMMM.

[1] M. Griswold, B. Swihart, B. Caffo, S. Zeger, Stat, 2(1), 2013, 129-142.
[2] S. Zeger, K. Liang, P. Albert, Biometrics, 44, 1988, 1049-1060.

Parallel Sessions | Tuesday 29 August 2023

TO9.4 Use of priors in automated model building strategies for nonlinear mixed effects models

Prague M.*[1], Lavielle M.[2]
[1]Inria SISTM, Inserm Bordeaux Population Health ~ Bordeaux ~ France, [2]Inria XPOP, Ecole Polytechnique ~ Paris ~ France

Nonlinear mixed effects models are widely used in life sciences [1]. Automated model building strategies have been developed to efficiently select the "best model" from a large set of candidate models. Most of these methods, such as stepwise covariate modeling method (SCM), COnditional Sampling use for Stepwise Approach based on Correlation tests (COSSAC) consider only the covariate model. We recently developed the Stochastic Approximation for Model Building Algorithm (SAMBA) algorithm [2], a fast and efficient algorithm for building the complete statistical model (including variability, correlations and error models) by minimizing an information criteria (such as corrected BIC [3]). Iteratively, the individual parameters are sampled from the conditional distribution defined under the current model and used in combination with the observed data to select a new statistical model. Minimizing a given information criterion can lead to both false discoveries and false nondiscoveries. A "discovery" for a mixed effects model is an element of the statistical model, such as a relationship between a covariate and an individual parameter, the variability of a parameter or a correlation between parameters. Then, a clinical use of such a model may require controlling the respective rates of these two types of error. In screening studies, the goal is to detect as many true positives as possible even at the price of false positives. In contrast, when the goal is personalized medicine, false positives should be avoided in order to prevent erroneous predictions. In this work, we show how the use of prior knowledge on known mechanistic relationships can help in customizing a parameter-specific penalization in the objective criterion. In doing so, a posterior distribution is implicitely defined over the set of models, which can be either maximized or sampled. This method has been implemented in the R package Rsmlx (interfacing Monolix) implementing the SAMBA algorithm. We use both simulated data and real word applications (pharmacokinetics of cisplatine, vaccine pre-clinical study) to investigate the impact of the penalization term and the use of priors on the selected model.

[1] Lavielle. Mixed effects models for the population approach: models, tasks, methods and tools. CRC press, 2014.
[2] Prague and Lavielle. Samba: A novel method for fast automatic model building in nonlinear mixed-effects models. CPT:PsP, 2022.
[3] Delattre et al. A note on BIC in mixed-effects models. Electron. J. Statist.

TO9.5 Variable selection with 'too' many zero-inflated predictors: a Nonnegative garrote approach

Gregorich M.G.*, Kammer M., Heinze G.
Medical University of Vienna ~ Vienna ~ Austria

Fitting parsimonious prediction models with mass-spectronomy features combines the problem of high-dimensional variable selection with the presence of excess zeros in the predictors. Generally, the bimodal distribution of a zero-inflated feature demands that a prediction model should incorporate parameters for both of its two components: the binary component indicating the presence or absence of a measurable intensity value and the continuous component representing the (probably log2-transformed) intensity. We evaluate three possible approaches to explicitly address this problem when using penalized likelihood for variable selection. Each of the approaches involves a two-step estimation. In the "Lasso-Ridge" approach the first step uses a Lasso penalty applied to the continuous components, and a ridge penalty to the selected features now represented by both components. The "Ridge-Lasso" approach first uses a ridge penalty on the doubly represented features and uses its estimates to define joint penalty factors for the two components in the subsequent Lasso step. This approach is a special version of the adaptive Lasso [1]. (iii) The "Ridge- Garrote" approach also starts with a ridge penalty on the doubly represented features, but in the second step applies the Garrote [2] with joint shrinkage factors for the two components of each feature. We explain details of implementation of the three approaches by means of a real data application comprising 1333 proteomic variables used for the prediction of kidney function in 3,210 persons with chronic kidney disease. While the Lasso-Ridge and the Ridge-Garrote approaches guarantee that in the final model, each selected feature is represented with its binary and its continuous component, the Ridge-Lasso approach also selected features with only one component included. However, the Lasso-Ridge method ignores the zero inflation in the initial estimates resulting in suboptimal selection. The Ridge-Garrote model was as predictive as a random forest but has the advantage of full transparency and explainability of the roles of each feature in the final model. Extensions to accommodate nonlinear feature effects are straightforward. The Ridge-Garrote is a transparent, explainable, powerful, and extendable method to fit prediction models with zero-inflated predictors.

[1] Zou, H., The adaptive lasso and its oracle properties. Journal of the American statistical association, 2006. 101(476): p. 1418-1429.
[2] Breiman, L., Better subset regression using the nonnegative garrote. Technometrics, 1995. 37(4): p. 373-384.

PARALLEL SESSION TO10: CLINICAL TRIALS 6

TO10.1 Marginal odds ratios for cluster randomised trials: a novel analysis method

Thompson J.A.*[1], Kahan B.[2], Copas A.[2]
[1]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [2]University College London ~ London ~ United Kingdom

A cluster-level analysis is common in cluster randomised trials due to robust performance with a small number of clusters or because data are only available at the cluster-level. The method involves using the summary value (mean, rate, or proportion, depending on outcome type) for each cluster, transforming these with a suitable function if wanting an effect estimate other than the mean difference, and then testing for a difference between trial arms using a t-test. However, for non-collapsible effects measures such as an odds ratio, this method targets a cluster-specific effect.
In this talk we will introduce a novel alternative approach for the cluster-level analysis which targets a marginal effect. This approach uses a generalised linear model (GLM) with a link function used to transform the summary values in place of a t-test of transformed values. We present results of a simulation study assessing the performance of this method compared to the t-test approach to estimate an odds ratio. We simulated data from cluster randomised trials with 10-50 clusters, an ICC of 0.05 or 0.15, and constant or varying cluster size. Each trial was analysed with a cluster-level GLM with a logit link and either a normal or gamma distribution, a t- test of cluster log-odds, and independence estimating equations with a logit link and binomial distribution using a Fay and Graubard small sample standard error correction. We show that the cluster-level GLM targets the marginal odds ratio and the cluster-level t-test targets the cluster-specific odds ratio. 95% confidence interval coverage was nominal in most scenarios using the cluster-level t-test and cluster-level GLM with a normal distribution but was often greater than 95% using a cluster-level GLM with a gamma distribution and independence estimating equations. Power was lower using independence estimating equations than the t-test or GLM methods. Using a GLM with a logit link and normal distribution is a robust method of analysis to estimate a marginal odds ratio in a cluster randomised trial.

TO10.2 Pseudo-values regression for restricted mean survival time in small sample cluster randomized trials

Le Vilain--Abraham F.*[1], Tavernier E.[1], Dantan E.[2], Desmée S.[1], Caille A.[1]
[1]Université de Tours, Nantes Université, INSERM SPHERE ~ Tours ~ France, [2]Nantes Université, Université de Tours, INSERM SPHERE ~ Nantes ~ France

Time-to-event outcomes are not rare in cluster randomized trials (CRTs), in which social units are randomized, introducing a correlation between the individual outcomes within a cluster [1]. In a previous work, we proposed the use of pseudo-values regression for estimating difference in restricted mean survival time (ΔRMST) between the intervention and control groups up to time t*. It consists in computing pseudo-values for each individual and considering them as the dependent variable of a linear model fitted by generalized estimating equations (GEE). The simulation study showed this method requires at least 50 clusters to control the type I error rate, but can be corrected using a permutation test. Nevertheless, permutation test only provides p-value, no variance estimator. Here we aim to compare several approaches to correct the variance estimator in pseudo-values regression for ΔRMST in CRTs with a small number of clusters. We evaluate the performance of four bias-corrected variance estimators developed by Mancl and DeRouen (2001), Kauermann and Carroll (2001), Fay and Graubard (2001) and Morel et al. (2003). Additionally, we combine these small-sample corrections with a Student distribution to account for the variability of the standard error estimator, instead of the normal distribution of the Wald test statistic. We compare the methods by using a simulation study under several scenarios (number of clusters, mean cluster size, coefficient of variation of the cluster sizes, intervention effect, degree of clustering). The relative bias in estimating the variance, in absolute value, for the four bias-corrected variance estimators, does not exceed 10% in most of the scenarios, except for the Mancl and DeRouen estimator with a very limited number of clusters (K = 10). The bias-corrected variance estimators combined with the Student distribution show better performance compared to the normal distribution. The type I error rates are closer to the nominal level and the coverage rates are closer to 95%. This work opens the way for estimating a ΔRMST in presence of clustered time-to-event. We specifically propose the use of pseudo-values regression to correctly assessed the intervention effect and its variance with CRTs with a limited number of clusters.

[1] A. Caille, E. Tavernier, M. Taljaard, et al., Methodological review showed that time-to-evnt outcomes are often inadequately handled in cluster randomized trials, J Clin Epidemiol, 134, 2021, 125-37.

Parallel Sessions | Tuesday 29 August 2023

TO10.3 Optimal staircase designs and when to use them

Grantham K.*[1], Forbes A.[1], Hooper R.[2], Kasza J.[1]
[1]Monash University ~ Melbourne ~ Australia, [2]Queen Mary University of London ~ London ~ United Kingdom

Staircase designs are emerging as an efficient cluster randomised trial design for testing interventions applied at the cluster level that cannot be removed once implemented, where a stepped wedge design may otherwise be used [1]. Visually, the trial design resembles a staircase: clusters are randomly assigned to sequences made up of a limited number of measurement periods (control periods followed by intervention), where sequences start measurement at different times. Recent work has found the basic staircase design, which has just one control period followed by one intervention period in each sequence, to be a particularly lean design with power that can rival that of the stepped wedge in certain situations [1]. However, to meet or exceed the efficiency of a stepped wedge, a basic staircase design typically requires clusters to measure more participants in each period, or the inclusion of additional clusters. In this talk we aim to find optimal staircase designs among those with more than two measurement periods in each sequence that could more closely rival the stepped wedge. We will identify optimal staircase designs via the variance of the treatment effect estimator using a linear mixed model under different trial settings. We will examine whether there is a benefit to having different numbers of control and intervention periods in a sequence, moving beyond the basic staircase design. We will also relax the previous assumption of a fixed number of clusters assigned to each treatment sequence to determine the optimal allocation of clusters to sequences for different trial configurations. Surprisingly, our results show that balanced designs are not always optimal, for certain common trial configurations and modelling assumptions. Furthermore, in a result mirroring that of the stepped wedge literature [2], the optimal allocation of clusters to sequences in a staircase design does not assign the same number of clusters to each sequence. The most efficient staircase design for a given trial setting depends on a number of factors. Rather than opt for an existing standard cluster randomised trial design such as a stepped wedge, trialists may follow the guidance from this work to select an efficient staircase design.

[1] K.L. Grantham, A.B. Forbes, R. Hooper, J. Kasza. The staircase cluster randomised trial design: a pragmatic alternative to the stepped wedge. Under review.
[2] J. Lawrie, J.B. Carlin, A.B. Forbes. Optimal stepped wedge designs. Statistics & Probability ers 2015; 99: 210-214.

TO10.4 Finding cost-efficient incomplete stepped wedge designs using an iterative approach

Rezaei--Darzi E.*, Kasza J., Forbes A.B., Grantham K.L.
Monash University ~ Melbourne ~ Australia

Standard stepped wedge trials, in which clusters switch from the control to the intervention condition in a staggered manner, can be costly and burdensome to both clusters and individuals. In these designs each cluster provides data in each period. Recent work has investigated the iterative removal of cluster-period cells of the design that contribute relatively small amounts of information, producing a sequence of candidate incomplete designs [1]. Many of these incomplete designs retain high power to detect effects of interest. In this talk, we extend this work to investigate the cost-efficiency of these designs, seeking to identify incomplete stepped wedge designs that retain high power while limiting trial costs. We provide a framework that incorporates the costs per cluster and per individual, and of restarting data collection in a cluster following a pause in data collection.
We consider linear mixed models for continuous outcomes, with constant cluster-period sizes, categorical period effects, and assume repeated cross-sectional sampling and a discrete-time decay within-cluster correlation structure. We obtain progressively reduced stepped wedge designs by iteratively removing pairs of cells with the lowest contribution to the precision of the treatment effect estimator. For each of the incomplete designs we then assess the total cost, variance of the treatment effect estimator and study power. We show that incomplete designs with approximately half the number of cluster-period cells as a complete design tend to have minimal reductions in precision and study power for a substantially lower total cost. However, for incomplete designs in which a cluster has a pause in data collection, total costs may not necessarily decrease with progressively reduced designs.
Our methods enable trialists to examine the trade-off between the total trial cost and the power of an incomplete stepped wedge design. Designs incorporating only half as many cluster-period cells may be preferable to complete designs - they reduce both the data collection burden and the study costs, while having the potential to maintain high levels of power. "Staircase"-type designs, where clusters only contribute measurements immediately before and after the treatment switch are indicated as particularly cost-efficient variants of the stepped wedge design.

[1] Rezaei-Darzi E, Grantham KL, Forbes AB, Kasza J. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. Under review. 2023.

TO10.5 Joint modelling for phase III clinical trial primary endpoint estimation: Simulation study and application

Pitoy A.*[1], Desmée S.[2], Thai H.[1], Cerou M.[1], Semiond D.[4], Veyrat--Follet C.[1], Bertrand J.[3]
[1]Translational Disease Modelling Oncology, Sanofi ~ 91380 Chilly-Mazarin ~ France, [2]Université de Tours, Université de Nantes, UMR 1246 SPHERE, INSERM ~ 37000 Tours ~ France, [3]Université de Paris, UMR 1137 IAME, INSERM ~ F-75018 Paris ~ France, [4]Sanofi Translational Medicine & Early Development ~ Cambridge ~ United States of America

Nonlinear joint modelling has been increasingly used to characterize the relationship between biomarker kinetics and time-to-event in therapeutic development [1]. At the individual level, such models have been shown to improve patient follow-up by providing dynamic predictions. However, at the population level, the benefit of joint modelling to inform and support decision-making remains to be assessed. Here, we evaluated the ability of a joint model selected from Phase I/II clinical trials data to provide an earlier estimate of the final hazard ratio of a subsequent Phase III study in interim analysis context, compared with Cox and classical parametric models. We first performed a simulation study followed by an application on Phase III data.
Using a Claret tumor growth inhibition model for the biomarker longitudinal data and a parametric proportional hazard model with slope of the biomarker as link function for the survival data, we simulated 100 clinical trials under 2 scenarios to estimate type I error and power according to a spending function (e.g., Pocock, O'Brien and Fleming, Haybittle-Peto) at interim and final analyses. All the approaches in this study are controlling for the type I error. Assessment of the power is ongoing, but we expect to better detect treatment effect using longitudinal information through joint model. We also conducted a retrospective analysis on 256 serum M-protein patients from the Phase III ICARIA-MM clinical trial (NCT02990338) that compared progression-free survival (PFS) of isatuximab (anti-CD38 monoclonal antibody) plus pomalidomide/dexamethasone versus pomalidomide/dexamethasone alone in patients with relapsed and refractory multiple myeloma [2]. Using a joint model previously developed from Phase I/II data, we applied approaches on final analysis data at 162 PFS events and on interim analyses data retrospectively scheduled when 33% and 65% of the expected PFS events occurred. Joint modelling allowed for an earlier estimate of PFS improvement in the isatuximab arm from the second interim analysis data. This work encourages the systematic development of joint model using data from early phases of clinical trials to guide decision-making during Phase III.
Funding: Sanofi

[1] Kerioui, M., Bertrand, J., Bruno, R., et al. Modelling the association between biomarkers and clinical outcome: An introduction to nonlinear joint models. Br J Clin Pharmacol. 2022; 88: 1452-1463.
[2] Attal, M., Richardson, P.G., Rajkumar, V. S., et al. Isatuximab plus pomalidomide and low- dose dexamethasone versus pomalidomide and low-dose dexamethasone in patients with relapsed and refractory multiple myeloma (ICARIA-MM): a randomised, multicentre, open-label, phase 3 study. Lancet. 2019; 394: 2096–2107.

Parallel Sessions | Tuesday 29 August 2023

PARALLEL SESSION TO11: ITR-IBS & ITALIAN STATISTICAL SOCIETY

TO11.1 Exploring the relationship with the digital self-image: integrating model-based clustering and graphical model approaches

Chiara Brombin*[1], Federica Cugnata[1], Clelia Di Serio[1]
[1]*University Center for Statistics in the Biomedical Sciences (CUSSB), Faculty of Psychology, Vita-Salute San Raffaele University, Italy*

Introduction and Objectives: Digital revolution has deeply changed not only the way people interact with each other but also the relationship with the self-image. Increased data availability and computational power have significantly improved algorithms for facial feature detection which have been also successfully applied to develop ("beautifying") face filters/apps. Potential of these filters in altering facial appearance has raised concerns not only among digital technology experts but also in health professionals as they promote unrealistic beauty standards. This is especially crucial when it comes to younger users with low awareness in the use of digital technologies. However, in psychological and social phenomena, behaviors are rarely homogeneous and properly modelling data heterogeneity is crucial to derive unbiased results and to generalize findings in ***clinical practice*** settings. Hence uncovering natural clusters in the data while simultaneously effectively modelling and visualizing relationships among psychological constructs represent an appealing strategy to (i) identify more vulnerable users; (ii) aid in developing tailored training programs enhancing digital wellbeing; (iii) uncover new data-driven relationships thus generating new hypotheses
Method and Results: An approach integrating model-based clustering and graphical models[1,2] has been applied to copula transformed data collected on a sample of 229 middle school (pre)adolescents which took part to the online survey investigating selfie-sharing/editing behaviour, the relationship with digital self-image, problematic use of social network and possible internalizing symptoms, including both validated and ad-hoc realized scale. A two-clusters solution was selected as best based on BIC criterion: the two clusters actually showed different covariance network and different management of online self-image and psychological status. Participants in the cluster displaying a worse management of online self-image and psychological status were mainly female reporting higher use of social networks. To better examine the relationships among variables within each cluster, partial correlation networks were estimated separately for the two clusters and compared using both global and local network statistics and inferential procedure for network comparison.
Conclusions: Combining potentials of model-based clustering with graphical models represents a promising approach to investigate selfie behavior identifying target variables on which leverage to enhance digital wellbeing.
Keywords: Model-based clustering, graphical models, selfie behaviour

[1] *Fop, M., Murphy, T.B. and Scrucca, L. Model-based clustering with sparse covariance matrices. Statistics and Computing, 29(4), 2019, 791-819.*
[2] Kashihara, J., Takebayashi, Y., Kunisato, Y. and Ito, M. Classifying patients with depressive and anxiety disorders according to symptom network structures: A Gaussian graphical mixture model-based clustering. Plos one, 16(9), 2021, p.e0256902

TO11.2 The average uneven mortality index: building on the "e-dagger" measure of lifespan inequality

Bonetti M.*[1], Basellini U.[2], Nigri A.[3]
[1]*Carlo F. Dondena Research Center, Bocconi University ~ Milan ~ Italy,* [2]*Max Planck Institute for Demographic Research ~ Rostock ~ Germany,* [3]*University of Foggia ~ Foggia ~ Italy*

In recent years, lifespan inequality has become an important indicator of population health. We revisit the "e-dagger" measure of lifespan inequality, introduced in [1]. Specifically, we introduce a novel mortality indicator, which enlarges the toolbox of available methods for the study of mortality dynamics. We note that when conditioning on surviving at least until age a, the conditional e-dagger index can be written as the covariance between the conditional lifespan random variable and its transformation through its own cumulative hazard function (hence generalizing an earlier result). Leveraging this result, we obtain an upper bound for the index, and exploit it to introduce the "Average Uneven Mortality" (AUM) index, a novel relative index that can be used to analyze mortality patterns. We discuss some general features of the new index, including its relationship with a constant ("even") force of mortality. We develop exact formulas for the calculation of the two indices from life table data, under the assumption of piece-wise constant hazard function within each age class. Preliminary results suggest that the new functions have higher precision when compared to conventional and available functions, particularly for calculations involving older ages. The use of the AUM index is illustrated through an application to observed period and cohort death rates, as well as to period life-table death rates from the Human Mortality Database. The new AUM index can be related to the more general study of the shape of the age-at-death distribution, which has gained increasing attention in most recent decades (see, e.g., [2]). We believe that this novel indicator may provide additional insights on human mortality, enlarging the toolbox of available methods for the analysis of mortality developments.

[1] Vaupel, J. W. and Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. Demography, 40(2):201–216.
[2] Bonetti, M., Gigliarano, C., and Basellini, U. (2021). The Gini Concentration Index for the Study of Survival. In The Gini Inequality Index, pages 107–124. Chapman and Hall/CRC.

TO11.3 Simultaneous directional inference

Heller R.[2], Solari A.*[1]
[1]*University of Milano-Bicocca ~ Milano ~ Italy,* [2]*Tel-Aviv University ~ Tel-Aviv ~ Israel*

We consider the problem of inference on the signs of n > 1 parameters. Within a simultaneous inference framework, we aim to: identify as many of the signs of the individual parameters as possible; provide confidence bounds on the number of positive (or negative) parameters on subsets of interest. Our suggestion is as follows: start by using the data to select the direction of the hypothesis test for each parameter; then, adjust the one-sided p-values for the selection, and use them for simultaneous inference on the selected n one-sided hypotheses. The adjustment is straightforward assuming that the one-sided p-values are conditionally valid and mutually independent. Such assumptions are commonly satisfied in a meta-analysis, and we can apply our approach following a test of the global null hypothesis that all parameters are zero, or of the hypothesis of no qualitative interaction. We consider the use of two multiple testing principles: closed testing and partitioning. The novel procedure based on partitioning is more powerful, but slightly less informative: it only infers on positive and non-positive signs. The procedure takes at most a polynomial time, and we show its usefulness on a subgroup analysis of a medical intervention, and on a meta-analysis of an educational intervention.

Heller, R. and Solari, A. (2022) Simultaneous Directional Inference. arXiv preprint arXiv:2301.01653.

TO11.4 Treatment effect assessment in observational studies: a propensity Score method based on bayesian networks

Vicard P.[1], Rancoita P.M.V.[2], Cugnata F.[2], Briganti A.[3], Mecatti F.[4], Di Serio C.*[2], Conti P.L.[5]
[1]Roma Tre University ~ Rome ~ Italy, [2]University Centre for Statistics in the Biomedical Sciences, Vita-Salute San Raffaele University ~ Milan ~ Italy, [3]Vita-Salute San Raffaele University ~ Milan ~ Italy, [4]University of Milano- Bicocca ~ Milan ~ Italy, [5]Sapienza University of Rome ~ Rome ~ Italy

In medical research evaluating treatments effect on a given outcome using data from observational studies has become increasingly important. However, in these data there could be relevant differences between treatment groups and these differences can lead to biased estimates of treatment effects. Propensity score (PS) is a widely used tool to account for reducing this confounding. Usually, the PS is estimated through logistic regression within different methodological frames for treatment evaluation. In spite of its positive aspects, the logistic model relies on pre-specified assumptions on the dependency among the variables. If any of these are violated, the misspecified PS may fail to achieve covariate balance between treatment groups, which could subsequently bias treatment effect estimates. To overcome this issue, we propose to estimate the PS by using Bayesian Networks (BNs). The main advantages of this approach are flexibility and the possibility to learn the dependence structure that involves jointly treatments and covariates from data. The proposed estimator of the PS is then used to estimate the Average Treatment Effect (ATE) through two inverse probability weighting estimators: the Horvitz-Thompson (HT) type and the Hajek (H) type estimators. On the theoretical side, the two estimators are shown to be asymptotically equivalent when the model for PS is correctly specified. In case of misspecification, the H-type estimator proves to be better than the HT-type estimator. The properties of the proposed approach are also studied through two simulation studies. The first one is inspired by real data setting, by generating data mimicking real data of prostate cancer patients, while the second one uses a more general data generating mechanism. The simulation results show that the proposed approach outperformed the standard logistic modelling of PS especially in presence of a complex dependence structure of treatments from covariates. Estimating propensity score via BNs offers substantial theoretical and practical advantages. The more complex the dependence of treatments from covariates is, the greater is the gain due to BNs. Since the dependence structure of treatments from covariates is generally unknown, it is particularly important to account for complexity of such a structure in the model for propensity score.

1 R G. Cowell, A P. Dawid, S L. Lauritzen, and D J. Spiegelhalter. Probabilistic Networks and Expert Systems. Springer Verlag, New York, 1999.
2 K. Hirano, G W. Imbens, and G. Ridder. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica, 71:1161–1189, 2003.

WIN1 INVITED SESSION

RECURRENT EVENTS AND THEIR USE IN MEDICAL STUDIES

ORGANIZER | CHAIR: THOMAS SCHEIKE

WIN1.1 Estimands for recurrent event endpoints

Akacha M.*
Novartis Pharma AG ~ Basel ~ Switzerland

Treatment effects on recurrent events are of primary clinical interest in many diseases such as asthma or chronic heart failure. When death makes it impossible to experience further events, defining an appropriate measure of the treatment effect, the estimand, is challenging. For example in chronic heart failure, patients may experience repeated hospitalizations, but are also at an increased risk of death. For a test treatment which reduces mortality compared to a control treatment, one may observe more hospitalizations under the test treatment simply because patients with a high risk of hospitalizations may die earlier under the control treatment.
Many statistical analysis procedures have been proposed for such data, however it is often unclear which estimands these imply. As emphasized in the causal inference literature and the ICH E9(R1) guideline, the definition of the estimand of interest should precede any statistical analysis. We focus here on estimands for recurrent events terminated by death that have a causal interpretation. We also discuss whether common statistical analyses imply causal estimands of interest. Chronic heart failure is used as a motivating example.
A suite of recurrent event methods exist but it is not always clear which estimands these approaches target. We advocate for taking a step back and to first precisely specify estimands which are clinically meaningful. More cross functional discussion and alignment is needed to reach a conclusion. Heinz Schmidli, James H. Roger & Mouna Akacha (2021) Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2021.1895883
H. M. James Hung, Sue-Jane Wang. (2022) Comment on "Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event". Statistics in Biopharmaceutical Research 0:0, pages 1-2. Jiawei Wei, Tobias Mütze, Antje Jahn-Eimermacher, James Roger. (2021) Properties of Two While- Alive Estimands for Recurrent Events and Their Potential Estimators. Statistics in Biopharmaceutical Research 0:0, pages 1-11. Arno Fritsch, Patrick Schlömer, Franco Mendolia, Tobias Mütze, Antje Jahn-Eimermacher, on behalf of the Recurrent Event Qualification Opinion Consortium. (2021) Efficiency Comparison of Analysis Methods for Recurrent Event and Time-to-First Event Endpoints in the Presence of Terminal Events— Application to Clinical Trials in Chronic Heart Failure. Statistics in Biopharmaceutical Research 0:0, pages 1-12.
https://www.ema.europa.eu/en/documents/other/draft-qualification-opinion-clinically-interpretable-treatment-effect-measures-based-recurrent-event_en.pdf

Invited Sessions | Wednesday 30 August 2023

WIN1.2 Estimating the marginal and conditional means of recurrent events in presence of terminal events

Cortese G.*[1], Scheike T.[2]
[1]University of Padova ~ Padova ~ Italy, [2]University of Copenhagen ~ Copenhagen ~ Denmark

In many clinical and epidemiological studies, to assess disease progression over time, it is of great interest to study recurrent events, i.e., the repeated occurrence of the same event over time, when a terminal event may also be experienced. This problem is particularly relevant in practice when the rate of the terminal event is high. A very relevant measure for studying disease progression is the marginal mean of the number of recurrent events, experienced prior to the terminal event. Its functional form can be studied over time and compared for different covariate levels. Another interesting key measure is the conditional mean number of recurrent events for the survivors of the terminal event, and for those who have died before a specific time. First, we present a class of efficient IPCW estimators based on a dynamic prediction augmentation (dynamic AIPCW estimators), derived using semi-parametric efficient estimation theory for right- censored data. We discuss the amount of efficiency gain provided by these estimators and show that standard estimators are efficient in settings with no heterogeneity, but, in other settings with different sources of heterogeneity, the efficiency can be greatly improved when dynamic AIPCW are employed, at no extra cost to robustness. Moreover, we show that various regression models can be used to assess the impact of covariates in reducing or increasing the mean number of recurrent events. When a terminal event is present, the same covariates may affect simultaneously the recurrent and terminal events, and play a different role for survivors and for those who had a terminal event during the observed time period. In regard to this, we present dynamic AIPCW estimating equations that provide an efficient estimation of regression coefficients, associated with smaller estimated variance. As a worked example, we apply the proposed approach to study the mean number of catheter-related bloodstream infections in heterogeneous patients with chronic intestinal failure who can possibly die. Here, we highlight the efficiency gain that results in narrower pointwise confidence intervals.

Cortese G., Scheike T. (2022). Efficient estimation of the marginal mean of recurrent events. Journal of the Royal Statistical Society-series C, 71:1787-1821. Cortese G., Scheike T. (2023). Regression models for recurrent events in presence of terminal events with efficient estimation.

WIN1.3 Dealing with competing risks in the analysis of recurrent events

Andersen P.K.*
Biostatistics, University of Copenhagen ~ Copenhagen ~ Denmark

Recurrent events outcomes are frequent in both clinical and epidemiological studies and their analysis is often complicated by the presence of competing risks in the form of terminating events. A challenge is that a frequent terminating event will reduce the occurrence of the recurrent event and, thereby incorrectly, make a group appear more beneficial.
Several approaches to this challenge have been proposed, including:

1 Ignore it by treating terminating events as censoring
2 Ignore it by focusing on the recurrent events, only
3 Study a composite end-point consisting of both recurrent and terminating events and, thereby, reduce to a one-dimensional problem
4 Study a 'while alive' estimand and, thereby, reduce to a one-dimensional problem
5 Acknowledge that the outcome is, indeed, bivariate

A review will be given of these different approaches, recommending to treat the problem as a bivariate one. A method based on pseudo-values is applicable for this purpose. Furberg, J.K., Andersen, P.K., Korn, S., Overgaard, M., Ravn, H. Bivariate pseudo-observations for recurrent event analysis with terminal events. Lifetime Data Analysis (in press).

WIN2 INVITED SESSION

QUANTIFICATION OF SAFETY SIGNALS IN CLINICAL TRIALS: ESTIMAND, ESTIMATION, AND HOW WOULD GOOD LOOK LIKE IN TEN YEARS?

ORGANIZER | CHAIR : KASPAR RUFIBACH

WIN2.1 Principled approach to time-to-event endpoints with competing risks, with a focus on analysis of aes

Rufibach K.O.B.O.S.W.G.*
Product Development Data Sciences, F. Hoffmann-La Roche Ltd, Basel, Switzerland ~ Basel ~ Switzerland

The assessment of safety is an important aspect of the evaluation of new therapies in clinical trials, with estimation of adverse event risk being an essential part of this. Standard estimators for such a probability of an adverse event (AE), such as the incidence proportion defined as the number of patients with a specific AE out of all patients in the treatment group of interest or one minus the Kaplan-Meier estimator of time-to-adverse event, do not account for varying follow-up times between arms and/or competing risks. Based on a sample of 17 randomized controlled trials (RCT) the SAVVY project (Survival analysis forAdVerse events with VarYing follow-up times, an academia – pharma consortium) aimed at quantifying the bias of several estimators of the AE probability. We will discuss common estimators of the AE probability and the assumptions they are making. Based on a sample of 186 AEs from the 17 RCTs we will empirically illustrate how large biases of various estimators can become for estimation of AE probabilities in one arm. In addition, the bias of estimators of relative AE risk between two arms will be quantified. We propose to use the Aalen-Johansen estimator to estimate AE risks, as it is a common nonparametric estimator of an event probability that properly accounts for varying follow-up times and competing risk that is implement in any standard software package. We will also discuss how key guidelines would need to be updated in light of our findings. Standard estimators of AE probabilities are biased up to a factor of five in the presence of varying follow-up times between patients or arms and/or competing risks. We advocate switching to the Aalen- Johansen estimator and to reflect this change in regulatory and reporting guidelines. More on SAVVY, incl. links to papers and markdown with code: https://numbersman77.github.io/savvy/

Invited Sessions | Wednesday 30 August 2023

WIN2.2 Estimands for safety – one size fits all?

Loos A.*
Merck ~ Darmstadt ~ Germany

With ICH E9(R1), the estimand framework has made its way into practice for efficacy endpoints in clinical trials [1]. Discussions continue on the application of the estimand concept for exploratory analyses including safety.
Traditionally, most of clinical studies are designed in support of efficacy evaluations. In contrast to efficacy, safety objectives are in most cases rather unspecific and of exploratory nature. Unless specific safety topics are already identified, based on e.g. (pre-)clinical observations, class effects, or mode of action, safety objectives focus primarily on safety surveillance or signal detection. With accumulation of safety data for a medicinal product, more specific safety questions can be raised based on the evolving knowledge about risks potentially associated with treatment. While standard safety analyses primarily focus on the detection of potential risks, regulatory authorities expect at marketing authorization the latest, an identification, characterization and quantification of adverse reactions common enough to be detected in an appropriately large clinical trial safety database [2-5]. Criticism on standard safety analyses is not new. The estimand framework has contributed to a re- ignited discussion on the current practice of characterizing the safety profile during clinical development. Even if the status of a safety profile has not yet evolved to enable specific questions, the estimand concept can be helpful to inform discussions on study design and data retrieval. In this talk, I will critically evaluate the applicability of the estimand framework in the light of the different safety objectives and report on current status of discussions in the scientific community.

[1] ICH E9(R1) – Addendum on estimands and sensitivity analysis in clinical trials to the guideline onstatistical principles for clinical trials. November 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf, last accessed Feb 2023.
[2] ICH E1 – The extent of population exposure to assess clinical safety for drugs intended for longtermtreatment of non-life-threatening conditions. https://database.ich.org/sites/default/files/E1_Guideline.pdf, last accessed Feb 2023.
[3] FDA. Guidance for Industry – Premarketing risk assessment. March 2005.https://www.fda.gov/media/71650/download, last accessed Feb 2023.
[4] European Commission. Guideline on summary of product characteristics (SmPC). September2009. https://health.ec.europa.eu/system/files/2016-11/smpc_guideline_rev2_en_0.pdf, last accessedFeb 2023.[5] FDA. Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products— Content and Format. January 2006. https://www.fda.gov/media/72139/download, last accessed Feb2023.WIN2.2_ 48

WIN2.3 Adverse events with survival outcomes: from clinical questions to methods for statistical analysis

Tassistro E., Valsecchi M.G., Bernasconi D.P., Antolini L.*
Università Milano Bicocca ~ Milano ~ Italy

When studying a novel treatment with a survival time outcome, failure can be defined to include an adverse event (AE) among the endpoints typically considered, for instance relapse. These events act as competing risks, where the occurrence of relapse as first event and the subsequent treatment change exclude the possibility of observing AE related to the treatment itself.
In principle, the analysis of AE could be tackled by two different approaches:
1. It requires a competing risk framework for analysis: the clinical question relates to the observed occurrence of AE as first event, in the presence of the event "relapse";
2. It requires a counterfactual framework for analysis: the clinical question relates to the treatment causing AE occurrence as if relapse could not occur.
This work has two aims: the first is to critically review the standard theoretical quantities and estimators with reference to their appropriateness for dealing with approaches 1 or 2 and to the following features: (a) estimators should address for the presence of right censoring; (b) theoretical quantities and estimators should be functions of time. The second aim is to define a strategy to relax the assumption of independence between the potential times to the competing events of the commonly used estimators when counterfactual approach 2 is of interest. After reviewing the standard methods[1] we clarify the impact of the crucial assumption of independence between potential times to competing events of the standard estimators used in the counterfactual approach. We propose the use of regression models, stratified Kaplan-Meier curves and inverse probability of censoring weighting[2] to relax the assumption of independence by achieving conditional independence given covariates and we develop a simulation protocol to show the performance of the proposed methods. The proposed methods overcome the problem due to the dependence between the two potential times. In particular, one can handle patients' selection in the risk sets, and thus obtain conditional independence between the two potential times, adjusting for all the observed covariates that induce dependence. The proposed methods can be also extended to the case of repeated adverse events.

[1] A. Allignol, J. Beyersmann, C. Schmoor (2016). Statistical issues in the analysis of adverse eventsin time-to-event data. Pharmaceutical Statistics, 15, 297-305
[2] S.J.W. Willems, A. Schat, M.S. van Noorden, M. Fiocco (2018). Correcting for dependent censoringin routine outcome monitoring data by applying the inverse probability censoring weighted estimator.Biometrical Journal, 62, 836-851

WIN2.4 Regulatory perspective on the analysis of safety in clinical trials and beyond

Roes K.*
Radboudumc ~ Nijmegen ~ Netherlands

From a regulatory perspective, the assessment of safety of new drugs is crucially important, as decision making is based on establishing that the benefit-risk balance is positive. This safety assessment goes well beyond estimation and quantification of signals from adverse events and laboratory data in clinical trials, and typically includes data, evidence and biological and pharmacological insights from a range of resources. By the very nature analysis and assessment includes signals on effects that were not foreseen, and hence cannot be a priori defined with the same rigor as estimands for primary efficacy outcomes. Nevertheless, estimation and quantification of the associated uncertainty (potential bias and variance) are important, and it is a long existing challenge that there has not been addressed statistically very well, while there is in principle no lack of methodology (e.g., in survival analysis). The estimand framework (1) projected on the analysis of safety data may help improve this, but it could be argued it does so because it highlights that the urgent improvement needed is proper estimation of comparative effects (irrespective of which estimand is targeted). In this presentation, I will explore current practice and pitfalls, what could be done to improve (jointly across the different stakeholders) and look ahead on how safety profiles are to be communicated (e.g. In SmPCs). I will also address potential use of external (real world data) in light of DARWIN-EU (2) and bridge to good practice guidance from pharmacoepidemiology (3).

(1) ICH E9(R1) – Addendum on estimands and sensitivity analysis in clinical trials to the guideline onstatistical principles for clinical trials. November 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf, last accessed Feb 2023.
(2) https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-worldinterrogation-network-darwin-eu
*(3) https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtmlWIN2.4_*

Invited Sessions | Wednesday 30 August 2023

WIN3 INVITED SESSION

MARGINAL VERSUS CONDITIONAL EFFECTS IN CLINICAL TRIALS

ORGANIZER | CHAIR: JONATHAN BARTLETT

WIN3.1 A value system for evaluating estimands in randomized trials

Benkeser D.*
Emory University Rollins School of Public Health ~ Atlanta ~ United States of America

There is a long-standing debate in the statistical community as to which estimands and estimators should be preferred in the context of randomized trials. This debate has led to confusion and acrimony amongst statisticians, patients, and providers. This talk aims to provide structure to this debate in the form of a value system for evaluating the relevance of estimands for various stakeholders in the drug evaluation process. We outline several axes of values along which estimands may be evaluated. We find that the role of that each stakeholder plays as well as individually held beliefs may influence which values are deemed of greatest importance in the regulatory process. We conclude that there are no "right" or "wrong" estimands in the context of a randomized trial. Any uniform recommendation as to which estimands should be preferred is likely to disparage one set of values in favor of another. Estimands should be selected to explicitly reflect the values held by key stakeholders, while transparently recognizing limitations that may be present in the view of other stakeholders.

WIN3.2 Conditional vs. marginal effects in randomized trials: tradeoffs

Rosenblum M.*[1], Wang B.[2]
[1]*Johns Hopkins University ~ Baltimore ~ United States of America,* [2]*University of Michigan ~ Ann Arbor ~ United States of America*

The target of inference (estimand) in a randomized trial could be chosen to be a marginal (average) or conditional treatment effect. A 2021 FDA draft guidance on covariate adjustment for trials of drugs and biologics discusses both types of treatment effect. We present tradeoffs between these two estimands in terms of the following: interpretation, assumptions required, robustness to model misspecification, power/precision, and transportability.
This is based in part on a comment submitted to the FDA by Michael Rosenblum and Bingkai Wang: https://downloads.regulations.gov/FDA-2019-D-0934-0040/attachment_1.pdf (see below)
An important advantage of the marginal treatment effect is that it is well defined and interpretable without having to make model assumptions (such as a linear or logistic regression model being correctly specified). In contrast, the conditional treatment effect typically requires such model assumptions or assumptions such as the treatment effect being constant across all strata of the relevant baseline variables. When the latter type of assumption is false, the conditional treatment effect is not a single number but instead is a function that may differ across strata of baseline variables; it therefore may be challenging to estimate with high precision. On the other hand, conditional effects provide more fine-grained information than marginal effects and could be more useful for clinical decisions.
Comment on FDA Draft Guidance for Industry (FDA-2019-D-0934): Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics, Michael Rosenblum1 and Bingkai Wang https://downloads.regulations.gov/FDA-2019-D-0934-0040/attachment_1.pdf

WIN3.3 Why do we worry about marginal inferences?

Senn S.*
University of Sheffield ~ Sheffield ~ United Kingdom

"A statistician is one who with one foot in ice and the other in boiling water says that on average they are comfortable, " goes a common criticism of statistics. It seems that the more we get beyond simple averages, the more relevant to practical action our inferences are. A common criticism of marginal estimates, made by many 1,2,3 is that compared to conditional estimates they seem to have little relevance for decision making. I shall consider what defence, if any, can be made of marginal estimates.
There are at least three justifications. The first is that the marginal estimates are applied to an agreed common ground on which different approaches can be compared. The second is that marginal estimates are in a sense calibrating: a correct conditional approach ought to be able to produce a valid marginal prediction. The third is that there may be circumstances (although such circumstances as rare) under which a decision has to be made at a level of a population. A relevant side-issue is that the linear model seems to be an odd one out when it comes to conflict between expectations for marginal and conditional models, although this conflict may be less important when predictions are considered. It may be that we ought to use single parameter models rather less than we do. A further issue is that patients should not be considered as being a random sample of some target population. In conclusion, conditional estimates are nearly always more relevant for the purpose of making decisions, since decisions in which we can decide who in a group of patients should get treatment are superior to those in which we merely decide whether everybody in a group should get treatment. Nevertheless, for calibrating our inferences marginal estimates may have some value.

1. Lindsey JK, Lambert P. On the appropriateness of marginal models for repeated measurements in clinical trials. Statistics in Medicine. 1998;17(4):447-69.
2. Lee Y, Nelder JA. Conditional and marginal models: Another view. Statistical Science. 2004;19(2):219-28.
3. Senn S. Conditions for success and margins of error: Estimation in clinical trials. Stat Med. 2022;41(28):5586-8. doi:10.1002/sim.9497

WIN3.4 Covariate adjustment and exploiting ordinality: simulations of power and a review of neurological trials

Steyerberg E.*
Leiden University Medical Center ~ Leiden ~ Netherlands

Randomized controlled trials for rare neurological diseases have a disappointing lack of success, possibly due to inefficient statistical analyses. We aimed to evaluate the impact of covariate adjustment for baseline characteristics and ordinal analysis on statistical power in randomized controlled trials with ordinal outcome measures. We also assessed current practice with respect to covariate adjustment and ordinal analysis in neurological trials. We reanalysed a previous trial in Guillain-Barré Syndrome (GBS) [1]. A small gain in power was achieved by covariate adjustment for two well-known prognostic factors, and a larger gain by exploiting ordinality instead of dichotomizing the ordinal scale [2]. The gains translated to a potential reduction in sample size of 10 and 26 % respectively. More systematic simulations confirmed these increases in power. The gains in power were only slightly smaller under plausible violations of the proportional odds assumption. Next, a scoping review included 50 randomized controlled trials published between 2015 and 2021 for treatment of acute neurological diseases with an ordinal scale as primary or secondary end point. This review showed ordinal scale analyses, commonly with a proportional odds model in 30 trials. In 20 trials, the ordinal outcome was dichotomized as poor versus good (e.g., unfavourable versus favourable), with disagreement about the cut off value within fields. Optimal analysis of ordinal scales should adjust for baseline characteristics (covariate adjustment) and should respect the ordinality of the outcome measure. While already common in acute neurological diseases, further implementation of this approach as the primary statistical analysis requires attention.

[1] van Leeuwen, N., Walgaard, C., van Doorn, P. A., Jacobs, B. C., Steyerberg, E. W., & Lingsma, H. F. (2019). Efficient design and analysis of randomized controlled trials in rare neurological diseases: an example in Guillain-Barre syndrome. PloS one, 14(2), e021140.
[2] Senn, S., & Julious, S. (2009). Measurement in clinical trials: a neglected issue for statisticians? Statistics in medicine, 28(26), 3189-3209.

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO1: CLINICAL TRIALS 7

WO1.1 Methods for assessment of frequentist operating characteristics in Bayesian trials

Golchi S.*
McGill University ~ Montreal ~ Canada

Bayesian posterior and posterior predictive probability statements are commonly used as test statistics in Bayesian clinical trials. Despite the popular use of Bayesian inference in clinical trials, however, the design needs to be assessed with respect to frequentist operating characteristics such as power and type I error rate that are defined according to the sampling distribution of the test statistic. Evaluation of frequentist error rates either for the purpose of strictly controlling them or to evaluate integrated risk functions in a decision theoretic framework requires estimation of sampling distribution of the Bayesian test statistic. However, unlike in the basic and sometimes oversimplified hypothesis tests employed for conventional randomized clinical trials, the sampling distribution of the test statistic is not available analytically. Computationally intensive simulation studies are routinely performed to estimate the sampling distribution and assess the frequentist operating characteristics in Bayesian clinical trial. The goal of the present talk is to address the computational complexity by modelling the sampling distribution of the "Bayesian test statistic" that are derived from the marginal posterior distribution of the "effect". The proposed approach takes advantage of the known behaviour of the test statistic as the effect size (the parameter of interest) and sample size grow to interpolate the sampling distribution based on significantly smaller number of simulation scenarios. The proposed methods are illustrated by exploring the operating characteristics of Bayesian adaptive trials with covariate adjustment. The proposed approach significantly reduces the computational burden in design of Bayesian clinical trials and thereby enables exploration of a larger variety of designs with complex analysis models.

- Gelfand, A. E. and Wang, F. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. Statistical Science, 17(2):193-208.
- Golchi, S. (2022). Estimating design operating characteristics in Bayesian adaptive clinical trials. Canadian Journal of Statistics, 50(2):417-436. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.11699.
- M uller, P. and Parmigiani, G. (1995). Optimal Design via Curve Fitting of Monte Carlo Experiments. Journal of the American Statistical Association, 90(432):1322-1330.

WO1.2 Optimal adaptive designs for time-to-event data: a simulation study

Bruder N.*, Meis J., Kieser M.
Institute of Medical Biometry, University of Heidelberg ~ Heidelberg ~ Germany

Time-to-event endpoints are frequently applied in clinical trials measuring the patients' survival time or progression-free survival. As they are often used in the context of serious diseases, rapid evaluation of an efficient therapy has the potential of saving many patient years. Therefore, an adaptive two-stage design with the option for early stopping could be a promising way of designing such a trial. Recently, a lot of research has been done on finding optimal adaptive designs. In the approach by Pilz et al. [1], the optimal design is determined via a numerical optimization procedure such that the best performance according to a chosen criterion is achieved. The optimization problem can be solved complying with various constraints, for example with respect to type I error rate or power. Since the performance of an optimized design cannot be improved for a given set of constraints, this approach is an attractive way of choosing design parameters. These optimal designs are most efficient when the determined information rates are followed exactly as specified. Adhering to this principle would require constant monitoring of trial participants and performing the interim analysis right after the specified number of first-stage events is reached, which is often impractical. Operationally, it is usually much easier to plan the interim analysis for a fixed point in time. However, when initial estimates of the event rates from the planning phase turn out to be wrong, the observed information rate at the interim analysis can be markedly different from the optimal information rate of the design. In this work, we will compare optimal adaptive designs with various other methods of designing a two-stage clinical trial for time-to-event endpoints. We will evaluate their performance in terms of type I error, power, and average sample size under a variety of different assumptions on event-rate, hazard ratio, and variations in baseline-hazard. By comparing different methods for designing a two-stage trial with time-to-event endpoints, we aim to identify design candidates which are efficient while still being reasonably robust to misspecification of the event rate.

[1] Pilz, M, Kunzmann, K, Herrmann, C, Rauch, G, Kieser, M., Statistics in Medicine. 2021; 40: 3196– 3213

WO1.3 Confirmatory adaptive enrichment designs with a normally distributed outcome

Stallard N.*
Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick ~ Coventry ~ United Kingdom

With the growing importance of clinical trials of targeted medicine there has been much recent interest in two-stage adaptive enrichment designs [1,2,3]. In this design patients from the first stage are used to identify a biomarker-defined population in which a treatment effect is anticipated. In the second stage the trial population is 'enriched' by restricting recruitment to patients from this selected population. At the end of the trial a hypothesis test is conducted of the treatment effect in the selected population. The data-dependent selection leads to statistical challenges if data from both stages are for this hypothesis test. This talk considers the case in which the outcome is normally distributed. Subgroup selection using the stage 1 data can then be based on a linear model, with the subgroup selected to comprise patients with biomarker values that correspond to a positive predicted treatment effect. By considering the joint distribution of the test statistic comparing the treatment groups and the size of the selected subgroup, the distribution of a test statistic based on data from patients in the selected subgroup in both stages is derived. This enables calculation of a p-value corrected for the selection, allowing use of data from both stages. The proposed method provides a valid analysis method enabling use of data from both stages of a two-stage adaptive enrichment design in a confirmatory setting. The method can be used for either continuous or dichotomous biomarkers.

[1] N. Simon, R. Simon. Biostatistics, 14, 2013, 613-625.
[2] N. Stallard. Biometrics, 79, 2023, 9-19.
[3] R. Frieri, W.F. Rosenberger, N. Flournoy, Z. Lin. Biometrics, early view. DOI: 10.1111/biom.13805

WO1.4 Adaptive enrichment clinical trial designs using joint modelling of longitudinal and time-to-event data

Burdon A.*, Jaki T.
University of Cambridge ~ Cambridge ~ United Kingdom

Adaptive enrichment allows for pre-defined subgroups to be investigated throughout the course of a clinical trial. Many trials which measure a long-term time-to-event endpoint often also routinely collect repeated measures on biomarkers which may be predictive of the primary endpoint. We aim to make greater use of these data to increase efficiency and improve interim decision making. We present a joint model for longitudinal and time-to-event data and methods for creating standardised test statistics based on this joint model. We can use the estimates to define subgroup selection rules and efficacy and futility early stopping rules for a flexible efficient clinical trial with possible enrichment. We shall apply the methodology to a trial for the treatment of metastatic breast cancer where repeated ctDNA measurements are available and the subgroup criteria is defined by patients' HER2 status. A key feature of this methodology is the necessity to accurately predict the correlation between the two endpoints and we discuss how the timing of the interim analysis plays a role. Using simulation, we show the benefits of incorporating biomarker information and the effects on interim decision making.

Rizopoulos, D., 2012. Joint models for longitudinal and time-to-event data: With applications in R. CRC press.
Stallard, N., 2010. A confirmatory seamless phase II/III clinical trial design incorporating short-term endpoint information. Statistics in medicine, 29(9), pp.959-971.

Parallel Sessions | Wednesday 30 August 2023

WO1.5 A two-stage bayesian adaptive umbrella design borrowing information over the control data

Ouma L.*[1], Wason J.[1], Grayling M.[1], Zheng H.[2]
[1]Newcastle University ~ Newcastle Upon Tyne ~ United Kingdom, [2]MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom

Umbrella trials are a type of precision medicine trial design comprising a series of subtrials run in parallel within a single disease setting, each evaluating a unique targeted therapy. To date, most umbrella trials are designed and analysed independently – an approach that only provides operational efficiency gains under the master protocol framework.
Motivated by the efficiencies guaranteed by novel Bayesian methodology in related designs such as basket trials, we propose a new two-stage Bayesian adaptive umbrella design, coupled with a joint analysis of the substudies, wherein a common control arm is embedded in each subtrial for comparison with experimental treatment arms. Specifically, the proposed design features adaptive assignment in favour of an experimental treatment if adequate evidence suggests it outperforms the control at an interim analysis. For the efficient use of a common control across substudies, we employ commensurate priors to facilitate borrowing of information. Bayesian predictive power (BPP) is utilised to decide how far a deviation from equal allocation should be considered after the interim analysis. The choice of the post-interim allocation ratio is set to either maximise the BPP, number of patients on treatment, match the expected power given by equal allocation or guarantee at least a pre-specified BPP. Numerical results demonstrate that the proposed Bayesian adaptive design can increase allocation to experimental arms while maintaining statistical power and error rate control at desirable levels.
Novel Bayesian methodology presents a useful framework to introducing considerable statistical efficiencies to the umbrella design.

Ouma, L.O., Grayling, M.J., Wason, J.M. and Zheng, H., 2022. Bayesian modelling strategies for borrowing of information in randomised basket trials. Journal of the Royal Statistical Society. Series C: Applied Statistics.
Zheng, H. and Wason, J.M., 2022. Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. Biostatistics, 23(1), pp.120-135.

PARALLEL SESSION WO2: SURVIVAL ANALYSIS 6

WO2.1 The shape of the relative frailty variance induced by discrete random effects in time-to-event models

Bardo M.*[1], Unkel S.[2]
[1]University Medical Center Göttingen ~ Göttingen ~ Germany, [2]University of Siegen ~ Siegen ~ Germany

In statistical models for the analysis of time-to-event data, individual heterogeneity is usually accounted for by means of one or more random effects, also known as frailties. In the vast majority of the literature, the random effect is assumed to follow a continuous probability distribution. However, in some areas of application, a discrete frailty distribution may be more appropriate. We investigate and compare various existing families of discrete univariate and shared frailty models by taking as our focus the variance of the relative frailty distribution in survivors. The relative frailty variance (RFV) among survivors (Hougaard, 1984) provides a readily interpretable measure of how the heterogeneity of a population, as represented by a frailty model, evolves over time. We explore the shape of the RFV for the purpose of model selection and review available discrete random effect distributions in this context. We find non-monotone trajectories of the RFV for discrete univariate and shared frailty models, which is a rare property. Furthermore, we proof that for discrete time-invariant univariate and shared frailty models with (without) an atom at zero, the limit of the RFV approaches infinity (zero), if the support of the discrete distribution can be arranged in ascending order. Through the one-to-one relationship of the RFV with the cross-ratio function in shared frailty models, which we generalize to the higher-variate case, our results also apply to patterns of association within a cluster. Extensions and contrasts to discrete time-varying frailty models and contrasts to correlated discrete frailty models are discussed. Hougaard P. (1984) Life table methods for heterogeneous populations: Distributions describing neity. Biometrika 71(1):75–83. https://doi.org/10.2307/2336399

WO2.2 Family history in breast cancer development

Vinattieri M.V.*[1], Bonetti M.[1], Czene K.[2]
[1]Bocconi University ~ Milano ~ Italy, [2]Karolinska Institutet ~ Stockholm ~ Sweden

Risk prediction models for breast cancer develop- ment can help identify the highest-risk families for tailored, more intensive, screen- ing. We implement a multivariate model to better exploit the family breast cancer history information, with the intention to increase the accuracy of risk prediction. We assume that family members are characterized by a true common risk (multiplicative frailty) of developing breast cancer that is latent and unchanged from birth. We build a Multivariate Shared Frailty Cox (MSFC) model, with Gamma frailty, for the age at onset [1, 2]. We compare this model to the widely used univariate Cox model with a covariate that describes the pres- ence of any previous first-degree (mother and sisters) family history (FH) of breast cancer, as a binary replacement for the true frailty risk. Comparisons are made on model-generated data in terms of accuracy in risk prediction, focusing on the posterior frailty mean and median. In addition, we explore the use of the estimated family-specific posterior probability of belonging to the highest-risk families. We im- plement the model on a multi-generational Swedish dataset merged with the breast cancer registry of Sweden and present some preliminary results. We show that the MSFC model has a good fit and better performance in risk prediction compared to the FH model. We also discuss some features of cure rate models when applied to this setting. The appropriate use of family information is crucial in risk prediction models for breast cancer. It contributes to identifying and targeting the highest-risk families for tailored, more intensive, screening and prevention strate- gies. A rough summary of family information is not enough to capture the latent family risk component for disease development.

[1] Theodor A Balan and Hein Putter. A tutorial on frailty models. Statistical methods in medical research, 29(11):3424–3454, 2020.
[2] Germán Rodriguez. Multivariate survival models, 2010.

WO2.3 Flexible time-to-event models for double-interval-censored data with a competing event

Ramjith J.*, Andolina C., Bousema T., Jonker M.
Radboud University Medical Centre ~ Nijmegen ~ Netherlands

In routinely followed-up infectious disease data, the observed induction time from an incident infection to a secondary event-of-interest is often double-interval-censored and prevented from being observed by a competing risk. Usually, the times at which events are detected are used as a proxy for the exact times. Interpretation has to be made on these times and not the actual induction times. We aim to develop proportional hazards (PH) models with different baseline hazards specifications to estimate covariate effects and the baseline hazard. Through a simulation study we aim to evaluate the impacts of: mis-specifying the baseline hazard function on estimated covariate effects; or unnecessarily fitting too complex models; and the models' performances compared with the Cox proportional hazards (CPH) model when using exact times. We illustrate the methodology with a malaria example. We developed the likelihood function using exponential, Weibull and a B-splines baseline hazards in a PH framework using double-interval-censoring. The maximization of the likelihood, and estimation of penalty factors (B-splines hazard) are described. In the simulation study, we simulated the baseline hazard to be "flatter" for the event-of-interest and unimodal for the competing event across all scenarios. For the unimodal hazards, the exponential model underestimated the magnitude of the covariate effect while the Weibull model overstated the effect, and the B-splines model was unbiased. The exponential model's estimates were much closer to a null effect compared with the CPH model, while the magnitude of the Weibull model estimates were mostly larger. Estimates from the B-splines model correlated nearly perfectly with the CPH model. For the "flatter" hazards we saw that the estimated effects from all models were closely correlated with those from the CPH model. In the malaria example we found that the peak latent time from incident malaria to gametocyte initiation was around 2 weeks, with human sickle-cell trait as a significant risk factor. Understanding how the hazard evolves over the latent times, in the case of double-interval- censoring, is important. A flexible B-splines model is useful in estimating the shape of the baseline hazards as well as estimating unbiased effects without unreasonable assumptions of the baseline hazards.

Parallel Sessions | Wednesday 30 August 2023

WO2.4 Model assessment in regression with a doubly truncated response

De Uña--Álvarez J.*
CINBIO, Universidade de Vigo ~ Vigo ~ Spain

In Survival Analysis and Epidemiology, among other fields, doubly truncated data may appear. Double truncation means that the target variable is observed only when it falls within two random limits, which are also available in such a case. An important example of double truncation is interval sampling, where the lifetime data correspond to the individuals undergoing the event of interest between two given calendar dates. Under double truncation the target variable is subject to a sampling bias, and ordinary statistical methods may be inconsistent [1]. In this work an omnibus goodness-of-fit test for a regression model with a doubly truncated response is introduced. This is important when one is willing to assess the fit of a given structure, e.g. linearity, for the predictor. The test statistic is based on the distance between two empirical integrated regression functions: one purely nonparametric, and the other one driven by the model to be tested. The underlying process is a marked empirical process based on weighted residuals, where the weights remove the observational bias induced by the double truncation. The asymptotic null distribution of the test statistic is obtained for both a fully specified and a parametric regression model. A bootstrap algorithm is proposed in order to approximate the null distribution of the test in practice. This adapts the basic ideas in [2] to the doubly truncated framework. The method is illustrated with both simulated and real data. The proposed method is well calibrated, that is, the nominal level of the test is well respected. On the other hand, it is seen how the power of the test increases with the sample size and the degree of violation of the postulated regression model. Situations with heavy truncation are found to deteriorate the behaviour of the test, as expected. The application to regression for AIDS incubation times as depending on age at VIH infection shows how taking the double truncation into account is critical for a proper identification of the regression model.

[1] J. de Uña-Álvarez, C. Moreira, and R.M. Crujeiras, The Statistical Analysis of Doubly Truncated Data: With Applications in R, Wiley, Hoboken, 2021.
[2] W. Stute, W. Gonzalez-Manteiga, and M. Presedo-Quindimil (1998) Bootstrap approximations in model checks for regression. Journal of the American Statistical Association
93, 141-149.

WO2.5 Modelling excess mortality comparing to a control population: a combined additive and relative hazards model

Andersson T.M.*[1], Crowther M.J.[2], Weibull C.E.[3]
[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet ~ Stockholm ~ Sweden, [2]Red Door Analytics ~ Stockholm ~ Sweden, [3]Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet ~ Stockholm ~ Sweden

In medical research, exposures are often assumed to act multiplicatively on mortality. However, in some instances it is more biologically plausible to assume an additive effect. The best-known example is in population-based cancer patient survival, where the presence of cancer is assumed to have an additive effect on mortality. This excess mortality rate is typically measured using relative survival, where the observed mortality rate in the cancer population is compared to that in a similar cancer-free population (the expected mortality rate). In such analyses, a publicly available population mortality file stratified on sex, year, and age, is matched to the cancer population, and included in the relative survival model as an offset, and hence assumed to be measured exactly (i.e., without uncertainty). However, situations exist where this standard approach is not optimal or even possible. For example, it might be necessary to stratify the expected mortality on additional factors, such as socio-economy or comorbidity. Or, a suitable population mortality file might simply not exist. We propose a flexible parametric excess hazard model on the log hazard scale, incorporating a modelled expected rate from a control population (e.g., matched comparators). By modelling the expected rate, we appropriately allow for uncertainty. Covariate effects are assumed to be multiplicative within the expected and the excess hazard, while the presence of disease among the studied population (e.g., cancer patients) has an additive effect. The model is further extended to include time-dependent effects, multiple time-scales, and more. Following estimation, results are quantified through prediction of the survival, hazard, and cumulative incidence functions, as well as transformations of these, and crucially with associated confidence intervals on all measures. The proposed method is implemented in the Stata package stexcess (github.com/RedDoorAnalytics/stexcess). We illustrate the method using a population-based dataset of colorectal cancer patients diagnosed 2007-2016, with comparators matched 1:6 on: sex, age, country, and being colorectal cancer-free. Analyses are ongoing with no results available at this stage, but will be presented at the meeting. The proposed method, together with user-friendly software implementation, offers an alternative in situations when standard relative survival methods do not suffice.

PARALLEL SESSION WO3: LONGITUDINAL ANALYSIS 3

WO3.1 How resampling methods can improve variable selection in longitudinal Models

Rancoita P.M.*
Vita-Salute San Raffaele University ~ Milano ~ Italy

Longitudinal data are commonly modeled through mixed-effects (ME) models, which properly account for observations belonging to the same subject and for heterogeneity among subjects, by suitably specifying random effects terms. Longitudinal trends are usually analyzed accounting for baseline characteristics, by allowing the model parameters (related to fixed and/or random effect terms) to depend on them. Thus, often a variable selection procedure is needed to obtain a simpler parsimonious model which can be effectively used in the clinical practice. In this setting, different strategies based on parameter testing and/or goodness of fit measure evaluation could be applied. Methods based on penalizations were also defined. Moreover, the complexity of the model selection procedure increases when it involves also the random effect terms (instead of being fixed a priori). In the context of standard regression modelling, model selection approaches based on resampling methods, like subsampling and bootstrap, were evaluated, showing some advantages of subsampling over bootstrap due to its characteristics [1]. The performance of these methodologies in improving ME model selection has not been studied yet. A simulation study is set up to deeply evaluate when and how subsampling and bootstrap approaches allow to improve model selection for linear ME models. Several standard model selection strategies are considered (e.g. backward selection based on parameter testing) and compared with the corresponding approach based on resampling. The simulation scenarios are defined not only by setting different model parameters, but also by varying the sample size and the number of longitudinal observations per subject. Moreover, the resampling approaches are explored by varying the subsample size and the number of repetitions. Results show that, depending on the simulation scenario (especially, the combination of sample size and number of longitudinal observations, or the need of random effects selection), the subsampling strategy requires different subsample size and number of repetitions to outperform the corresponding standard selection method. Moreover, in more complex scenarios, subsampling outperformed bootstrap. A similar evaluation on public real clinical datasets is also performed. Resampling methods (especially subsampling) can improve model selection for linear ME models, if subsample size and number of repetitions are carefully set.

[1] R. De Bin, S. Janitza, W. Sauerbrei, A.L. Boulesteix, Subsampling versus bootstrapping in resampling-based model selection for multivariable regression, Biometrics, 72(1), 2016, 272-80

Parallel Sessions | Wednesday 30 August 2023

WO3.2 Dynamic prediction of an event using multiple longitudinal markers: a model averaging approach

Hashemi R.[1], Baghfalaki T.*[2], Philipps V.[2], Jacqmin--Gadda H.[2]
[1]Department of Statistics, Razi University, Kermanshah, Iran ~ Kermanshah ~ Iran, Islamic Republic of, [2]Centre Inserm Bordeaux Population Health U1219 ISPED Université de Bordeaux ~ Bordeaux ~ France

Dynamic event prediction, using joint modeling of survival time and longitudinal variables, is extremely useful in personalized medicine. However, the estimation of joint models including many longitudinal markers is still a computational challenge because of the high number of random effects and parameters to be estimated. We propose a model averaging (MA) strategy to combine predictions from several joint models for the event, including one longitudinal marker only or pairwise longitudinal markers. The prediction is computed as the weighted mean of the predictions from the one-marker or two-marker models, with the time-dependent weights estimated by minimizing the time-dependent Brier score. This method enables us to combine a large number of predictions issued from joint models to achieve a reliable and accurate individual prediction. Advantages and limits of the proposed methods are highlighted in a simulation study by comparison with the predictions from well-specified and misspecified all- marker joint models as well as the one-marker and two-marker joint models. The method is used to predict the risk of death in patients with PBC. MA has been proposed to increase predictive abilities and cope with model uncertainty, which is most often neglected after model selection. MA consists of estimating several possible models and combining their estimates or predictions using some weighting methods. In Bayesian MA, the quantities of interest are computed as the mean of the estimates from several candidate models weighted by the posterior probability of each model given the data.cOur approach avoids estimating a joint model including all markers that may be untractable as the number of markers increases. The time-dependent weights are computed to minimize the prediction error in each time window as measured by the dynamic Brier score1. Among the two MA approaches we proposed, MA of two-marker joint models often has better performances than its one-marker counterpart because it better accounts for the dependence between the markers but it is at the price of much more computation time when the number of markers is large. On the other side, the one-marker MA has better performances when the dependence structure between the event and the markers may be misspecified.

[1] Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. Biometrics 2015; 71(1): 102–113.

WO3.3 Non-parametric clustering of multivariate longitudinal data: identifying sub-phenotypes of alzheimer's disease

Rouanet A.*, Helmer C., Proust--Lima C.
U1219 Bordeaux Population Health research Center ~ Bordeaux ~ France

Many diseases are characterized by highly heterogeneous progression patterns across patients. This is the case in Alzheimer's disease and related dementias (ADRD) [1-5]. Despite the abundance of biomarkers now available in ageing cohorts to describe pathological changes involved in AD (such as neurodegeneration, cognitive impairment, functional dependency), this heterogeneity is statistically difficult to apprehend. It requires clustering methods that could handle a large number of biomarker data measured repeatedly at irregular visits over time. In this work, we developed a non-parametric Bayesian clustering model to identify latent clusters of subjects from multivariate longitudinal outcomes and additional cross-sectional variables. The objective was to uncover latent ADRD sub-phenotypes from repeated biomarker data and to characterize their specific physio pathological pathways. We extended the profile regression approach developed by Liverani et al. [6], that links non- parametrically a response vector and cross-sectional variables through cluster membership, to handle multiple longitudinal outcomes measured irregularly over time. The cluster-specific trajectories are described by flexible random-effect models, and the profiles of cross-sectional variables are modeled using cluster-specific generalized linear models. A Dirichlet Process prior was adopted for the mixture distribution to deal with an unconstrained number of clusters, and a variable selection based on importance weighting was used to identify markers that best discriminate between clusters. Parameter estimation is achieved using Monte Carlo Markov Chains.
This method will be applied to the French Three-city cohort [7] to uncover latent sub-phenotypes of ADRD, based on repeated cognitive tests and cross-sectional brain imaging volumes. We expect to identify a small number of meaningful sub-phenotypes that differ according to the sequence and speed of neuropathological degradations. Each group will be associated with a specific pattern of cognitive functions decline and a specific profile of brain atrophy. The model will also highlight the key markers for sub-phenotyping that inform best on the future disease progression. By combining machine learning and biostatistical modeling, this approach extends clustering techniques to large-dimensional longitudinal data encountered in health cohorts. Although motivated by ADRD, it applies far beyond as a mean to identify profiles of trajectories

[1] Reitz, C. Toward precision medicine in Alzheimer's disease. Annals of translational medicine. 2016; 4(6):107.
[2] Rouanet A, Joly P, Dartigues J-F, Proust-Lima C, Jacqmin-Gadda H. Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia. Biometrics, 2016; 72(4):1123 1135.
[3] Proust-Lima C, Philipps V, Dartigues, J-F. A joint model for multiple dynamic processes and clinical endpoints: Application to Alzheimer's disease. Statistics in Medicine, 2019; 38(23):4702- 4717.
[4] Ten Kate M, Dicks E, Visser PJ, van der Flier WM, Teunissen CE, Barkhof F, Scheltens P, Tijms BM; Alzheimer's Disease Neuroimaging Initiative. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain, 2018; 141(12):3443-3456.
[5] Eavani H, Habes M, Satterthwaite TD, An Y, Hsieh MK, Honnorat N, Erus G, Doshi J, Ferrucci L, Beason-Held LL, Resnick SM, Davatzikos C. Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. Neurobiology of Aging, 2018; 71:41-50.
[6] Liverani S, Hastie DI, Azizi L et al. PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes. Journal of Statistical Software, 2015; 64(7):1-30.
[7] 3C Study Group. (2003). Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. Neuroepidemiology, 2003; 22(6):316–325.

Parallel Sessions | Wednesday 30 August 2023

WO3.4 Shared-parameter modelling of longitudinal data allowing for possibly informative visiting process and dropout

Thomadakis C.*[1], Meligkotsidou L.[2], Pantazis N.[1], Touloumi G.[1]
[1]Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens ~ Athens ~ Greece, [2]Department of Mathematics, National and Kapodistrian University of Athens ~ Athens ~ Greece

Dropout in longitudinal studies has been widely studied but the frequency of observations/visits is usually ignored. When the visiting probabilities depend on past marker values and visit times, the visiting process is "at random" (VAR), thus ignorable, whereas, if visiting depends on unknown marker values/characteristics, visiting is "not at random" (VNAR) and should be considered [1]. To handle a VNAR process using linear mixed models (LMMs) for the marker, shared-parameter models (SPMs) are frequently applied. SPMs assume that the marker model and the visiting process are independent conditionally on the random effects. However, it has been shown that SPMs, applied for informative dropout, can lead to seriously biased marker estimates under random dropout [2]. We aim to investigate the performance of standard SPMs for handling VNAR and dropout and propose an alternative SPM. In the proposed SPM, we jointly model marker data along with the visiting and dropout processes. The gap times between visits are modelled conditionally on both a) previously observed marker values and visit times (VAR mechanism) and b) the random effects (VNAR mechanism). A similar sub-model is considered for dropout. We performed a simulation study assuming the visiting probability depends on the most recent marker value, previous gap times, observation time and random effects. Similarly, dropout depended on the two most recent observed marker values and the random effects. The proposed model, being correctly specified, yielded negligible biases (<1.2%) with nominal coverage rates for all parameters. Simplifying the model considering only the most recent marker value for both processes (misspecified visiting/dropout models) resulted in a biased marker's slope estimate (14.7%). Further restricting associations of both visiting and dropout processes only to random effects, yielded a hugely biased slope estimate (61.5%). However, completely ignoring the visiting process but correctly specifying the dropout process led to moderately biased slope estimates (16.4%). Applying the above-mentioned approaches to real CD4 data during untreated HIV infection led to similar conclusions. Ignoring and/or oversimplifying the modelling of an informative visiting process can lead to biased marker estimates although the marker and dropout sub-models are correctly specified.

[1] E. Pullenayegum, L. Lim, Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design, Statistical Methods in Medical Research, 25, 2016, 2992-3014
[2] C. Thomadakis, L. Meligkotsidou, N. Pantazis, G. Touloumi, Longitudinal and time-to-drop- out joint models can lead to seriously biased estimates when the drop-out mechanism is at random, Biometrics, 75, 2019, 58-68

WO3.5 Impact of partial information in longitudinal group-sequential designs on probability of success calculations

Wheeler G.*, Mander A.
GSK ~ London ~ United Kingdom

Interim analyses of clinical trials may help investigators stop a study early if data indicate substantial evidence of efficacy, futility, or harm. The number and timing of interim analyses are usually prespecified at the design stage and may be based on calendar time (e.g. 12 months after first site initiation visit) or the number of participants who have reached a particular follow- up duration (e.g. once 100 participants have follow-up data for the primary endpoint of 12 months). In the latter case, the interim analysis is performed on the subset of trial participants who have completed the required follow-up. However, there will be participants who, having only recently enrolled onto the study, may have observable outcome data at earlier timepoints that will not be accounted for in the interim analysis. The information from these participants may impact our estimation of the interim treatment effects, and more importantly, probability of success calculations.
We derive closed-form expressions for the covariance matrix of the fixed effects parameters from a mixed effects model for repeated measurements, based on the number of patients observable at each visit and treatment arm. We then use simulation to assess the gain in precision, and therefore the increase in the effective sample size at an interim analysis, of including partial information from on-study participants who are not interim-completers relative to the standard interim analysis approach of using interim-completers only. These methods are applied to an ongoing trial comparing Verapamil versus placebo in people with type 1 diabetes, where AUC C-peptide measurements after a mixed meal tolerance test are taken every three months up to the primary endpoint at 12 months. Our methods will allow rapid assessment of how the timing of interim analyses affects the operating characteristics of a trial design, and therefore how these impact the end-of-study outcome predictions and the probability of success for the drug development pathway.

PARALLEL SESSION WO4: CAUSAL INFERENCE 3

WO4.1 Data-driven model building for life-course epidemiology

Helby Petersen A., Ekstrøm C.*
University of Copenhagen ~ Copenhagen ~ Denmark

Life-course data are common in epidemiology where individuals are followed over time, and variables have a known partial temporal ordering. We propose a temporal extension of the PC algorithm[1] for causal discovery of the underlying life-course model from observational data. The method is applied to a dataset analysing the development of depression in Danish men.
We develop the temporal PC algorithm (TPC) for incorporating temporal information in causal discovery. We provide specific suggestions on how these ideas can be used in practice when no oracle property is available, and statistical tests are needed to evaluate conditional independencies to learn the underlying causal structure. We propose a regression-based information loss test to test a necessary (but not sufficient) criterion for conditional independence. Further, we show how a sequence of analyses with varying significance levels may be necessary to infer a "useful" life-course model; and show very high retention rates for edges in the inferred temporal partially directed acyclic graph. Analysis of 2928 Danish men born in 1953, followed from birth until age 65 years with information from several contacts throughout their lives. We apply the TPC algorithm to a total of 33 variables measured in 5 periods over their life course. Depression in early old age is found to be conditionally independent of all remaining variables in the data set given information about depression history in adulthood. Thus, there is no benefit in including any of the childhood information, or other variables measured in adulthood, to understand why a person develops depression in early old age. No causal effects of birth weight or birth length that span longer than youth are found. This is in contrast to myriad studies linking these factors to diabetes, death from ischemic heart disease, and mental health outcomes. We developed the temporal PC algorithm to produce life-course models from observed data. Information from the whole life course is considered jointly and allows for exploratory model building. This facilitates building global models that can provide empirical evidence about presence or absence of causal links between exposures occurring in different periods

[1] P. Spirtes, C. Glymour, Social Science Computer Review, 9 (1), 1991, 62-72

WO4.2 Outcome- versus exposure-wide framework in molecular epidemiology: false positive findings due to correlation

Cadiou S.*
Fundazione Human Technopole ~ Milano ~ Italy

Vanderweele (2017) [1] has underlined the opportunities for causal inference purpose offered by "outcome-wide epidemiology" designs, i.e the simultaneous assessment of associations between a single exposure and multiple outcomes. Within an environmental epidemiology setting, using a rigorous causal analysis of directed acyclic graphs (DAG), he has shown that "outcome-wide" (OW) studies do not suffer from some of the biases arising in the "exposure- wide" (EW) studies, which are now more and more commonly used: especially, in an EWS setting, simultaneous confounding control is almost impossible and results interpretation is complicated because some exposures are likely to affect and mediate the effects of other exposures. Both of these pitfalls are avoided in an OW setting. Here, we would like to analyze how the "outcome-wide" framework is relevant for the challenges of "omics" and molecular epidemiology. Indeed, high-dimensional omics biological layers, such as proteome or methylome, can be considered as intermediate layers between, on the one hand, the environmental and genetic drivers and, on the other hand, the outcome(s) of interest: thus, the EW and the OW settings are both of interest to answer epidemiological questions. Especially, when using multi-steps strategies such as mediation or Meet-in-the-Middle, [2] both EW and OW designs are usually used within the same study on the same data. Here, we wanted through systematic DAG analyses to compare the impact of correlation within a multidimensional layer in terms of expected false positive findings in both an EW and an OW scheme. We showed the superiority of the OW design to avoid false positive findings due to correlation. To our knowledge, this result has not been clearly described; it should be taken into consideration for both the design and the interpretation of environmental and molecular epidemiology studies with high dimensional biological layer(s). We showed especially how this result could help to choose how to order steps of analysis in multi-steps studies, such as the ones using mediation or Meet-in-the-Middle approaches.

[1] Vanderweele TJ. Outcome-wide Epidemiology. Epidemiology (Cambridge, Mass). 2017;28(3):399. doi:10.1097/EDE.0000000000000641
[2] Cadiou S, Basagaña X, Gonzalez JR, et al. Performance of approaches relying on multidimensional intermediary data to decipher causal relationships between the exposome and health: A simulation study under various causal structures. Environment International. 2021;153:106509. doi:10.1016/j.envint.2021.106509

Parallel Sessions | Wednesday 30 August 2023

WO4.3 Resampling-based confidence intervals and bands for the average treatment effect in time-to-event data

Rühl J.*, Friedrich S.
University of Augsburg ~ Augsburg ~ Germany

The g-formula can be used to estimate treatment effects while accounting for confounding bias in observational studies. For time-to-event endpoints, statisticians need to take additional difficulties into account. It is for example not advisable to answer causal questions by hazard ratios, which is why we consider the risk difference instead. This way, a competing risks framework is accommodated on top. The distribution of the associated stochastic process is rather complicated, and hence, confidence intervals are commonly constructed by means of the bootstrap. In certain situations, e.g., when the data lack independence, the classical bootstrap suffers from limitations, though. Furthermore, its execution can be rather time-consuming. This work investigates the performance of different resampling methods in terms of the resulting confidence intervals and bands for the average treatment effect. Apart from Efron's classical bootstrap, we consider an approach that proceeds from the influence function [1] and, since counting processes are inherent to time-to-event analysis, a bootstrap version based on martingales. It is shown that the bootstrap methods approximate the distribution of the stochastic process at hand. We further compare the precision of the different techniques in a simulation study. The results indicate that the wild bootstrap generally yields the most accurate confidence intervals and bands, while for larger sample sizes, all approaches perform similarly. In scenarios where competing events are likely to occur before the event of interest, the influence function approach attains slightly more precise coverage levels, however. Our simulations imply that the wild bootstrap should in general be preferred if the sample size is small and sufficient data on the event of interest have been accrued.

[1] B. M. H.Ozenne, T. H. Scheike, L. Staerk, T. A. Gerds, Biometrical Journal, 62, 2020, 751-763.

WO4.4 Simulating collider stratification bias and an application to the inverse obesity paradox in prostate cancer

Fritz J.*, Stocks T.
Lund University ~ Malmö ~ Sweden

Collider stratification bias is a potential explanation when the association of obesity with the development of a disease is of different magnitude than the association of obesity with disease outcome. The extreme case, where obesity is a risk factor for the disease, but is associated with improved survival in the diseased, as observed for example for cardiovascular disease and several types of cancer, is known as the "obesity paradox" and has been well studied. For prostate cancer (PCa), observational research has consistently reported an "inverse" form of the obesity paradox; body mass index (BMI) is inversely associated with PCa diagnosis (hazard ratio [HR] ~0.9 per 5-kg/m2 increase), but positively associated with PCa-specific death (HR~1.2). Collider bias as a potential explanation in this specific setting is unexplored. Simulating multiple scenarios with different choices of input parameters (i.e., effect of obesity on disease diagnosis; strength of the unadjusted disease diagnosis and outcome confounder; disease and outcome incidence), we investigated the potential for collider bias. Using plausible choices of input parameters for the PCa example, simulations showed that collider bias is unlikely to distort the PCa death HR by more than ±0.02 in case only analyses. The reason for this finding is that the magnitude of collider bias is sensitive to the strength of the effect of BMI on disease diagnosis. In addition to assuming the existence of a strong unmeasured risk factor for both PCa diagnosis and death (which could be genetical risk, for example), a substantially stronger effect of BMI on PCa diagnosis (i.e., HR≤0.67 instead of HR~0.9) would be needed to introduce relevant collider bias. Our simulations suggest that collider bias is an insufficient explanation for the inverse obesity paradox in PCa. Even in large observational studies with a couple of hundred thousand participants, random variability in effect estimates of the relationship between obesity and PCa- specific death in case only analyses outweighs the maximum magnitude of collider bias. Classical confounding and other sources of bias, such as detection bias, heterogeneity of disease bias, and model misspecification, bear a higher

WO4.5 Evaluate application of causal machine learning to adaptive enrichment clinical trials

Yin J.*[1], Ngufor C.[2], Zhang N.[3], Ross J.[4], Noseworthy P.[5], Yao X.[3]
[1]Division of Clinical Trials and Biostatistics, Mayo Clinic ~ Rochester, MN ~ United States of America, [2]Department of AI and I, Mayo Clinic ~ Rochester, MN ~ United States of America, [3]Department of Health Care Delivery Research, Mayo Clinic ~ Rochester, MN ~ United States of America, [4]Department of Medicine, School of Medicine, Yale University ~ New Haven, CT ~ United States of America, [5]Department of Cardiology, Mayo Clinic ~ Rochester, MN ~ United States of America

Pre-specifying patient subgroups of interest in randomized clinical trials (RCTs) can be challenging without prior knowledge of biological mechanisms that lead to heterogeneous treatment effects (HTE). Unlike in conventional RCTs, where HTE evaluation occurs after the trial is completed, machine learning (ML) could potentially identify HTE as the data accumulate during the trial and facilitate subsequent targeted enrollment to enrich certain subgroups that are more likely to respond to the intervention. We used a hybrid causal tree [1] and targeted maximum likelihood (TMLE) method to discover HTE subgroups for adaptive enrichment. The algorithm partitioned patients into clusters such that it maximizes HTE between clusters while minimizing it within the clusters. We identified two completed RCTs as use cases: an oncology treatment trial and a cardiovascular pragmatic trial. The proposed method identified key factors that determine HTE and estimated the treatment effect within each subgroup, thereby facilitating subsequent adaptive enrichment. Resampling method was also used to simulate treatment effect as if the subgroups were enriched based on the HTE identified during the early trial period. The ML-identified HTE at interim analysis is consistent with the HTE identified at the end of the RCT, although some subgroups with signals may not be detected at interim analysis due to their smaller sample size and/or smaller effect size. We therefore recommend excluding patient subgroups who would be predicted to have a detrimental effect to avoid patient attrition in clinical trials.

[1] Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences. 2016 Jul 5;113(27):7353-60.

PARALLEL SESSION WO5: MISSING DATA

WO5.1 Substantive model compatible multilevel multiple imputation: a joint modeling approach

Quartagno M.[1], Carpenter J.*[2]
[1]MRC Clinical Trials Unit at UCL ~ London ~ United Kingdom, [2]London School of Hygiene & Tropical Medicine ~London ~ United Kingdom

Motivated by data from a cluster randomised trial of interventions to improve the paediatric admission process in Kenyan hospitals [1], we present a method for multilevel substantive model compatible multiple imputation (SMC-MI) [2]. We report selected results of an extensive simulation strategy to explore the performance of the method when we have (a) missing data are at the individual or cluster level, (b) non-linear effects in the substantive model and (c) a moderately mis-specified imputation model. Finally, we apply the imputation methods to the motivating example. Our results show superior results compared to standard joint-model imputation, particularly in presence of large variation in random slopes, non-linearities, and interactions. The results appear robust to slight mis-specification of the imputation model for the covariates. When imputing level 2 data, sufficient clusters have to be observed in order to obtain unbiased estimates of the level 2 parameters. We conclude that our new approach is preferable when the substantive analysis has complexities such as non-linearities, interactions or random slopes. Our approach is implemented in the R package jomo.

[1] Ayieko P, Ntoburi S, Wagai J, Opondo C, Opiyo N, Migiro S, Wamae A, Mogoa W, Were F, Wasunna A, Fegan G, Irimu G and English M (2011) A multifaceted intervention to implement guidelines and improve admission paediatric care in Kenyan district hospitals: a cluster randomised trial. PLoS Med, 8:15.
[2] Quartagno, M and Carpenter, JR (2022) Substantive model compatible multilevel multiple imputation: a joint modelling approach. Statistics in Medicine, https://doi.org/10.1002/sim.9549

Parallel Sessions | Wednesday 30 August 2023

WO5.2 Handling missing data in binary variables with low prevalence

Gao H.*, Pavlou M., Omar R.
University College London ~ London ~ United Kingdom

Multiple imputation by chained equations (MICE) is frequently used to handle missing data, a common and challenging issue in health studies. Missingness in binary explanatory variables with low prevalence categories combined with large number of variables in the imputation model may cause model overfitting and separation problems resulting in extreme and inaccurate imputed values. This may result in biased and inefficient estimates of the regression coefficients in analysis models. The aim is to compare the performance of MICE based approaches for handling missing values in presence of model overfitting and separation. Penalised regression is used to alleviate problems of model overfitting and fitting problems due to separation. Additionally, an augmentation procedure [1] has been proposed where extra observations are added to the data before imputation to avoid separation. We conducted simulation studies to compare (a) standard MICE with logistic regression using only variables from the analysis model, all imputation variables (some of which are true predictors of missingness) with and without augmentation and including only variables selected by univariate screening and (b) penalised imputation models (Ridge, Lasso and Firth) using all candidate imputation variables. The performance of the imputation methods is evaluated by:1) quality of the imputed values and 2) bias and efficiency in the estimates of the regression coefficients in the analysis models. A comparison is done with complete case analysis. The simulation study results showed that penalised methods improved the accuracy of imputed values, and the bias and efficiency of parameter estimates of the analysis model in presence of overfitting and separation problems. Lasso and Ridge performed better than Firth when there is model overfitting without separation. In presence of separation, Firth method performed better when the variables in the imputation model were weakly correlated while Lasso and Ridge performed better when variables were strongly correlated. Augmentation procedure produced biased estimates in presence of separation problems. Imputation using penalised regression improved the accuracy of imputed values and consequently the bias and efficiency of regression parameters in the analysis model in presence of overfitting and separation. White, I.R., Daniel, R. and Royston, P., 2010. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. Computational statistics & data analysis, 54(10), pp.2267-2275.

WO5.3 Advanced bayesian joint modelling for time-to-event subgroup analysis with partially missing subgroup status

Bratton D.*[1], Psioda M.[2]
[1]*GSK ~ Stevenage ~ United Kingdom,* [2]*GSK ~ Philadelphia ~ United States of America*

Complete-case analysis of subgroup-specific treatment effects may lead to biased estimates when subgroup status is missing for some participants. Multiple imputation can be used to impute missing subgroup status, however, typical multiple imputation approaches might not always allow the use of an imputation model which is compatible with the analysis model, potentially leading to bias. In addition, predictors of the subgroup variable should be included in the model to make the missing at random assumption more plausible, but this can be challenging if there is no a priori knowledge. In an example, we apply a Bayesian joint model to the outcome of overall survival (OS) and a binary baseline subgroup variable X to perform simultaneous imputation and analysis of the subgroup treatment effects. The joint model is factorised as a conditional model for OS given X, analysed using a stratified Cox proportional hazards model, and a marginal logistic regression model for X. To identify predictors of X, a predictive modelling approach is incorporated into the joint model by including a large number of baseline covariates in the model for X and applying regularised horseshoe shrinkage priors to the regression coefficients to address potential overfitting. This approach allows the use of an imputation model which is compatible with the primary analysis model, and the dynamic approach to variable subset selection appropriately handles instability due to flat likelihood of the fitted marginal model for X and avoids separate modelling of X (i.e., a multi-step process). Finally, the fully Bayesian solutions allows for shrinkage estimation of the subgroup effects to be scaffolded on top of the joint-modelling solution, further demonstrating the versatility of Bayesian joint modelling for complex modelling challenges.

WO5.4 Imputation of longitudinal patient reported outcomes in the presence of death and other intercurrent events

Thomassen D.*[1], Roychoudhury S.[2], Delphin Amdal C.[3], Liu L.[4], Musoro J.[5], Sauerbrei W.[6], Le Cessie S.[7], Goetghebeur E.[4]
[1]*Department of Biomedical Data Sciences, Leiden University Medical Center ~ Leiden ~ Netherlands,* [2]*Pfizer Inc. ~ New York ~ United States of America,* [3]*Department of Oncology, Oslo University Hospital ~ Oslo ~ Norway,* [4]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University ~ Ghent ~ Belgium,* [5]*European Organisation for Research and Treatment of Cancer (EORTC) Headquarters ~ Brussels ~ Belgium,* [6]*Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg ~ Freiburg ~ Germany,* [7]*Department of Clinical Epidemiology, Leiden University Medical Center ~ Leiden ~ Netherlands*

Patient reported outcomes (PROs) such as health-related quality of life and symptom severity can correlate with the occurrence of intercurrent events (e.g., death, disease progression): for instance, a decrease in health-related quality of life may precede death. At the same time, intercurrent events may impede PRO measurement, leading to missing data. We encountered this in a single arm trial where health-related quality of life was repeatedly measured and where the mortality rate was high. Since PROs after death are not defined, our estimand was the mean PRO at each time point in those who were still alive, regardless of treatment discontinuation or disease progression[1]. Marginal means from a standard linear mixed model (LMM) were not appropriate here, as they are based on implicit imputation of missing PROs, ignoring the relation with intercurrent events and extrapolating PROs after death. Therefore, our aim was to derive methods to deal with missing PRO data before death, assuming missingness at random conditional on the longitudinal trajectory of the PRO; the individual timing of progression of disease (PD), treatment discontinuation (TD) and death.
An important challenge was censoring: the times of PD and death were not observed for all patients. We imputed missing PRO-data until death or censoring, after which we reweighted for censoring in the analysis. To this end, we formulated single and multiple imputation models: LMMs, multivariate normal models in MICE and a predictive mean matching-model in MICE[2]. The models conditioned on time-varying variables specifying the time to death, PD and TD; time-varying indicators of the event, and censoring indicators. We varied the coding of the time- to-event variables (linear, splines, a change of slope at event times) and the addition of interactions with the indicators in model specification. Imputation conditional on intercurrent events generally led to lower estimated mean health- related quality of life (while alive) over time compared to available data; particularly with increasing incidence of intercurrent events. Variations in the imputation models led to noteworthy differences in their estimates and standard errors. We demonstrate how the imputation models' assumptions about the relation between the PRO and intercurrent events ct the resulting estimates. Abstract submitted on behalf of SISAQOL-IMI Work Package 3.

[1] Kurland, B. F., Johnson, L. L., Egleston, B. L., & Diehr, P. H. (2009). Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. Statistical science : a review journal of the Institute of Mathematical Statistics, 24(2), 211. https://doi.org/10.1214/09-STS293
[2] Huque, M.H., Carlin, J.B., Simpson, J.A. et al. A comparison of multiple imputation methods for missing data in longitudinal studies. BMC Med Res Methodol 18, 168 (2018). https://doi.org/10.1186/s12874-018-0615-6

Parallel Sessions | Wednesday 30 August 2023

WO5.5 The midoc r package: providing expert guidance and methodology for multiple imputation

Curnow E.*[1], Carpenter J.[2], Heron J.[1], Cornish R.[1], Tilling K.[1]
[1]*Department of Population Health Sciences, Bristol Medical School, University of Bristol ~ Bristol ~ United Kingdom,*
[2]*Department of Medical Statistics, London School of Hygiene and Tropical Medicine, University of London ~ London ~ United Kingdom*

Clinical studies often have missing data, which are commonly handled by multiple imputation (MI). In practice, using MI can be complex, involving multiple decisions which are rarely justified or even documented. For example, what is the optimal structure for the imputation model(s) in a clinical study where the analysis model has an outcome and multiple covariates, each of which is incomplete and needs imputing, plus a choice of many potential variables to include in each imputation model, some of which may themselves be incomplete? Guidance is needed, particularly for researchers with little formal training in statistical analysis of missing data. We are developing the midoc (multiple imputation doctor) R package to address this, focusing on the following objectives: Resolve questions around bias due to incorrect specification of the imputation model Develop methodology and diagnostics to identify optimum variable selection for the imputation model We show how the midoc R package will work in practice using examples from the Avon Longitudinal Study of Parents and Children [1, 2], a UK biomedical birth cohort, illustrating situations in which mis-specification of the imputation model can lead to bias, where bias is due to: Assuming linear relationships in the imputation model (often the default in software packages) Inclusion of collider variables in the imputation model Inclusion of variables predictive of missingness but not the missing data We demonstrate that by correctly specifying the imputation model, several sources of bias can be avoided. The midoc R package guides researchers through their analyses with missing data, examining the structure of the dataset to advise on whether multiple imputation is needed, and if so how to perform it. By providing documented decisions and code, use of midoc will increase reproducibility and transparency of analyses.

[1] A. Boyd et al., "Cohort Profile: The 'Children of the 90s'; the index offspring of the Avon Longitudinal Study of Parents and Children (ALSPAC)," Int. J. Epidemiol., vol. 42, no. 1, pp. 111- 127, 2013.
[2] A. Fraser et al., " Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort," Int. J. Epidemiol., vol. 42, pp. 97-110, 2013.

PARALLEL SESSION WO6: EPIDEMIOLOGY 2

WO6.1 A case-control study to evaluate blood bacterial dna in the intestinal adenoma-carcinoma sequence

Speciani M.C.[1], Mignozzi S.[1], Gargari G.[2], Guglielmetti S.[2], Bonzi R.[1], Ferraroni M.[1], La Vecchia C.[1], Rossi M.*[1]
[1]*Department of Clinical Sciences and Community Health, Università degli Studi di Milano – Milan, Italy ~ Milano ~ Italy,* [2]*Department of Food, Environmental and Nutritional Sciences (DeFENS), Division of Food Microbiology and Bioprocesses; Università degli Studi di Milano – Milan, Italy. ~ Milano ~ Italy*

Microbe-associated molecular patterns exacerbate inflammation in colorectal carcinogenesis, mediating the loss of epithelial barrier function that can lead to an increase of bacterial translocation of intestinal microbes into bloodstream. A new case-control study approach was used to test this hypothesis and to evaluate whether early diagnosis of CRC may be defined by mean of metagenomic analyses on blood, with relevant implications on public health level. Appropriate statistical techniques will be applied to address typical methodological challenges of metagenomics data and a focus of those deriving from this original study design will be presented. We conducted a case-control study recruiting 100 incident histologically confirmed CRC, 100 intestinal adenoma (IA) and 100 free from IA/CRC (hereafter referred to as healthy) subjects, frequency–matched (1:1:1) by centre, sex and age in two hospitals of Milan, Italy, during 2017- 2019. Subjects were selected among patients aged 20-85 with a scheduled colonoscopy. Participants were interviewed through a validated questionnaire and blood samples were collected before colonoscopy. qPCR quantification and taxonomic profiling by Illumina MiSeq sequencing of the 16S rRNA gene copies were performed. We applied conditional logistic regression and further statistical analyses based on negative binomial distribution normalization and Random Forest algorithm to evaluate blood bacterial DNA load and profiling and selected epidemiological information according to CRC cancer, IA and heathy subjects. Our data confirm the presence of bacterial DNA in blood in healthy adults and an overrepresentation of blood 16S rRNA gene copies in colon cancer, especially in right colon, as compared to tumor-free controls. A set of factors discriminating between CRC and controls/IA with an accuracy of 0.70 was identified. This research can serve as a pilot study to offer a validated framework of standardised procedures in order to elaborate microbiological data and to design future epidemiological studies involving microbial ecosystem analysis of potential interest for early CRC diagnosis. Strengths and limitations of the original aspects of the study design as well of the corresponding statistical methods used will be discussed. This research is also designed to contribute with original information to the on-going international scientific debate on the causes of CRC and its prevention.

Mutignani M, Penagini R, Gargari G, Guglielmetti S, Cintolo M, Airoldi A, Leone P, Carnevali P, Ciafardini C, Petrocelli G, Mascaretti F, Oreggia B, Dioscoridi L, Cavalcoli F, Primignani M, Pugliese F, Bertuccio P, Soru P, Magistro C, Ferrari G, Speciani MC, Bonato G, Bini M, Cantù P, Caprioli F, Vangeli M, Forti E, Mazza S, Tosetti G, Bonzi R, Vecchi M, La Vecchia C, Rossi M.
Blood Bacterial DNA Load and Profiling Differ in Colorectal Cancer Patients Compared to Tumor-Free Controls. Cancers (Basel). 2021;13:6363.

Speciani MC, Cintolo M, Marino M, Oren M, Fiori F, Gargari G, Riso P, Ciafardini C, Mascaretti F, Parpinel M, Airoldi A, Vangeli M, Leone P, Cantù P, Lagiou P, Del Bo' C, Vecchi M, Carnevali P, Oreggia B, Guglielmetti S, Bonzi R, Bonato G, Ferraroni M, La Vecchia C, Penagini R, Mutignani M, Rossi M. Flavonoid Intake in Relation to Colorectal Cancer Risk and Blood Bacterial DNA. Nutrients. 2022;14:4516.

Parallel Sessions | Wednesday 30 August 2023

WO6.2 Integrating data across multiple sites to examine associations between a metal mixture and child cognition

Rosa M.[1], Foppa Pedretti N.[1], Goldson B.[1], Mathews N.[1], Merced--Nieves F.[1], Xhani N.[1], Bosquet Enlow M.[2], Gershon R.[3], Ho E.[3], Huddleston K.[4], Wright R.[1], Wright R.[1], Colicino E.*[1]
[1]Mount Sinai ~ New York ~ United States of America, [2]Harvard University ~ Boston ~ United States of America, [3]Northwestern ~ Chicago ~ United States of America, [4]George Mason University ~ Fairfax ~ United States of America

Exposure to multiple metals may impact child neurodevelopment. Data integration of diverse epidemiologic studies can provide enhanced exposure, contrast and statistical power to examine associations between metal mixture exposures and child cognition. Prior studies combined studies identifying an overall mixture-outcome association without accounting for differences across studies. To develop and apply a novel Hierarchical Bayesian Weighted Quantile Sum (HBWQS) regression to combine data from two different studies across three sites in order to examine associations between prenatal exposure to metals and cognitive functioning in early to middle childhood.
Analyses included study participant data from 419 mother-child dyads enrolled in the dual-site PRISM cohort, based in New York City and Boston, and in the First Thousand Days of Life (FTDL) cohort, based in Northern Virginia, participating in the Environmental influences on Child Health Outcomes (ECHO) national consortium. Arsenic (As), Cadmium (Cd), Manganese (Mn), Lead (Pb) and Antimony (Sb) were measured in maternal urine collected during the second or third trimester of pregnancy. Child cognitive functioning was assessed using the National Institute of Health (NIH) Toolbox Cognition Battery when children were between 3 and 11 years of age. We examined associations between urinary metal mixtures and cognition scores using the HBWQS and compared the findings to those of the individual cohorts and traditionally pooled analyses.
The HBWQS regression showed a negative association between the metal mixture and the early childhood cognition composite score in all cohorts: PRISM Boston (β: -2.93; 95% CrI: - 5.66, 0.44; 90% CrI: -5.17, -0.28), PRISM NYC (β -3.40; 95% CrI: -7.07, 0.37; 90% CrI: -6.25, - 0.23) and FTDL (β: -3.63; 95% CrI: -6.96, 0.02; 90% CrI: -6.30, -0.86). The largest contributor to the mixture effect was As (48%), followed by Pb (17%). We did not detect these associations in the individual cohorts or traditionally pooled models.
We demonstrate the utility of this novel statistical approach, which allowed us to increase power while accounting for study heterogeneity and detect associations between prenatal metal mixtures exposure and cognitive outcomes in childhood. Given the ubiquity of metals exposure, interventions aimed at reducing exposure during pregnancy may improve cognitive outcomes in children.

Gibson EA, Goldsmith J, Kioumourtzoglou M-A. Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results. Current environmental health reports 2019. Valeri L, Mazumdar MM, Bobb JF, et al. The Joint Effect of Prenatal Exposure to Metal Mixtures on Neurodevelopmental Outcomes at 20-40 Months of Age: Evidence from Rural Bangladesh. Environ Health Perspect 2017;125(6):067015.
Tanner E, Lee A, Colicino E. Environmental mixtures and children's health: identifying appropriate statistical approaches. Curr Opin Pediatr 2020;32(2):315-320. Wright RO, Baccarelli A. Metals and neurotoxicology. J Nutr 2007;137(12):2809-13.
Colicino E, Foppa Pedretti N, Busgang S, Gennings C. Per- and poly-fluoroalkyl substances and bone mineral density: results from the Bayesian weighted quantile sum regression. medRxiv 2019:19010710. DOI: 10.1101/19010710.

WO6.3 Hierarchical clustering for the evaluation of transitivity assumptio in a network of interventions

Spineli L.*
Hannover Medical School ~ Hannover ~ Germany

Transitivity, also known as similarity, is the cornerstone assumption underlying network meta- analysis [1]. Transitivity states that pre-specified clinical and methodological characteristics of the synthesised trials that act as effect modifiers are similarly distributed across the observed comparisons of the network. The validity of transitivity is necessary to ensure that the results obtained from network meta-analysis are credible [2]. There are currently no universal recommendations for the methods to assess the transitivity assumption regarding the distribution of the effect modifiers. We propose hierarchical clustering to evaluate the transitivity assumption in a network of interventions. Based on a set of effect modifiers, we will investigate if there are distinct homogeneous clusters of observed comparisons that may signal possible intransitivity in the network. We also propose the framework of pseudo-studies to address comparisons with a single study when performing clustering. Finally, we introduce the network of comparisons as a visualisation tool to aid in detecting hot spots of intransitivity based on dissimilarity measures. Then the impact of intransitivity on the network meta-analysis results is investigated by conducting network meta-regression with the clusters as the covariate. We apply our proposed approach to a collection of published systematic reviews with network meta-analysis for the primary outcome. Our approach offers a simple framework for objectively evaluating the transitivity assumption. The proposed framework considers all effect modifiers simultaneously using well-established statistical methods, offering numerous advantages over multiple statistical testing.

[1] S. Baker, B. Kramer. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? BMC Med Res Methodol. 2, 2002, 13.
[2] G. Salanti. Indirect and mixed-treatment comparison, network, or multiple-treatments meta- analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 3, 2012, 80–97.

WO6.4 Bayesian unanchored additive models for component network meta-analysis

Wigle A.*, Béliveau A.
University of Waterloo ~ Waterloo ~ Canada

Component Network Meta-Analysis (CNMA) models are an extension of standard network meta-analysis models which account for the use of multicomponent treatments. Popularity of CNMA models to estimate relative treatment effects has increased in recent years [1]. Limited work has been done to rigorously compare established CNMA models. The aim of this research is to clarify an important distinction between existing models which was previously overlooked and show the implications of each approach. First, by introducing a unified notation, we establish that currently available CNMA methods use two different approaches to defining the relative effects, a distinction which was unnoticed in the literature until now. We term the two approaches anchored and unanchored. We show that an anchored model can provide a poor fit to the data if it is misspecified, while an unanchored model is robust to misspecification. In light of these findings, we present two Bayesian unanchored CNMA models and present them under the unified notation. An extensive simulation study examining bias, coverage probabilities, treatment rankings, and model selection methods confirms the favourable performance of unanchored models and the poor performance of misspecified anchored models. The use of anchored and unanchored CNMA models is demonstrated on a real dataset, where we discuss the strengths and weaknesses of each approach. There are two existing approaches to CNMA, which are anchored and unanchored models. Anchored models for CNMA are more prone to misspecification, which leads to biased estimates, poor coverage probability, and incorrect treatment rankings. Unanchored models are robust to misspecification. Practitioners should be aware of whether the model they use is anchored or unanchored and avoid using a misspecified anchored model.

[1] M. Petropoulou, O. Efthimiou, G. Rücker, G. Schwarzer, T. A. Furukawa, A. Pompoli, H. L. Koek, C. Del Giovane, N. Rodondi, D. Mavridis, PLoS ONE, 16, 2, 2021, e0246631.

Parallel Sessions | Wednesday 30 August 2023

WO6.5 Generalized fused lasso for treatment pooling in network meta-analysis

Beliveau A.*, Kong S.
University of Waterloo ~ Waterloo ~ Canada

This work develops a generalized fused lasso (GFL) approach to fitting network meta-analysis (NMA) models that penalize pairwise differences between treatments. Treatments that are not significantly different end up being pooled, which is an interesting avenue for improving interpretability of NMA models and potentially avoiding biases in treatment rankings [1] or multiple comparison issues [2]. From a statistical perspective, fitting NMA models within the GFL framework requires that we express NMA models as contrast-based weighted regression models because GFL assumes equal variances. Incorporating multi-arm trials and heterogeneity (while preserving the equal variance assumption) presents challenges. In fixed-effect models that do not have multi-arm studies, we can define least square weights as the inverse of the variance for each contrast. Multiplying the data vector and design matrix by the square root of the weight matrix allows us to fit contrast based NMA models within the GFL framework. We demonstrate how to construct the GFL penalty matrix such that every pairwise difference is penalised. We extend the approach to random-effect models by estimating heterogeneity and adding it to the diagonal of the weight matrix. We expect to be able to extend the approach to multi-arm models by incorporating the correlation between contrasts into the weight matrix as well. Cholesky decomposition will be used to decompose the weight matrix so it can be used to linearly transform the data vector and design matrix for use within the GFL framework. We applied our approach to the Parkinson's and Diabetes datasets available from the netmeta R package. For each dataset, we used a selection criterion such as the AICc to identify the tuning parameter(s) that fit the data best. The full model (which is standard in NMA and equivalent to setting the tuning parameter to 0) was not favored compared to the best fitting GFL models. Generalized fused lasso can be used to fit NMA models to allow treatment pooling. This pooling can facilitate interpretations of the results and streamline treatment rankings and comparisons. The method is straightforward to implement in R and runs quickly.

[1] T. Kribet, D. Richer, J. Beyene, Clinical Epidemiology, 6, 2014, 451-460.
[2] O. Efthimiou, I.R. White, Research Synthesis Methods, 11, 2020, 105-122.

PARALLEL SESSION WO7: CLINICAL TRIALS 8

WO7.1 Utilizing co-primary endpoints to test for clinically significant differences in progression-free survival

Leblanc M.*
Fred Hutchinson Cancer Center ~ Seattle ~ United States of America

Immunotherapies and targeted therapies have brought improvements in survival for cancer patients but introduced opportunities for the design and interpretation of clinical trials. The hazard ratio or the associated statistical tests do not address treatment effects well when treatments have delayed effects or cure fractions, or even transient effects. For many cancers, clinically significant changes in therapies often hinge on long-term improvements in progression- free survival (PFS). Furthermore, having an endpoint that is less sensitive to precise long-term scheduling and scanning for PFS would be desirable to keep the trial conduct more practical. We study a co-primary endpoint strategy for randomized trials. One endpoint evaluates the treatment effect using a powerful test for proportional hazards, and another co-primary endpoint that focuses on a clinically meaningful and interpretable impact of the new treatment. To address practical needs of the trial – we also want the second endpoint to be less sensitive to modest variation in the assessment schedules than the logrank test. For example, we study the second co-primary objective based on a comparison of a landmark long-term survival (of progression-free survival) for two treatment arms. However, based on our prior work (Zhao et al Stat in Med, 2019) we also consider alternative measures, which represent the chance of longer life with the new treatment. The key aspect of the second co-primary is that is does not depend on proportional hazards assumption. We use of the joint distribution of associated co-primary test statistics to ensure overall type I error rate control while maximizing statistical power. Modified group-sequential monitoring for strategies are studied by simulating from the joint distribution.
We evaluate the performance based on designs motivated by an upcoming SWOG Cancer Research Network Phase III trial and we explore the impact on several previous trials conducted in our group. The co-primary testing provides a strategy to assess the treatment effects in terms of clinically meaningful long term treatment effects - yet retaining the power to detect a large proportional hazards effect. Depending on the alpha allocation, only modest increases in the trial sample sizes are needed comparted to the logrank testing alone. Reference- Zhao YQ, Redman, MW, LeBlanc M, Quantifying treatment effects using the personalized chance of longer survival, Statistics in Medicine, 2019 38:5317-5331

WO7.2 Analysis of multicentre trials: limiting the effect of centre heterogeneity on the marginal treatment effect

Payne M.*, Emsley R.
King's College London ~ London ~ United Kingdom

Multi-centre trials are conducted to increase generalisability and ensure timely recruitment of participants. However, it is possible that the treatment effect varies across centres, known as treatment effect heterogeneity. In the presence of this heterogeneity, the marginal treatment effect (MTE) for the whole population is no longer equal to the conditional effect, within centres. The impact of centre effects may be highly influential on the marginal effect and the impact of ignoring this heterogeneity is not generally considered. The aim of this paper is to investigate the impact of heterogeneity on the MTE and to suggest the most unbiased estimator for a parallel group trial. A Monte Carlo simulation study was conducted to compare the performance of five statistical approaches that allow for between-centre heterogeneity when considering the MTE: A fixed effect model, two fixed effects models with interactions, a random intercept model and a random intercept and slope model. We vary the number of centres, size of centres and treatment effect to replicate real trial scenarios and elicit practical advice. The degree of site heterogeneity was varied in our data generating mechanisms. We estimated the MTE for all scenarios and evaluated the bias, efficiency, mean squared error and coverage of our estimators. In a majority of scenarios, all models gave unbiased estimates; however, we highlight some situations where it is not advisable to use particular models due to poor performance. We apply the methods to a recent clinical trial in mental health. In terms of trial design, multi-centre trials should aim to recruit an approximately equal number per centre to limit the effect of informative cluster size. When analysing a trial, fixed effects models with unweighted interactions and random intercept models are unlikely to incur bias in all scenarios, even in the presence of treatment effect heterogeneity. When specifying our primary analysis at the start of trial, we advise using a model that will be unbiased irrespective of heterogeneity. ICH E9 guidelines advise not to include an interaction in the main model, therefore we advise using a random intercept model for estimating the marginal treatment effect. B. C. Kahan, T. P. Morris, Statistics in Medicine, 32, 2013, 1136-1149.

WO7.3 Determining the minimum duration of treatment in tuberculosis: an order-restricted non-inferiority design

Serra A.*[1], Mozgunov P.[1], Davies G.[2], Jaki T.[3]
[1]University of Cambridge ~ Cambridge ~ United Kingdom, [2]University of Liverpool ~ Liverpool ~ United Kingdom, [3]University of Regensburg ~ Regensburg ~ Germany

In this talk, we will present a non-inferiority adaptive design based on a recent proposal of an adaptive order-restricted superiority design (Serra et al. 2022) that employs the ordering assumptions within various treatment durations of the same drug. The proposed design allows to select all durations that are non-inferior to a standard duration of the treatment with high probability and to incorporate the order of treatment effects in the decision-making when no parametric duration-response model is assumed.
The focus of this talk will be on the implementation and evaluation of this design in a specific Tuberculosis (TB) clinical trial setting. Indeed, the challenge in TB setting is not only the identification of drugs that can provide benefits to patients but also the optimisation of the duration of these treatments. While conventional duration of treatment in TB is 6 months, there is evidence that shorter durations might be as effective but could be associated with fewer side effects and will result in better adherence. Together with the general construction of the hypothesis testing and expression for type I and type II errors, we focus on how the novel design was proposed and discussed with the clinical team. We will cover several practical aspects such as choice of the design parameters, randomisation ratios, and timings of the interim analyses. The proposed design has been shown to be capable of evaluating multiple nested durations of a treatment in an efficient manner compared to conventional multi-arm multi-stage adaptive designs. Serra A, Mozgunov P, Jaki T. An order restricted multi-arm multi-stage clinical trial design. Statistics in Medicine 2022. doi: 10.1002/sim.9314

Parallel Sessions | Wednesday 30 August 2023

WO7.4 On the design of biomarker-driven trials with measurement error for time to event outcomes

Halabi S.*, Guo S.
Duke University ~ Durham ~ United States of America

Innovations in molecular and genetic laboratory assays have advanced our understanding of cancer biology by allowing identification of specific molecular and cellular characteristics of tumors (biomarkers). These advances in the understanding of cancer biology have additionally led to therapeutic advances and importantly advances in the type of trial designs used to evaluate these "biomarker-driven" therapies in possibly biomarker-defined populations. Many biomarkers, however, are derived from tissues from patients, and hence their levels may be heterogeneous. As a result, biomarker levels may be measured with error and this would have an adverse impact on the power of a biomarker-driven clinical trial. We propose methods for two types of biomarker-driven trials (enriched and the stratified biomarker designs) to adjust for misclassification errors.
We assume that the prevalence of the biomarker, sensitivity and specificity are known. In both biomarker-driven designs, patients will have their specimens obtained at baseline and the biomarker status will be assessed prior to random assignment. While enriched designs test the efficacy of treatment among biomarker positive patients, the stratified biomarker design can be used to test for a treatment-biomarker interaction in predicting outcome, such as time-to-event outcomes. Through simulations, we show that the naïve tests are biased and provide bias- corrected estimators for computing the sample size and the 95% confidence interval when testing for a treatment-biomarker interaction. We propose sample size formulae that adjust for misclassification and apply it in the designs of phase III clinical trials in renal cancer and prostate cancer.
The proposed methods work well. However, one needs to evaluate other assumptions for the biomarker-driven designs. We have assumed that the prevalence, sensitivity and specificity are known. These are important assumptions, and if these are unknown these parameters could be estimated using a pilot data or empirically from the data. Another point to consider is the reliability of the prevalence of the biomarker. Lastly, we have focused on a single biomarker. If a panel of biomarkers are of interest (such as a classifier), these biomarker-driven designs may be also utilized as long as the cut-off point for the classifier has been validated.

Halabi S, Lin CY, Liu A. On the design and the analysis of stratified biomarker trials in the presence of measurement error. Stat Med. 2021 May 30;40(12):2783-2799. doi: 10.1002/sim.8928.

WO7.5 A superlearner-enforced approach for the estimation of treatment effect in pediatric trials

Azzolina D.*[3], Comoretto R.[1], Gregori D.[2]
[1]University of Turin ~ Turin ~ Italy, [2]University of Padua ~ Padua ~ Italy, [3]University of Ferrara ~ Ferrara ~ Italy

Randomized Clinical Trials (RCT) represent the gold standard among scientific evidence. An RCT is tailored to control selection bias and the confounding effect of baseline characteristics on the treatment effect. However, the trial conduction and enrolment procedures could be a challenging issue, especially n pediatric research. In these research frameworks, the estimation of the treatment effect could be compromised. A potential countermeasure is to develop, on previously collected observational data, predictive models on the probability of the baseline disease. In addition, machine learning (ML) algorithms have recently become attractive in clinical research because of their flexibility and improved performance compared to standard statistical methods. This manuscript proposes an MLT-enforced treatment effect estimation procedure based on an ensemble SuperLearner (SL) approach[1], trained on historical observational data, to control the confounding effect. Method(s) and Results: The RESCUE (REnal SCarring Urinary infEction)[2] served as a motivating example; the study is designed to evaluate the effect of antibiotics combined with steroids in preventing renal scarring on pediatric patients affected by Urinary Tract Infections (UTI)[1]. Based on published information on the prevalence of renal scars, the historical observational study data have been simulated through 10000 Monte Carlo (MC) runs. Hypothetical RCTs have been simulated, for each MC run, by assuming different treatment effects of antibiotics combined with steroids. For each MC simulation, the SL tool has been applied to the simulated observational data. Furthermore, the average treatment effect (ATE), has been estimated on the trial data and adjusted by weighting a binomial logistic regression model for the SL predicted probability of renal scar according to the characteristics of the patient. The simulation results revealed a gain in the ability to truly detect the ATE by considering the SL-enforced estimation compared to the unadjusted estimates for all the algorithms composing the ensemble SL and all sample sizes. The MLT disease prediction tool could be easily implemented and validated by considering observational data; the developed model could be used in an RCT to enforce the treatment effect estimation by adjusting the final estimate for a patient-specific disease risk profile.

[1] Lanera, C.; Berchialla, P.; Lorenzoni, G.; Acar, A.Ş.; Chiminazo, V.; Azzolina, D.; Gregori, D.; Baldi, I. A SuperLearner Approach to Predict Run-In Selection in Clinical Trials. Computational and Mathematical Methods in Medicine 2022, 2022, doi:10.1155/2022/4306413.
[2] Da Dalt, L.; Bressan, S.; Scozzola, F.; Vidal, E.; Gennari, M.; La Scola, C.; Anselmi, M.; Miorin, E.; Zucchetta, P.; Azzolina, D.; et al. Oral Steroids for Reducing Kidney Scarring in Young Children with Febrile Urinary Tract Infections: The Contribution of Bayesian Analysis to a Randomized Trial Not Reaching Its Intended Sample Size. Pediatr Nephrol 2021, doi:10.1007/s00467-021-05117-5.

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO8: SURVIVAL ANALYSIS 7

13:20 - DEDICATED TO PROF. ETTORE MARUBINI

WO8.1 Competing risks, the fine-gray model, and pseudovalues

Therneau T.*
Mayo Clinic ~ Rochester ~ United States of America

For competing risks data, we have found that if there is a moderate fraction of subjects who experence the competing event (1/4 or more) and there are important covariates that modify the risk of that competing event, then the Fine-Gray (FG) model does not work well. Namely, the key underlying assumption of proportional odds (but using c-loglog scale) is badly violated, so badly as to throw into question any utility for an "average" coefficient over time. This is, however, exactly the situation in which understanding the impact of the competing event is critical. The same overall model as is assumed by the Fine-Gray, ordinal c-loglog regression, can also be assesed using pseudo values. This approach allows us to understand the underlying issue with the FG model and suggest more suitable alternatives. Hazard ratios from a multistate hazards model provide complementary information. The issue and solution is illustrated using dementia and death-without-dementia outcomes in the Mayo Clinic Study of Aging.
The Fine-Gray model fails in exactly the situation where competing risks are a serious issue. It is time to retire it in favor of alternatives that are both better and simpler.

WO8.2 Analyzing restricted mean survival time curves using pseudo-values and machine learning

Di Maso M.*, Bravi F., Ferraroni M., Ambrogi F.
Department of Clinical Sciences and Community Health, Branch of Medical Statistics, Biometry and Epidemiology "G.A. Maccacaro", Università degli Studi di Milano ~ Milano ~ Italy

The restricted mean survival time (RMST) is a measure of average survival until a specified time point during the follow-up. Although the ubiquitously use of hazard ratios is invaluable for hypothesis testing, differences in RMST between patient groups may have a plainer clinical interpretation and help to fully elucidate study results. Several recent contributions have been proposed focusing on estimates of differences in RMST at multiple time points, instead of using a single follow-up time horizon. The resulting curves can be used to quantify the association between patient groups through follow-up time, making differences in RMST an advocate tool to measure the association in time to event studies. Furthermore, regression models have been developed to directly evaluate RMST on covariate patterns. These methods are based either on the inverse probability of censoring weighting (IPCW) or on the pseudo-values (PV). In particular, the method based on PV is easily implementable in most of available statistical software and therefore it could be extended to Machine Learning techniques, such as the Deep Neural Network (DNN) proposed by Zhao [1]. We investigated the ability of DNN to account for complex covariate patterns, such as interactions, using literature data as in [2]. The DNN appears to be flexible enough to reproduce the results of the nonparametric approaches when no baseline covariates are considered. Further examples where adjustment for covariates are required were comprised to illustrate the differences in methods using both real data and simulations. Regression methods using standard statistical models and Machine Learning techniques were compared to highlight the differences in their implementation and results.

WO8.3 A comparison of kaplan--meier-based inverse probability of censoring weighted regression methods

Overgaard M.*
Aarhus University ~ Aarhus ~ Denmark

An approach to deal with right-censored survival outcomes in regression analyses is to use weighting with the inverse probability of censoring. This would allow for, for instance, regression analysis of the risk of death within a certain time after treatment. The objective of this study is to compare three separate approaches involving the idea of inverse probability of censoring weighting where the Kaplan--Meier estimator is used for estimating the censoring probability, perhaps in strata of some variable. In more detail, the three approaches involve weighted regression, regression with a weighted outcome, and regression of a jack-knife pseudo-observation based on a weighted estimator. The focus is on the study of large-sample properties. Specifically, it is of interest to investigate which approach will produce the smallest variance of estimators. Furthermore, the challenge of variance estimation as well as the potential benefits of stratification will be investigated. Two of the approaches have been studied in a special case by Blanche et al., [1]. The results of Blanche et al. for these two approaches can be generalized using much the same arguments. The third approach, which is the pseudo-observation approach, can be seen as a special case of what is studied in Overgaard et al., [2]. The influence functions of the three approaches can be stated in a certain way where it is clear how censoring will increase the asymptotic variance in comparison to the uncensored case. A common structure will also reveal exactly what is estimated by the usual sandwich variance estimator in each of the three approaches. Simulations demonstrate the performance of the three approaches in finite samples in certain scenarios. From the asymptotic variance expressions in the three cases, it is clear that examples can be found where either of the three approaches will produce the smallest variance. Which approach will produce the smallest variance will depend heavily on when censoring occurs. The usual sandwich variance estimator will be upwards biased in each of three approaches under the model assumption. Stratification should be useful in weaking the censoring assumption, reducing the asymptotic variance, reducing bias from inappropriate weighting, and reducing the bias of the sandwich variance estimator.

[1] P.F. Blanche, A. Holt, T. Scheike, Lifetime Data Analysis, 29, 2023, pp. 441–482.
[2] M. Overgaard, J. Pedersen, E.T. Parner, Journal of Statistical Planning and Inference, 202, 2019, pp. 112-122

WO8.4 Quadratic inference functions as a new approach to analyze pseudo-observations in survival analysis

Orsini L.*[1], Brard C.[1], Lesaffre E.[3], Dejardin D.[2], Le Teuff G.[1]
[1]CESP, INSERM U1018, Université Paris-Saclay, UVSQ ~ Villejuif ~ France, [2]Product Development, Data Sciences, F. Hoffmann-La Roche AG ~ Basel ~ Switzerland, [3]I-Biostat, KU-Leuven ~ Leuven ~ Belgium

The analysis of pseudo-observations offers an alternative to the Cox model and is particularly interesting for complex survival modeling [1]. Its advantage lies in overcoming the complexity of censored data modeling. Pseudo-observations can be analyzed using Generalized Estimating Equations (GEE). We propose a new approach, the Quadratic Inference Functions (QIF), to analyze Kaplan-Meier-based pseudo-observations. In the frequentist framework, QIF has better theoretical efficiency than GEE. It also allows for goodness-of-fit tests and model selection. In the Bayesian framework, the analysis of pseudo-observations using QIF may offer a new attractive alternative to the Bayesian survival analysis, avoiding the hazard function specification. The QIF approach is an extension of GEE based on the generalized method of moments. A Bayesian version has been proposed [2] using a pseudo-likelihood function. We extended this approach to pseudo-observations analysis with a cloglog link function under different working correlation matrices. We compared the performances of QIF approaches (frequentist and Bayesian) to the Cox, GEE, and Bayesian piecewise exponential models through a simulation study of two-arm randomized clinical trials. Initial simulation results showed convergence issues when using non-informative priors, which have been resolved by truncating priors (except for the treatment effect parameter) to reasonable values. The frequentist QIF gave similar performances compared to GEE. With the Bayesian QIF, a slight overestimation of the treatment effect was observed for small sample sizes. Higher variances were observed with pseudo-observations-based models. For illustration, we used three randomized clinical trials involving Ewing-Sarcoma patients with different sample sizes and prognostics regarding overall survival. QIF approaches gave valid estimates compared to the benchmark approaches. We propose to use the QIF approach to analyze Kaplan-Meier-based pseudo-observations. In the frequentist framework, this approach gave similar estimates compared to GEE. Unlike GEE, the analysis of pseudo-observations using the QIF approach can be extended to the Bayesian framework, creating an alternative to Bayesian survival models without relying on any assumption on the baseline hazard function. Bayesian analysis of pseudo-observations also opens new perspectives in the statistical analysis of complex survival models.

[1] P. K. Andersen, J. P. Klein, and S. Rosthøj, Biometrika. 90, 2003, 15 – 27.
[2] G. Yin, Bayesian Analysis. 4, 2009, 191 – 208.

Parallel Sessions | Wednesday 30 August 2023

WO8.5 Clinical impact and disease dynamics in competing risks: an analysis of two historical clinical trials

Biganzoli G.*, Demicheli R., Marano G., Boracchi P.
Unit of Medical Statistics, Biometry and Epidemiology, Department of Biomedical and Clinical Sciences (DIBIC) "L. Sacco" & DSRC, LITA Vialba campus, Università degli Studi di Milano ~ Milano ~ Italy

In a longitudinal clinical trial the aim is to evaluate treatment and prognostic/predictive factors effects. This implies measures of clinical impact (e.g., relative risk during follow-up) and measures of disease dynamics (the shape of hazard function). In presence of competing risks, the endpoint is based only on subsets of possible events. Relative risk (RR) is the ratio of crude cumulative incidence (CCI) and it is directly linked to sub-distribution hazard (SDH), whose definition in discrete time is interpreted as the probability of observing an event in each time interval given that the event was not observed in the interval before, allowing to evaluate the contribution of disease dynamics to the CCI. The aim was the proposal of a coherent analysis framework on CCI and SDH to assess the patterns of relative risk and the dynamics of distant metastasis (DM).
Regression models on pseudo-observations [1] and models on discrete-time sub-distribution hazard [2] were adopted. Two historical randomized clinical trials of the Milan Cancer Institute were considered: "Milan 1" trial, which compared mastectomy (MAST) with breast conserving therapy (BCT) plus radiation therapy (QUART), and "Milan 3" trial which compared BCT with (QUART) or without (QUAD) radiation therapy. Clinical features such as primary tumour size, axillary lymph node status (N), and menopausal status were accounted. DM competing events were other cancer and all-cause death occurrence. A significant interaction was found between N and treatment. N+ subjects undergoing MAST or QUAD had a higher risk for developing DM compared to N+ subjects receiving QUART. In N- subjects, QUART resulted in a higher risk of DM compared to MAST, after about 8 years, while no evidence of a difference in risk between QUART and QUAD was found. SDH analysis revealed a non-monotonic peaked hazard pattern of DM, with highest hazards associated with N+ subjects receiving MAST or QUART.
The joint analysis of the CCI and the SDH of DM allowed us to speculate that individuals receiving extended surgery are more at risk for DM in the first ten years with peaks of risks according to the clinical dormancy hypothesis.

[1] Ambrogi F, Biganzoli E, Boracchi P. Model-based estimation of measures of association for time-to-event outcomes. BMC Med Res Methodol. 2014 Aug 9;14:97.
[2] Berger M, Schmid M, Welchowski T, Schmitz-Valckenberg S, Beyersmann J. Subdistribution hazard models for competing risks in discrete time. Biostatistics. 2020 Jul 1;21(3):449-466.

PARALLEL SESSION WO9: PREDICTION MODELS 3

WO9.1 Causal blind spots in risk-based decision making

Van Geloven N.*[1], Van Amsterdam W.[3], Cinà G.[4], Didelez V.[5], Krijthe J.[6], Luijken K.[3], Magliacane S.[8], Morzywolek P.[9], Van Ommen T.[3], Peek N.[7], Putter H.[1], Sperrin M.[7], Wang J.[10], Weir D.[10], Keogh R.[2]
[1]Leiden University Medical Center ~ Leiden ~ Netherlands, [2]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [3]University Medical Center Utrecht ~ Utrecht ~ Netherlands, [4]Amsterdam University Medical Centers ~ Amsterdam ~ Netherlands, [5]Leibniz Institute for Prevention Research and Epidemiology - BIPS ~ Bremen ~ Germany, [6]Delft University of Technology ~ Delft ~ Netherlands, [7]University of Manchester ~ Manchester ~ United Kingdom, [8]University of Amsterdam ~ Amsterdam ~ Netherlands, [9]Ghent University ~ Ghent ~ Belgium, [10]Utrecht University ~ Utrecht ~ Netherlands

"Association does not imply causation" is an age-old warning that highlights the importance of distinguishing between prediction tasks and causal inference tasks. While it is widely appreciated that regression coefficients from clinical prediction models can in general not be interpreted causally, many have suggested that risks from prediction models can still support treatment decisions and they are used for this purpose in different clinical practices. The argument behind this is based on the intuition that high-risk patients may be in greater need of treatment compared to low-risk patients, and potentially they could derive greater benefit from treatment. For example in the US the SOFA score is recommended for informing prioritization of scarce resources in the intensive care setting. In this work we study the validity of different types of risk-based decision making in medical applications by placing them in a causal framework. We establish different pitfalls of using individual predictions from standard prediction models to inform treatment decisions. Some pitfalls are easy to spot such as inappropriately interpreting regression coefficients from prediction models causally. Others are more subtle, including the issue of treatment drop-in: treatments that are applied during follow-up of patients in the training data that change the interpretation of the resulting predictions. A second nuanced issue is that of changes in treatment policies over time in the application setting (dataset shift), with a special case of this being changes in treatment decisions related to the use of the prediction model itself (prediction paradox). We demonstrate the potential impacts of ignoring the above pitfalls using examples of risk models used in clinical practice. We also discuss the circumstances under which it may still be valid to use standard risk prediction models, and in which cases temporal updating is helpful. We explain how some of the issues can be resolved by predicting outcomes under hypothetical interventions, for example targeting the outcome risk if a patient would or would not initiate treatment. We conclude that outcome predictions under current treatment policies do not in general offer support for decisions about the prescription of those very same treatments.

Parallel Sessions | Wednesday 30 August 2023

WO9.2 Risk-based decision making: formulating estimands for prediction under hypothetical interventions

Luijken K.*[1], Van Amsterdam W.[1], Cinà G.[2], Didelez V.[3], Peek N.[4], Keogh R.[5], Krijthe J.[6], Magliacane S.[7], Morzywolek P.[8], Van Ommen T.[9], Putter H.[10], Sperrin M.[4], Wang J.[9], Weir D.[9], Van Geloven N.[10]

[1]University Medical Center Utrecht ~ Utrecht ~ Netherlands, [2]Amsterdam University Medical Center ~ Amsterdam Netherlands, [3]Leibniz Institute for Prevention Research and Epidemiology - BIPS & University of Bremen ~ Bremen Germany, [4]University of Manchester ~ Manchester ~ United Kingdom, [5]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [6]Delft University of Technology ~ Delft ~ Netherlands, [7]University of Amsterdam ~ Amsterdam ~ Netherlands, [8]Ghent University ~ Ghent ~ Belgium, [9]Utrecht University ~ Utrecht ~ Netherlands, [10]Leiden University Medical Center ~ Leiden ~ Netherlands

Prognostic models are used to inform medical decisions on interventions. Individuals with high outcome risks could be advised to undergo an intervention while those at low risk could be advised to refrain from it. Standard prognostic models may not provide accurate risks to inform such decisions: the reason for the low estimated risk of an individual may be that similar individuals in the past received the intervention which lowered their outcome risk [1, 2]. Therefore, prognostic models supporting treatment decisions should preferably be estimated under hypothetical interventions, i.e., provide outcome risks belonging to certain defined intervention strategies. In this work, we present an overview of risk estimands to inform intervention decisions. The overview draws from causal inference literature but focuses on individual risk prediction under interventions rather than intervention effect estimation. We distinguish three elements of the decision context that inform the choice of a risk estimand. First, the medical condition may dictate a very short time window during which the decision must be made, or it may allow a time range. For example, in case of an acute stroke, treatment decisions need to be made instantly, whereas many chronic conditions (e.g., diabetes) allow re-evaluated decisions and/or deferred interventions. Second, the relevant prediction horizon can vary. Risks of long-term outcomes may for instance be necessary for side effects associated with long term use of medication. Other outcomes may only require short-term information until the next decision moment. Third, the intervention regime under which predictions are made can have various forms, such as a point treatment, sustained treatment, or dynamic treatment rule. Sometimes a mix of strategies is defined by fixing the intervention until the next decision moment after which 'care as usual' is continued. For example, the decision when a kidney patient should be placed on the waiting list for a donor organ could be informed by the risk of not receiving a transplant in the coming year and after that receiving it under current allocation policies.
Our overview provides guidance on how to formulate risk estimands and illustrates their applicability in different medical contexts.

[1] van Geloven, N., Swanson, S. A., Ramspek, C. L., Luijken, K., van Diepen, M., Morris, T. P., ... & le Cessie, S. (2020). Prediction meets causal inference: the role of treatment in clinical prediction models. European journal of epidemiology, 35, 619-630.
[2] Sperrin, M., Martin, G. P., Pate, A., Van Staa, T., Peek, N., & Buchan, I. (2018). Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. Statistics in medicine, 37(28), 4142-4154.

WO9.3 Improving local prediction models using similarity based data pooling

Behrens M.*[1], Farhadizadeh M.[1], Pechmann A.[2], Kirschner J.[2], Rohde A.[3], Zöller D.[1]

[1]Institute of Medical Biometry and Statistics, University of Freiburg ~ Freiburg ~ Germany, [2]Department of Neuropediatrics and Muscle Disorders, Faculty of Medicine, Medical Center – University of Freiburg ~ Freiburg ~ Germany, [3]University of Freiburg, Department of Mathematical Stochastics ~ Freiburg ~ Germany

Multi-site data collection can provide a larger and more diverse dataset for analysis, leading to improved prediction models. However, differences in patient care, case mix, and other factors can create substantial obstacles in incorporating data from other sources. This variation can generate bias and lead to a decreased prediction performance for the target population if not addressed correctly. Additionally, dealing with limited sample sizes may render methods with a high number of parameters impractical. To resolve this issue, we propose measuring the similarity between the external site and the target site and employing this information to incorporate external sites in a weighted manner. For example, the SMArtCARE registry contains data on patients diagnosed with spinal muscular atrophy (SMA). Physiotherapy is a significant factor in treatment and disease progression evaluation, and it is highly dependent on the data site. Therefore, site-specific prediction models are necessary for predicting when SMA patients reach a mobility milestone or a mobility score at a particular time point. By using pairwise logistic regression models, we estimate the probability of an individual belonging to the target site. Higher weights are assigned to external individuals who are more similar to the target site individuals according to the estimated probability. We standardize these weights across all external sites to incorporate multiple external sites, and since we use weights, this approach can be applied to different types of outcomes and prediction models.
Our proposed method will be evaluated not only through the use of the SMArtCARE registry data but also through an extensive simulation study. We will compare it to classical approaches such as mixed models and regression models with interactions. Through our demonstration, we show that our proposed method can effectively address the challenges arising from the heterogeneity between sites in multi-site data settings. Additionally, it can enhance the prediction performance of models for a target site when dealing with small sample sizes. Our approach quantifies site similarity using logistic regression and utilizes this information to incorporate external data when developing prediction models. Pechmann A, König K, Bernert G, Schachtrup K, Schara U, Schorling D, Schwersenz I, Stein S, ogt S, et al. SMArtCARE-A platform to collect real-life outcome d

Parallel Sessions | Wednesday 30 August 2023

WO9.4 Similarity quantification for small data

Farhadizadeh M.*[1], Behrens M.[1], Rohde A.[2], Daniela Z.[1]
[1]*Institute for Medical Biometry and Statistics, University Hospital Freiburg, Faculty of Medicine, Albert-Ludwigs- University Freiburg ~ Freiburg ~ Germany,* [2]*Freiburg University, Department of Mathematical Stochastics ~ Freiburg ~ Germany*

There are several situations in clinical studies where the research question requires the identification of similarities between two data sets. For instance, suppose the goal is to improve predictive models for a data site with limited observations. In that case, one approach is to include weighted data from external sites based on their similarity while maintaining the original distribution of the target site. To determine these weights and quantify similarity, one can make use of the inverse probability of belonging to the target site using logistic regression. However, this approach may underestimate the similarity if the target site has significantly fewer observations than the external site, even if they are samples of the same population. To address this issue, we propose the use of oversampling techniques. We have developed an iterative process that involves oversampling the target site observations until we observe that the distributions of the target data, before and after including weighted data, remain consistent. To compare the distributions, we utilize Kullback-Leibler divergence and parametric methods. By monitoring the distribution of the target data, we can increase the sample size based on similarity without introducing bias. To evaluate our approach, we conducted a simulation study in two scenarios: one with similar observations in both datasets and another with dissimilar comments with varying degrees of similarity. We compared the results obtained with and without oversampling and assessed the impact of oversampling techniques on the performance in terms of prediction performance. Results indicate that oversampling the small target data can improve the quantification of similarity for obtaining weights in a prediction model, resulting in higher weights when the distributions of datasets are similar but not overestimating the weights when the datasets are dissimilar.
To demonstrate our approach, we applied it to the International Stroke Trial (IST) data, which includes patients with acute stroke from different countries. The goal is to include weighted data from other countries to a country with limited data, while the distribution of target data plus weighted external data remains similar.

[1] D. Hajage, F. Tubach, P. G. Steg, D. L. Bhatt, and Y. De Rycke, "On the use of propensity scores in case of rare exposure," BMC Medical Research Methodology, vol. 16, no. 1, p. 38, Mar. 2016, doi: 10.1186/s12874-016-0135-1.
[2] K. Madjar, M. Zucknick, K. Ickstadt, and J. Rahnenführer, "Combining heterogeneous subgroups with graph-structured variable selection priors for Cox regression," BMC Bioinformatics, vol. 22, no. 1, p. 586, Dec. 2021, doi: 10.1186/s12859-021-04483-z.

WO9.5 Predicting response under interventions in patients with rheumatoid arthritis: a methodological exploration

Gehringer C.*[1], Martin G.[2], Sperrin M.[2], Hyrich K.[1], Verstappen S.[1], Sergeant J.[3]
[1]*Centre for Epidemiology Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, Manchester Academic Health Science Centre, University of Manchester ~ Manchester ~ United Kingdom,* [2]*Centre for Health Informatics, Division of Informatics, Imaging and Data Sciences, University of Manchester ~ Manchester ~ United Kingdom,* [3]*Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester ~ Manchester ~ United Kingdom*

In rheumatoid arthritis (RA), predicting response under different interventions could improve treatment personalisation. Research to date has focused on developing prediction models for non-response to the first-line therapy methotrexate (MTX). However, those at high risk of non- response to MTX may also be unlikely to respond better to alternative therapies, so these models are inadequate for comparing treatment options. In UK clinical practice, as captured in observational data, patients are required to have failed MTX prior to the prescription of more advanced, and more expensive, biologic drugs(1). This makes it challenging to estimate the treatment-naïve (TN) risk of not responding to biologics, which is required to impact clinical practice and enable these drugs earlier on, as patients receiving biologics are already MTX treatment-inadequate-responders (TIR). The target population of TN patients is used in some randomised controlled trials (RCTs) though. The aim of this study is to compare causal prediction approaches for response to MTX and the biologic infliximab (INF) in TIR and TN populations. This will determine how models can be transferred between TIR patients and the target TN population and assess potential bias in our estimand of individual patient response. Observational data came from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis and RCT data from the YODA Project, a collaboration between Yale University and Janssen Pharmaceuticals. The three methodological approaches compared in this study are: 1) a marginal structural model of response to MTX vs INF developed in a TIR population and evaluated in the target population of TN patients, 2) a causal prediction model based on the "risk modelling"(2) approach using randomised data of TN patients and evaluating this in the TIR population, and 3) a causal prediction model that corrects for potential bias in the TIR population using the trial estimate(3). Results comparing these modelling approaches in terms of predictive performance and net benefit for treatment decision will be presented alongside methodological recommendations. This study contributes a novel methodological approach to predicting treatment outcomes in the field of rheumatology, as previously developed prediction models of MTX outcomes have not considered response to alternative interventions.

[1] National Institute for Health and Care Excellence. Rheumatoid arthritis in adults: management. 2020;33.
[2] Kent DM, van Klaveren D, Paulus JK, D'Agostino R, Goodman S, Hayward R, et al. The PATH Statement Explanation and Elaboration Document. Ann Intern Med. 2020 Jan 7;172(1):W1–25.
[3] Colnet B, Mayer I, Chen G, Dieng A, Li R, Varoquaux G, et al. Causal inference methods for combining randomized trials and observational studies: a review [Internet]. arXiv; 2022 [cited 2023 Jan 2]. Available from: http://arxiv.org/abs/2011.08047

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO10: CAUSAL INFERENCE 4

WO10.1 From data to decisions: how effects of intervening variables can guide policies

Wen L.[2], Sarvet A.[1], Stensrud M.*[1]
[1]*Ecole Polytechnique Federale de Lausanne ~ Lausanne ~ Switzerland,* [2]*University of Waterloo ~ Waterloo ~Canada*

This work is motivated by two major threats to valid causal inference: unmeasured confounding and ill-defined interventions. I will present new results on average causal effects in settings with unmeasured exposure-outcome confounding.
I will consider a class of estimands, frequently of interest in medicine and public health, that is currently not targeted by standard approaches for average causal inference. These estimands are queries about the average causal effect of a so-called intervening variable. As a practical motivation, I will consider chronic pain and opioid prescription patterns in the USA, and I will illustrate how conventional approaches will lead to unreplicable estimates with ambiguous policy implications. Then I argue that effects of intervening variables are replicable and have clear policy implications, and furthermore are non-parametrically identified by the classical frontdoor formula. As an independent contribution, I will present results on a new semiparametric efficient estimator of the frontdoor formula with a uniform sample boundedness guarantee. This property is unique among previously-described estimators in its class, and the estimator has superior performance in finite-sample settings. Finally, the theoretical results are applied to data from the National Health and Nutrition Examination Survey.
The new identification results justify the use of the frontdoor formula in new settings, reflecting questions of practical interest, where unmeasured confounding is a serious concern. These results do not rely on ill-defined interventions or cross-world assumptions. Furthermore, the new semiparametric estimator can be applied whenever the frontdoor formula identifies the parameter of interest, which e.g., could be an average causal effect, a population indirect effect or the new intervening variable estimand. The results also motivate future methodological work. In particular, generalizing the methods to longitudinal settings, involving time-varying treatments, is a major practical interest. Wen L, Sarvet AL, Stensrud MJ, Causal effects of intervening variables in settings with unmeasured confounding, 2023, Forthcoming

WO10.2 Methodology for systematic identification and analysis of multiple biases in causal inference

Wijesuriya R.*[1], Carlin J.B.[2], Peters R.L.[2], Moreno--Betancur M.[2]
[1]*Murdoch Children's Research Institute ~ Melbourne ~ Australia,* [2]*Murdoch Children's Research Institute and University of Melbourne ~ Melbourne ~ Australia*

Observational studies examining causal effects rely on unverifiable assumptions which, if violated, can imply multiple biases due to confounding, measurement and selection processes. Quantitative bias analysis (QBA) aims to examine the sensitivity of findings to assumption violations, generally by producing bias-adjusted estimates that would arise under alternative assumptions. Most QBA in practice addresses either a single source of bias or multiple sources separately, which may not accurately reflect the overall impact of the potential biases. In this work we aimed to address gaps in approaches for systematically identifying multiple biases and in guidance regarding the most appropriate way to analyse these together. First, we propose the "target trial" approach as a tool to help systematically list all potential biases in a study. This approach defines the causal effect of interest by specifying the protocol components of the ideal hypothetical trial, and then specifies how each of these components is emulated with the observational data. The emulation of each component is perfect only under certain assumptions, so considering the potential violation of each allows systematic identification of all biases. Second, we conducted a simulation study examining three possible approaches for analysing multiple biases: (i) individually analysing each source of bias, producing multiple bias-adjusted estimates, (ii) sequentially adjusting for biases in the assumed order in which they occurred, (iii) simultaneous adjustment for multiple biases. Throughout we consider both weighting and imputation bias-adjustment methods. Approaches are assessed in terms of how well they recover the true causal effect across different causal structures, varying bias strengths and misspecification of bias parameters. Simulations are based on a case study from a population-based investigation of food allergy in children, which is also used to illustrate the target trial approach and bias analysis approaches. We speculate, based on preliminary findings, that a simultaneous adjustment approach is the most appropriate, allowing assessment of overall impact whilst not relying on bias ordering assumptions. The target trial specification allows systematic identification of potential biases thus facilitating high-quality multiple bias analysis in causal inference, which may be best performed using a simultaneous adjustment approach.

[1] Brendel, Paul, Aracelis Torres, and Onyebuchi A. Arah. "Simultaneous adjustment of uncontrolled confounding, selection bias and misclassification in multiple-bias modelling." International Journal of Epidemiology (2023)
[2] Hernán, Miguel A., et al. "Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses." Journal of clinical epidemiology 79 (2016): 70-75.

Parallel Sessions | Wednesday 30 August 2023

WO10.3 Instrumental variable analysis with categorical treatment and ordinal instrument

Kazemi A.*, Olsen I.C.
Department of Research Support for Clinical Trials, Oslo University Hospital ~ Oslo ~ Norway

In the past two decades, the treatment of inflammatory arthritis diseases has been revolutionized by the introduction of Tumor Necrosis Factor Inhibitors (TNFis). The Norwegian Drug Procurement Cooperation (NDPC) manages a tender system where the health regions collaborate to procure favourable agreements to purchase TNFis. Each year, pharmaceutical companies are invited to offer prices for their TNFi, and the treatment with the lowest cost is selected as first line treatment for all health regions. Clinicians are expected to adhere to the cooperation's recommendations unless there are strong medical reasons not to. Therefore, price is the primary determinant of the preferred drug, based on the common, but largely untested assumption that TNFis are equally effective on a group level. Our objective is to employ the tender system results as a strong Instrumental Variable (IV) for head-to-head comparisons of the efficacy of five TNFis of interest. Current IV methodology requires assuming either homogeneity or monotonicity to derive causal point-estimators for the difference in treatment effects between two treatments. Homogeneity is a strong and often implausible assumption in many applications. Additionally, without limiting the number of principal strata, homogeneity is no longer sufficient on its own to enable the derivation of unbiased causal point-estimators for head-to-head comparisons when more than two relevant treatment alternatives are present. Generalizing monotonicity to a set of assumptions for more than two treatments, which are both plausible and enable effect estimation, is not trivial. By taking advantage of the instrument's ordinality, we propose a method for deriving causal point-estimators of the efficacy differences between several treatment pairs given plausible and well-defined rationality assumptions that limit the number of principal strata. We demonstrate that our methodology provides asymptotically unbiased estimates in the presence of unobserved confounding effects in a simulation study. We then apply the method to compare the effectiveness of the five TNFis of interest in the treatment of Rheumatoid Arthritis. The developed methodology provides an important addition to the toolbox for causal inference when comparing more than two interventions influenced by an IV.

J. Angrist and G. Imbens, "Identification and estimation of local average treatment effects", Econometrica, vol. 62, no. 2, 1994, pp. 467–75
S. A. Swanson, J. M. Robins, M. Miller, et al., "Selecting on treatment: A pervasive form of bias in instrumental variable analyses," American journal of epidemiology, vol. 181, no. 3, pp. 191– 197, 2015.
J. J. Heckman and R. Pinto, "Unordered monotonicity," Econometrica, vol. 86, no. 1, pp. 1–35, 2018.
T. K. Kvien, M. Heiberg, E. Lie, et al., "A norwegian dmard register: Prescriptions of dmards and biological agents to patients with inflammatory rheumatic diseases", Clinical and experimental
rheumatology, vol. 23, no. 5, S188, 2005.

WO10.4 Just what the doctor ordered: an evaluation of provider preference-based instrumental variable methods

Güdemann L.*, Dennis J., Shields B., Bowden J.
University of Exeter ~ Exeter ~ United Kingdom

The Instrumental Variable (IV) approach provides a possibility to address bias due to unmeasured confounding when estimating treatment effects with observational data. Healthcare provider prescription preference (PP) has been proposed as an IV. [1] As PP is hard to measure, IV analysis often relies on a surrogate measure Z constructed from data at hand. Different construction methods for Z are possible and their suitability rely on aspects such as data availability within provider, missing data in measured confounders and accounting for possible change in PP over time. We propose an extended version of the method by Ertefaie et al. [2], which aims to account for change in PP as well as maintaining its original feature of addressing nonignorable missingness. We evaluate this method together with already established approaches in a comprehensive simulation study including scenarios on the mentioned data conditions. Additionally, the methods are applied in a case study analysing the relative glucose lowering effect of two Type 2 Diabetes treatments in primary care data.
The proposed construction method utilized a mixed effect model with a random intercept for provider ID and a random slope for prescription time, to address between provider differences and change in PP. The method is applied as a two-step approach using different subsets of the available data to construct Z and estimate the treatment effect with the Two Stage Least Square approach. [1] Simulation results show that the method performs best in case of larger provider sizes, but is able to account for change in PP and nonignorable missingness. Results of the treatment estimation were slightly more efficient than its original version. Additionally, most of the established methods struggled with nonignorable missingness and some were biased in case of changing PP.
This state of the art evaluation showed that methods utilizing complex models are best applied with sufficiently large provider sizes. Only the method by Ertefaie et al. [2] and our extension method were able to address nonignorable missingness. It is valuable to triangulate results from different methods and discuss inconsistencies in order to identify sources of bias.

[1] E. L. Korn und S. Baumrind, „Clinician Preferences and the Estimation of Causal Treatment Differences," Statistical Science, Bd. 13, Nr. 3, pp. 209-235, 1998.
[2] A. Ertefaie, J. H. Flory, S. Hennessy und D. S. Small, „Instrumental variable methods for continuous outcomes that accommodate nonignorable missing

WO10.5 Practical considerations of using negative control exposures to detect residual confounding

Daniel N.*[1], Weinstein B.[1], Cohen K.[1], Ben--Menachem T.[2], Raz R.[2]
[1]Tel Aviv University ~ Tel Aviv ~ Israel, [2]The Hebrew University of Jerusalem ~ Jerusalem ~ Israel

The no unmeasured confounding assumption is the most highlighted and discussed assumption in observational studies. Researchers often struggle to justify the validity of this assumption. To this end, researchers increasingly utilize Negative Control Exposures (NCEs) to detect residual confounding. NCEs are variables known to have no effect on the outcome, for example, due to temporal ordering. If there is no unmeasured confounding, the NCE should be conditionally independent of the outcome. While NCEs are increasingly used in practice, key issues on how to choose and how to correctly use NCEs to detect residual bias are missing from the discussion.
We focus on two main issues: First, we show that the correlation between the exposure of interest and the NCE does not suffice to characterize how well the NCE will detect confounding. For example, even a seemingly valid NCE could have minimal or no power to detect residual bias, because conditioning on the exposure closes one non-causal path between the NCE and the outcome but may open another. Second, we argue that current practice of using hypothesis testing for detecting bias is problematic because the type I error, the error of announcing that that there is residual bias when in practice there is no confounding, is the wrong error to control. Therefore, we develop a two one-sided tests approach for bias detection, controlling for the appropriate type I error of announcing there is no residual bias when there is bias. We build on this approach to develop a sensitivity analysis. Focusing on linear and logistic regression models, we demonstrate both issues in simulations and data analysis in environmental epidemiology studies of the effect of air pollution exposures during pregnancy on outcomes at birth. The NCEs in these studies are typically air pollutants measured well after birth. Negative controls can be a powerful tool to detect unmeasured

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO11: MACHINE LEARNING 3

WO11.1 (Co-)clustering models for spatial transcriptomics

Andrea S.*, Davide R.
University of Padova ~ Padova ~ Italy

Spatial transcriptomics is a cutting-edge technology that, differently from traditional transcriptomic methods, allows researchers to recover the spatial organization of cells within tissues and to map where genes are expressed in space. By examining previously hidden spatial patterns of gene expression, researchers can identify distinct cell types and study the interactions between cells in different tissue regions, leading to a deeper understanding of several key biological mechanisms, such as cell-cell communication or tumour- microenvironment interaction. During this talk, we will be presenting novel statistical tools that exploit the previously unavailable spatial information in transcriptomics to coherently group cells and genes. First, we will introduce SpaRTaCo, a new model that clusters the gene expression profiles according to a partition of the tissue. This is accomplished by performing a co-clustering, that is, a simultaneous clustering of the genes using their expression across the tissue, and of the image areas using the gene expression in the locations where the RNA is collected. Then, we will show how to use SpaRTaCo when a previous annotation of the cell types is available, incorporating biological knowledge into the statistical analysis. Last, we will discuss a new modelling solution that exploits recent advances in sparse Bayesian estimation of covariance matrices to reconstruct the spatial covariance of the data in a sparse and flexible manner, significantly reducing the computational cost of the model estimation. Results obtained on human brain e prostate tissue samples reveal that the proposed models can detect specific patterns that appear only in some restricted tissue areas, confirming the importance of performing image clustering for revealing biological mechanisms. They also determine those genes that carry out specific biological functions in the discovered areas.

Sottosanti, A., and Risso, D. (2022). Co-clustering of Spatially Resolved Transcriptomic Data. The Annals of Applied Statistics, in press.
Kidd, B., and Katzfuss, M. (2022). Bayesian Nonstationary and Nonparametric Covariance Estimation for Large Spatial Data (with Discussion). Bayesian Analysis, 17(1), 291-351. Maynard, K. R., Collado-Torres, L., Weber, L. M., Uytingco, C., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nature neuroscience, 24(3), 425-436.

WO11.2 Bayesian rank-based clustering via mallows mixtures with covariates for cancer subtyping

Eliseussen E.*, Frigessi A., Vitelli V.
Oslo Centre for Biostatistics and Epidemiology, University of Oslo ~ Oslo ~ Norway

Rankings can be used to estimate individual behaviors in several areas, such as marketing and politics. Notably, rank-based models have recently been proposed as a useful tool for -omics analyses, as making use of data rankings instead of the actual continuous measurements increases the robustness of conclusions when compared to classical statistical methods (Vitelli et al., 2022). Furthermore, combining -omics information (rankings) with clinical information (covariates) can potentially lead to a better understanding of the patient's genetic and epigenetic makeup. The Mallows model is one of the most popular rank-based models, as it is flexible and it easily adapts to different types of preference data, and the previously proposed Bayesian Mallows Model (BMM) offers a computationally efficient framework for Bayesian inference also allowing capturing the heterogeneity across patients, via a finite mixture. However, BMM currently does not allow including covariate information on the patients. The main objective is to extend BMM to be able to utilize both rankings and covariate information in a joint modeling framework.
We develop a Bayesian Mallows-based finite mixture model that performs clustering while also accounting for patient-related clinical covariates. The proposed method is based on a similarity function that a priori favours the aggregation of people into a cluster when their covariates are similar, in line with the Product Partition models (PPMx) (Muller, 2011) proposal. We present two approaches to measure the covariate similarity: one based on an augmented model as in PPMx, and one based on a novel deterministic function measuring the covariates' goodness-of- fit to the cluster. We investigate the performance of the method, and compare it to alternative approaches, in both simulation studies and real-data examples.

Peter Muller, Fernando Quintana, and Gary L. Rosner. A Product Partition Model With Regression on Covariates. Journal of Computational and Graphical Statistics, 20(1):260–278, 2011. Valeria Vitelli, Thomas Fleischer, Jørgen Ankill, Elja Arjas, Arnoldo Frigessi, Vessela N. Kristensen, and Manuela Zucknick. Transcriptomic pan-cancer analysis using rank-based Bayesian inference. Molecular Oncology, 2022.

WO11.3 A clustering approach to multiple time-to-event data and application to multimorbidity associated with stroke

Delord M.*, Malek R., Learoyd A., Douiri A.
King's College London ~ London ~ United Kingdom

Multimorbidity refers to the co-occurrence of two or more chronic conditions in an individual. This state of health becomes increasingly prevalent and represent a public health challenge requiring novel analytic approaches. We present here a novel empirical approach to multiple time-to-event data, denoted as multiple state analysis (MSA) aiming at creating typologies of multimorbidity patterns based on patients' health record history.6 MSA is conducted according the following steps (1):
i) arrange patients individual records into multiple state indicators stacked in individual patients'state matrices and censoring indicators,
ii) compute pairwise patients' dissimilarities on individual state matrices (and censoringindicators) using the composite Jackard index and apply a clustering method,
iii) define a typology using partition quality indicators in the range of workable possibilities,
iv) display sequences characterising each cluster using a longitudinal graph representation andinterpret clusters in terms of the socio-demographic covariate and stroke risk-factors (SRF).
MSA was applied to routinely collected health records of 28 common acute and long-term conditions including stroke in patients aged over 18 and registered in 41 general practices in London between April 2005 and April 2021. Of 856,342 registered patients, 9,629 (1.1%) had a record of stroke. 8 clusters was identified of which 4 (67%) were characterised by at least two traditional SRF including hypertension (75.4%), diabetes (36.6%) and older age (median: 71.2 years). Osteoarthritis (29,8%) was the major non-traditional SRF in this part of the typology. In other clusters, hypertension and traditional SRF were less prevalent and sequences of non- traditional SRF, including mental and neurological conditions, asthma and multimorbidity emerged, in a younger population (median: 42 years) where female patients and White ethnicity were over-represented. Our results show how MSA discriminates between the main patterns of multimorbidity in terms of long lasting sequences of co-morbid conditions such as hypertension, diabetes or asthma, leading to stroke, in clusters of patients characterised by different ages, sex ratios, and ethnic compositions. Importantly, the propose application shows also how MSA allows to characterise non-traditional, but robust and recurrent sequences of risk-factors associated with common conditions (such as stroke), in atypical clusters in terms of traditional socio-demographic characteristics.

Delord, Marc, et al. "Multiple State Analysis, a Multidimentional Approach to Multiple Time-to- Event Data and Life Course Health Trajectories: Application to Patients with Myocardial Infarction." arXiv preprint arXiv:2209.11084 (2022).

Parallel Sessions | Wednesday 30 August 2023

WO11.4 Identification of novel dilated cardiomyopathy sub-phenotypes: unsupervised clustering for mixed-data type

Gandin I.*[1], Baj G.[2], Paldino A.[3], Zaffalon D.[3], Merlo M.[3], Dal Ferro M.[3], Barbati G.[1]
[1]*Department of Medical Sciences, University of Trieste ~ Trieste ~ Italy,* [2]*Department of Mathematics and Geosciences, University of Trieste ~ Trieste ~ Italy,* [3]*Cardiovascular Department, Azienda Sanitaria Universitaria Giuliano Isontina ~ Trieste ~ Italy*

Dilated cardiomyopathy (DCM) is a disease of the heart muscle that is associated with around 20% of 5-year mortality, for which accurate prognostic risk models are still lacking. It has a strong genetic component (around 30% of the cases are associated to pathogenic mutations) but also high variability in phenotypic expression and progression [1]. The aim of the present study is to identify novel sub-phenotypes based on the information collected at the first cardiological visit (clinical, ECG, Holter, imaging) using unsupervised clustering methods.
The study included detailed information of a longitudinal cohort of 409 DCM patients. The analysis presented two major challenges: the large number of variables (p=102) and their mixed-data type. An innovative two-step approach was applied: 1) all available variables went through a dimensionality reduction using a recent extension of principal component analysis for mixed data [2]; 2) agglomerative hierarchical clustering was applied to the first 11 components, after the identification of k=2 as optimal number of clusters (average silhouette width criteria). Using a post-hoc cluster representation technique, for the smaller group (Gr2, n=75) we obtained a clear clinical characterization for which Gr2 included mostly thickened heart muscles. Moreover, using a multivariate cause-specific Cox model, incorporating 9 well-known risk factors, we demonstrate that Gr2 is at lower risk for life-threatening arrhythmic events (HR=0.21, 95% CI [0.08,0.54]). Concordant results were obtained considering genetic information: Gr2 showed a lower yield of causative mutations compared to Gr1 (15% vs 47%, p<0.001) and none of those pathogenic variants were found in what are considered most arrythmogenic DCM- genes (DSP, PKP2, LMNA) [1]. Using a combination of dimensionality reduction and unsupervised clustering approach suitable for mixed-type data, two novel DCM subphenotypes were identified and characterized. The two groups differ in terms of progression of the disease, as well as in the genetic etiology. This findings could be useful for a future refining of the conventional phenotype-based classification of cardiomyopathies.

[1] A Paldino, M Dal Ferro, D Stolfo, I Gandin, K Medo et al. Prognostic Prediction of Genotype vs Phenotype in Genetic Cardiomyopathies. J Am Coll Cardiol. 2022 Nov, 80 (21) 1981–1994
[2] Marie Chavent, Vanessa Kuentz-Simonet, Amaury Labenne, Jérôme Saracco. Multivariate Analysis of Mixed Data: The R Package PCAmixdata. PREPRINT (v5) available at arXiv. 1411.4911

PARALLEL SESSION WO12: SYNTETIC DATA 1

WO12.1 From clinical trial simulations to in-silico trials

Friede T.*
University Medical Center Goettingen ~ Goettingen ~ Germany

The efficient conduct of clinical research is key for several reasons including resource constraints and ethical considerations, in particular in small populations and rare diseases [1].
Investigators make increasingly use of adaptive designs to increase the efficiency of their studies [2]. Recent work in this area also investigates the use of efficient optimization methods to conduct the simulation studies in a timely and resource economic manner [5]. Beyond the planning of experiments, simulations are used to assess statistical methods comparatively. Although this approach is very common in statistical science, in other areas such as computer science benchmarking based on established databases are more frequently used. Simulations and benchmarking will be contrasted and recommendations on their potentially combined used will be discussed [3]. Based on modelling and simulation techniques in-silico trials produce digital evidence complementing or replacing clinical trials [4]. An overview of the methodology for and applications of in-silico clinical trials will be presented.
Simulations play a key role in the efficient conduct of clinical research delivering robust results.

[1] Friede T, Posch M, Zohar S, Alberti C, Benda N, Comets E, Day S, Dmitrienko A, Graf A, Gunhan B, Hee SW, Lentz F, Madan J, Miller F, Ondra T, Pearce M, Röver C, Toumazi A, Unkel S, Ursino M, Wassmer G, Stallard N (2018) Recent advances in methodology for clinical trials in small populations: the InSPiRe project. Orphanet Journal of Rare Diseases 13: 186.
[2] Friede T, Stallard N, Parsons N (2020) Adaptive seamless clinical trials using early outcomes for treatment or subgroup selection: Methods, simulation model and their implementation in R. Biometrical Journal 62: 1264–1283.
[3] Friedrich S, Friede T (2023) On the role of benchmarking data sets and simulations in method comparison studies. Biometrical Journal (in press)
[4] Musuamba et al (2021) Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: Building model credibility. CPT: Pharmacometrics & Systems Pharmacology 10: 804–825.
[5] Richter J, Friede T, Rahnenführer J (2022) Improving Adaptive Seamless Designs through Bayesian optimization. Biometrical Journal 64: 948–963.

WO12.2 The challenges, feasibility and limits of statistical analysis on purely synthetic biomedical data

Carmisciano L.*[2], Montobbio N.[1], Signori A.[1], Sormani M.P.[1]
[1]*Università degli studi di Genova ~ Genova ~ Italy,* [2]*Università di Pisa ~ Pisa ~ Italy*

Synthetic data is an artificially generated set of observations with similar statistical properties and patterns of an existing real dataset. Privacy, data sharing and aid on statistical models training are among the most attractive applications of synthetic data in medicine. We aimed to replicate the results of SPI2, a published, Phase III, clinical trial evaluating the effect of biotin on progressive multiple sclerosis (MS) using synthetic data. To identify the most suitable strategy to answer three different clinical research questions we also compared the results from several synthetic data generation strategies. We used SPI2 as a data source. We applied five different methods for synthetic data generation (either based on additive random noise, uni and multivariate distribution modeling [1] or generative models [2]) to three different clinical research questions that were assessed using generalized linear models (as of today one of the most common tools in MS research). Results were averaged among 10 syntheses. All the synthetic data analyses, regardless of the generation strategy, concluded concordantly to the original SPI2 paper. For the primary SPI2 analysis the standardized mean difference of the treatment effect between the artificial and the real dataset analysis was lower than 9%. The mean confidence interval overlap over all coefficients between artificial and real dataset analysis was up to 85.0%*. It is feasible to use synthetic data generated from real clinical trials data to perform exploratory analyses using common statistical tools without the need of sharing the original data source. Not all the data generation procedures might capture the statistical properties that are relevant to a specific analytical method. Nevertheless, each of these approaches is expected to improve the future of data sharing.

[1] Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke Creation of Synthetic Data in R. Journal of Statistical Software, 74(11), 1–26. https://doi.org/10.18637/jss.v074.i11
[2.] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672–2680).
** not all results are yet available, some speculation are based on the literature and by results obtained on similar dataset.*

Parallel Sessions | Wednesday 30 August 2023

WO12.3 Data-generating models of longitudinal continuous outcomes and intercurrent events to evaluate estimands

Mitroiu M.*[1], Teerenstra S.[2], Oude Rengerink K.[4], Pétavy F.[3], Roes K.[2]
[1]Utrecht University ~ Utrecht ~ Netherlands, [2]Radboud University Medical Center ~ Nijmegen ~ Netherlands, [3]European Medicines Agency ~ Amsterdam ~ Netherlands, [4]Methodology Working Group, Medicines Evaluation Board ~ Utrecht ~ Netherlands

The ICH E9(R1) estimands addendum became public in December 2019, started to be adopted since, and is in the process of implementation for regulatory purposes in drug development and evaluation [1]. It provides a structured methodological framework for the planning, conducting, and interpreting randomised clinical trials for regulatory evaluation and approval. The main aim is to add clarity and a common understanding between all healthcare stakeholders, of the treatment effects targeted in clinical trials, using estimands. We aimed to develop and evaluate data-generating models to jointly simulate outcomes and intercurrent events for randomised clinical trials to enable the assessment of properties of estimands. We propose four data-generating models for the joint distribution of longitudinal continuous clinical outcomes and intercurrent events under the scenario where they are observable: a selection model, a pattern-mixture mixed model, a shared-parameter model and a joint model of longitudinally observed clinical outcomes and a survival model for intercurrent events. We present a case study in a short-term depression trial with repeated measurements of continuous outcomes and two types of intercurrent events, and compare the four proposed data-generating models. In our case study, we found that all four data-generating models can simulate different types of intercurrent events, their timing, and their associated longitudinal outcomes. These can be used to match envisaged patterns of intercurrent events and outcomes informed by prior available clinical trial data. For a given intercurrent event, the Shared-Parameter and Joint model tend to associate more similar longitudinal profiles (because of shared latent random effects), while the Selection Model and Pattern-Mixture model could allow more differences in associated profiles. We conclude that all four proposed data-generating models can be used to evaluate different estimands and to investigate their properties in-depth in the design stage. Thereby they are useful tools for the selection of estimands a priori.

[1] ICH E9(R1) EWG. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1) [Internet]. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

WO12.4 A simple yet effective approach for Synthetic clinical data generation with realistic marginal distributions

Farhadyar K.*[1], Bonofiglio F.[2], Zöller D.[1], Binder H.[1]
[1]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg ~ Freiburg im Breisgau ~ Germany, [2]National Research Council of Italy (CNR), Institute of Marine Sciences ~ La Spezia ~ Italy

Synthetic data generation is of great interest in diverse applications, such as privacy protection. Deep generative models, like variational autoencoders (VAEs) [1], are commonly used for creating synthetic datasets from original data. Despite the success of VAEs in image generation, there are limitations when it comes to clinical datasets, which are heterogeneous in terms of distribution and data type. In particular, VAEs have difficulty reconstructing the real distribution of variables when they deviate from the unimodal symmetric distributions, and that is because of the normality assumption typically used for the latent representations in VAEs. In these settings, it is essential to keep the architecture of VAE and the latent structure simple to avoid overfitting leading to disclosure risk and keep the possibility of pattern extraction and still be able to generate high-quality synthetic data. Therefore, we propose a novel method, pre-transformation variational autoencoders (PTVAEs), to specifically address bimodal and skewed data by employing pre-transformations at the level of original variables while preserving the simple, meaningful latent structure. Two types of transformations are used to bring the data close to a normal distribution by a separate parameter optimization for each variable in a dataset. We compare the performance of our method on a simulation design for an artificial, still realistic data set [2] and a real data example with other standard state-of-the-art methods (e.g., standard VAE and generative adversarial network) for synthetic data generation. In addition to the visual comparison, we use a utility measurement for a quantitative evaluation. At the same time, we investigate the latent structure of PTVAE to check for the patterns we expect from the datasets. The results show that the PTVAE approach can outperform others in both bimodal and skewed data generation. Furthermore, the simplicity of the approach makes it usable in combination with other extensions of VAE while we can still see the underlying patterns of used datasets. This method helps to generate high-quality synthetic data (in both utility and privacy aspects).

[1] Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
[2] Zöller, D., Wockner, L. and Binder, H., 2020. Modified ART study-Simulation design for an artifical but realistic human study dataset.

WO12.5 A simple-to-use r package for mimicking study data by simulations

Koliopanos G.[1], Ojeda F.[2], Ziegler A.*[1]
[1]Cardio-Care, Medizincampus Davos ~ Davos ~ Switzerland, [2]Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf ~ Hamburg ~ Germany

Data protection policies might prohibit the transfer of existing study data to interested research groups. To overcome legal restrictions, simulated data can be transferred which mimic the structure but are different from the existing study data. The aim of this work is to introduce the simple-to-use R package modgo which may be used for simulating data from existing study data for continuous, ordinal categorical, and dichotomous variables. The core is to combine rank inverse normal transformation with the calculation of a correlation matrix for all variables. Data can then be simulated from a multivariate normal and transferred back to the original scale of the variables. Unique features of modgo are that it allows to change the correlation between variables, to perform perturbation analysis, to handle multicenter data, and to change inclusion/exclusion criteria of a study by selecting specific values of one or a set of variables. Simulation studies on real data demonstrate the validity and flexibility of modgo. modgo mimicked the structure of the original study data. Results of modgo were similar with those from two other existing packages in standard simulation scenarios. modgo's flexibility was demonstrated on several expansions. The R package modgo is useful when existing study data may not be shared due to data protection issues. Its perturbation expansion permits to simulate truly anonymized subjects. The expansion to multicenter studies can be used for validating prediction models. Additional expansions can support the unraveling of associations in study data and can be useful in power calculations. G. Koliopanos, F. Ojeda, A. Ziegler 2023 A simple-to-use R package for mimicking study data by simulations. Methods Inf Med, in press

PARALLEL SESSION WO13: CLINICAL TRIALS 9

WO13.1 Basket trial designs based on power priors that incorporate overall heterogeneity

Baumann L.*, Kieser M.
Institute of Medical Biometry, University of Heidelberg ~ Heidelberg ~ Germany

Basket trials are used when a new treatment is tested simultaneously in several subgroups. They are usually uncontrolled phase II studies that investigate a binary endpoint. Most of the recently proposed designs for such trials utilize Bayesian tools to partly share the information between baskets depending on the similarity in order to increase the power compared to a separate analysis of each subgroup. A promising and computationally cheap design was proposed by Fujikawa et al. [1], where the subgroups are at first analyzed individually using a beta-binomial model. Information is then shared by calculating a weighted sum of the beta posterior parameters of the subgroups, where the weights are based on a similarity measure that is computed for the pairwise comparison of all individual posterior distributions. We aim to extend Fujikawa's design and thereby improve its operating characteristics. We show that the design by Fujikawa et al. is closely related to the approach of power priors, specifically to methods using empirical Bayes techniques. Power priors were originally proposed to incorporate historical data. For this application Gravestock & Held [2] showed that when information is borrowed from several historical studies, calculating weights based only on pairwise similarity with the current study is not optimal. Instead, they proposed an approach that incorporates all of the historical studies at once. We adapt the approach by Gravestock & Held for basket trials and consider other ways to extend Fujikawa's design by incorporating the overall heterogeneity between the different baskets in the calculation of the weights. In a simulation study, we compare the designs to other competing basket trial designs. The adaptations improve the operating characteristics of Fujikawa's designs in scenarios where only some of the baskets are active. The performance of the resulting designs is comparable to other basket trial designs. Fujikawa's basket trial design can be improved by considering the overall hetereogeneity between baskets when the weights that determine the amount of information sharing are computed. The extended designs are computationally cheap compared to other basket trial designs while having comparable operating characteristics.

[1] Fujikawa, K., Teramukai, S., Yokota, I., Daimon, T., Biometrical Journal, 62(2), 2020, 330- 338.
[2] Gravestock, I., Held, L., Biometrical Journal, 61(5), 2019, 1201-1218.

Parallel Sessions | Wednesday 30 August 2023

WO13.2 Application of constrained optimization techniques to bayesian basket trial designs

Sauer L.D.*[1], Ritz A.[2], Baumann L.[1], Freitag M.[3], Rauch G.[4], Kieser M.[1]
[1]*Institute of Medical Biometry, Heidelberg University ~ Heidelberg ~ Germany,* [2]*Clausthal University of Technology ~ Clausthal ~ Germany,* [3]*Institute of Biometry and Clinical Epidemiology, Charité - Berlin University of Medicine ~ Berlin ~ Germany,* [4]*Technical University of Berlin ~ Berlin ~ Germany*

In oncological research, rare tumor types are increasingly classified by common biomarkers, which encourages the development of new therapies targeted at these markers, independent of specific tumor sites. Phase II studies for such drugs are challenging: While separate studies for each site are often not feasible due to the low patient number, pooling all sites would disregard the site-specific heterogeneity. Basket trial designs address these issues: These trials recruit patients with all tumor sites of interest while grouping patients with specific tumor site as "baskets". In the analysis, baskets exhibiting similar response rates "borrow" more information from one another, whereas baskets exhibiting less similarity borrow less information or none at all. Recently, various frequentist and Bayesian methods for borrowing between baskets have been suggested. Many of them can be fine-tuned by using specific parameter values. The optimal parameter choice for maximizing power (or another utility function) while respecting a maximal type-I error rate (or another) constraint remains unclear up to now. We will formulate this question as a constrained optimization problem and tackle its solution using numerical optimization techniques. In order to investigate the feasibility of the constrained optimization framework, we will consider the Bayesian hierarchical basket trial design suggested by Berry et al. [1]. In this design, logit- transformed response rates are modeled on separate normal distributions for each basket, but the distributions' parameters are drawn from common prior distributions with identical hyperparameters. The drug is declared efficacious in a basket if the posterior probability of a sufficiently high response rate falls above a prespecified threshold. We will apply the simulated annealing algorithm [2] to find the optimal choice of hyperparameters and probability threshold. Berry et al. mention that their parameter choice is slightly over-powered, resulting in a moderate inflation of type-I error. With our approach, we strive to achieve an optimal balance between power and type-I error. Due to its broad applicability, numeric optimization is expected to be applicable to various basket trial designs including Bayesian hierarchical designs. This will facilitate informed planning of basket trials as well as objective comparisons between different designs.

[1] S.M. Berry, K.R. Broglio, S. Groshen, D.A. Berry, Clin Trials, 10 (5), 2013, 720-34. atrick, C.D. Gelatt Jr., M.P. Vecchi, Science, 220 (4598), 1983, 671-680.

WO13.3 Frequentist analysis of basket trials with one-sample mantel-haenszel procedures

Hattori S.*[1], Morita S.[2]
[1]*Osaka University ~ Osaka ~ Japan,* [2]*Kyoto University ~ Kyoto ~ Japan*

Recent substantial advances of molecular targeted oncology drug development are requiring new paradigms for early-phase clinical trial methodologies to enable us to evaluate efficacy of several subtypes simultaneously and efficiently. The concept of the basket trial is getting of much attention to realize this requirement borrowing information across subtypes, which are called baskets. Bayesian approach is a natural approach to this end and indeed the majority of the existing proposals relies on it. On the other hand, it required complicated modeling and may not necessarily control the type 1 error probabilities at the nominal level. In this research, we develop a purely frequentist approach for basket trials. We propose a framework to design and analyze basket trials based on one-sample Mantel- Haenszel procedure relying on a very simple idea for borrowing information under the common treatment effect assumption over baskets. We show that the proposed estimator is consistent under two limiting models of the large strata and sparse data limiting models (dually consistent) and propose dually consistent variance estimators. The proposed Mantel-Haenszel estimators are interpretable as the average treatment effect over baskets weighted by sample size even if the common treatment effect assumptions are violated. Then, we can design basket trials in a confirmatory matter. We also propose an information criterion approach to identify effective subclass of baskets. All the statistics are of closed-form expression. Simulation studies revealed that the proposed procedure was comparable to or outperformed existing Bayesian methods in terms of controlling type 1 error probabilities and identifying effective baskets. The proposed method gives a useful framework to design and analyze basket trials in a very simple frequentist matter.
Hattori S and Morita S (2023). Frequentist analysis of basket trials with one-sample Mantel- Haenszel procedures. arXiv:2302.08308 [stat.ME]

WO13.4 Non-concurrent controls in platform trials: separating randomised and non-randomised information

Schou M.*, Marschner I.
NHMRC Clinical Trials Centre, University of Sydney ~ Sydney ~ Australia

Flexibility to adapt randomised comparisons of multiple treatments by adding or dropping treatment arms under a single design framework is a key feature of platform trials. Each study stage may involve a different number of treatments. Reductions in sample size requirements and statistical efficiency improvements provided by a conventional multi-arm design with a common control also extend to a platform design in which, at each stage, multiple experimental treatments are compared with a common control. Additionally, the use of information contained in non-concurrent control subjects provides further efficiency gains in this design. A modelled approach which accounts for the treatment and stage has been proposed in the literature, facilitating the estimation of the effect of an experimental treatment versus a common control using information from subjects randomised concurrently and non-concurrently. This approach is based on a regression model adjusting for stage. We present research establishing an equivalence between this regression model approach and a mixed treatment comparison combining direct and indirect evidence. We decompose the total information into both direct and indirect evidence, presenting it as a weighted mean of the two components. The direct evidence will be based on an estimate of the evidence between the experimental treatment and the concurrent control, while the indirect evidence will be represented as a linear combination of the experimental treatment versus the control where the indirect evidence will be determined using other experimental treatments present in the platform as the common comparators. We first demonstrate this decomposition in a simple two-stage platform trial where the first stage is a standard two-arm design of treatment A versus a control. Experimental treatment B is introduced in the second stage. The decomposition of the total evidence for B versus the control will be presented as a weighted sum of the direct and indirect evidence (via A). We then extend this decomposition to more general designs. This approach is useful for understanding the contribution of direct and indirect evidence in a non-concurrent controls analysis. We recommend that platform trials incorporating non- concurrent controls report the decomposition of the overall treatment effect into its randomised and non-randomised components.

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO14: MISCELLANEA

WO14.1 Improve clinical and methodological research by adherence to reporting guidelines and structured reporting

Sauerbrei W.*, Kipruto E.
Medical Center - University of Freiburg ~ Freiburg ~ Germany

For many years, the quality of health science research has been heavily criticized. It is argued that significant improvement would be possible if research were better chosen, designed, executed, analyzed, regulated, managed, disseminated, and reported. Although, some of these issues may be difficult to address, there are appropriate guidance documents available to improve the reporting of research. Large weaknesses in this area are unnecessary and can be avoided. Concerning issues in reporting of health science, the EQUATOR (Enhancing the QUAlity and Transparency Of health Research, https://www.equator-network.org/) network acts as an umbrella organization. Although reporting guidelines for many types of observational studies were published over a decade ago, reviews of published studies have shown poor quality of reporting. This seriously biases the impression given by the published literature, and reduces the validity of results from systematic reviews and meta-analyses. Using the two-part REMARK profile (Sauerbrei et al 2018), a structured display summarizing key aspects of a study with an emphasize on all analyses, we will show that reporting of medical and methodological research can be improved. Selecting three articles each from five cancer journals, Sauerbrei et al (2022) created 15 REMARK profiles. Reporting of analyses was insufficient in nearly all studies and these profiles helped to identify severe weaknesses of analyses of many studies. The concept of structured reporting can be applied to various types of studies. For instance, we used simulated data to illustrate approaches that can help identify observations with an influence on function selection and the multivariable fractional polynomial (MFP) model. We investigated the effects of sample size and model replicability in MFP in eight subsamples, conducting 21 analyses. Using a structured display based on the REMARK profile, we distinguished between data description (D), analyses (A), and presentations (P), and provided a suitable illustration of all investigations. Additionally, we made recommendations for the sample size required to derive an MFP model that reflects the true underlying relationships.

The REMARK profile and adapted versions of structured reporting for other types of studies are suitable instruments to improve the reporting of various studies in clinical and methodological research.

Sauerbrei, W., Haeussler, T., Balmford, J., & Huebner, M. (2022). Structured reporting to improve transparency of analyses in prognostic marker studies. BMC medicine, 20(1), 1-19.
Sauerbrei, W., Taube, S. E., McShane, L. M., Cavenagh, M. M., & Altman, D. G. (2018).
Reporting recommendations for tumor marker prognostic studies (REMARK): an abridged explanation and elaboration. JNCI: Journal of the National Cancer Institute, 110(8), 803-811.

WO14.2 Confidence intervals using approximate propagation of imprecision

Ferguson J.*, Alvarez--Iglesias A.
HRB Clinical Research Facility Galway, University of Galway ~ Galway ~ Ireland

A common task is to calculate a confidence interval for a derived parameter that is a function, f, of K other parameters. For instance the population attributable fraction can be sometimes expressed as a function of relative risk and risk factor prevalence, both of which might be independently estimated. In these settings both the delta method and parametric bootstrap may be used to generate confidence intervals, but both have their drawbacks. For instance, delta method intervals require the derived estimator to be approximately normally distributed and bootstrap confidence intervals have the disadvantage that they are non-deterministic. Propagation of imprecision (propimp), as defined by [1], is an alternative method to construct confidence intervals that has advantages over the bootstrap in that it is deterministically calculated and over the delta method when the normality assumption is dubious. However [1] gave no general justification that propimp confidence intervals had correct asymptotic coverage, apart from in the trivial case of a difference in normally distributed means. Furthermore, if the derived parameter is a function of K other parameters the original algorithm requires a grid search over a K-1 dimensional hypersphere rendering it computationally infeasible for large K. Here, we show an asymptotic connection between propimp and delta method intervals, thereby justifying propimp in the settings where Newcombe recommended: f monotonic in all its arguments, and the K estimators that are 'plugged' into f are independent. In fact, this connection justifies the algorithm in more generality; for instance when the K estimators are correlated. To alleviate computational issues, we also present a new algorithm: approximate propimp (apropimp) that produces confidence intervals that closely replicate propimp intervals, without the grid search. We conclude by comparing propimp and apropimp intervals and associated coverage to the delta method and bootstrap for a range of differing estimands including Levin's PAF, the number needed to treat and the correlation coefficient. Apropimp is a viable way to construct confidence intervals, applicable in any scenario that the parametric delta method might be used, with better calibration in many settings. An R package to automate apropimp intervals will be available soon.

[1] Newcombe, Robert G. "Propagating imprecision: combining confidence intervals from independent sources." Communications in Statistics-Theory and Methods 40, no. 17 (2011): 3154-3180

WO14.3 Nonparametric bayesian analysis of survival data with spatially correlated cluster effects using soft-bart

Sinha D.*, Ghosh D.
Florida State University ~ Tallahassee ~ United States of America

We present a nonparametric Bayesian analysis of a study of racial disparity in breast cancer survival where the random cluster effects are spatially correlated, and some of the cluster-level covariates are estimated using a study different from the clustered survival study. Popular parametric and semi-parametric hazards regression models for clustered survival data are inappropriate and inadequate for this kind of situation with complex spatial effects on both clustering and covariates. In this article, we present a general nonparametric method for such clustered continuous survival responses under a paradigm of Bayesian ensemble learning called Soft Bayesian Additive Regression Trees or SBART in short. Our computationally feasible SBART method can incorporate unknown functional forms of the main effects, and interactions of various covariates, cluster specific effects and large number of clusters with variable cluster sizes. We illustrate the practical implementation of our method with an analysis of the effects of various intervenable and non-intervenable covariates of survival times of breast cancer patients from different counties (clusters) in Florida. For our analysis, the clustered survival data with patient-level covariates come from Florida Cancer Registry (FCR), and the data for one county-level intervenable covariate come from the Behavioral Risk Factor Surveillance Survey (BRFS). We compare our SBART based method with various existing analysis methods to demonstrate our advantage in assessing the impacts of intervention in some cluster/county level and patient- level covariates to eliminate racial disparity in breast-cancer survival in different Florida counties.

Basak, P, Linero, A, Sinha, D, Lipsitz, S. (2022) Semiparametric analysis of clustered interval- censored survival data using soft Bayesian additive regression trees (SBART). Biometrics, 78, 880– 893.

WO14.4 Survey sampling methods for partial verification bias in diagnostic evaluation studies

Thomas K.*[1], Meisner A.[2]
[1]University of Washington ~ Seattle ~ United States of America, [2]Fred Hutchinson Cancer Research Center ~ Seattle ~ United States of America

In studies of diagnostic accuracy of a test versus a given reference standard from a random sample of reference positive and reference negative patients, estimation of sensitivity and specificity is unbiased. However, partial verification bias occurs if selection of patients for reference testing is differential, for instance if likelihood of verification depends on an alternative, but related, reference test. Methods have been proposed to correct partial verification bias [1,2], usually when selection for verification depends on the result of the experimental test. Our objectives are to discuss design and analysis of diagnostic accuracy studies with multiple reference standards where one reference standard is not fully verified, and likelihood of verification depends on the result of a different reference standard. We consider standard survey sampling techniques to correct the bias by adjusting for the known sampling fractions. This approach, which has not been proposed in the literature for addressing verification bias in diagnostic test studies, does not involve regression modeling or propensity scores and is appealingly intuitive. We apply this to the evaluation of a test for SARS-CoV-2 against reference standards of polymerase chain reaction (PCR) and viral culture, where culture has scientific interest unique from PCR, but is more difficult to obtain. Among 257 participants given both the investigational test and the PCR reference standard, all 32 with available specimen from the 40 PCR-positive participants and a random sample of 20 of the 217 PCR- negative participants were tested by culture. Naïve analysis of the resulting 52 participants yielded sensitivity of 90.9% and specificity of 64.1% for the investigational test against culture. Recognizing that selection for verification by culture was based on PCR status, survey sampling techniques with inverse sampling weights of 40/32 and 217/20 for PCR-positive and -negative participants, respectively, were used and resulted in corrected sensitivity of 90.9% and specificity of 92.7% and appropriate standard errors. Comparisons to alternative approaches are discussed. In the case of partial verification bias from simple stratified sampling, naïve estimation of sensitivity and specificity will generally be biased. Standard survey sampling methods are a readily available solution in standard statistical software.

[1] C. Begg, R. Greenes. Biometrics, Vol 39, 1983, pp. 207-215. Assessment of Diagnostic Tests When Disease Verification is Subject to Selection Bias
[2] T. Alonzo, REVSTAT, Vol 12, 2014, pp. 67-83. Verification bias – impact and methods for correction when assessing accuracy of diagnostic tests

WO14.5 Optimal two-stage sampling for mean estimation in multilevel populations when cluster size is informative

Innocenti F.*, Candel M., Tan F., Van Breukelen G.
Maastricht University ~ Maastricht ~ Netherlands

In multilevel populations, there are two types of population means of an outcome variable: the average of all individual outcomes ignoring cluster membership, and the average of cluster- specific means.[1] To estimate the first mean, individuals can be sampled with simple random sampling or with two-stage sampling (TSS), that is, sampling clusters first, and then individuals within the sampled clusters. When cluster size varies in the population, three TSS schemes can be considered: sampling clusters with probability proportional to cluster size and then sampling the same number of individuals per cluster; sampling clusters with equal probability and then sampling the same percentage of individuals per cluster; and sampling clusters with equal probability and then sampling the same number of individuals per cluster. The aim of this research is to derive guidelines on how to design surveys to estimate the overall mean of a quantitative variable of interest in a multilevel population or to compare two multilevel populations in terms of their population means. Unbiased estimation of the average of all individual outcomes is discussed under each sampling scheme, allowing cluster size to be related to the outcome variable of interest (i.e. informative cluster size). [1] For each sampling scheme, optimal sample sizes are derived under a budget constraint.[2] The three optimal TSS designs are compared, in terms of efficiency, with each other and with simple random sampling of individuals. To overcome the dependency of the optimal sample sizes on some model parameters, maximin designs are derived.[2] A procedure for computing sample sizes for surveys to make cross-population comparisons is proposed and illustrated with the planning of a survey for comparing the average alcohol consumption among adolescents in France and Italy. Ignoring informative cluster size at the design stage is risky because it can lead to choosing a biased and inefficient sampling strategy, and underestimating the required research budget for the desired power level. If the cluster size distribution in the population is known, the best strategy to estimate the overall mean is to sample clusters with probability proportional to cluster size and then to take the unweighted average of cluster means.

[1] F. Innocenti, M. Candel, F. Tan, G. van Breukelen, Statistics in Medicine, 38, 2019, 1817– 1834.
[2] F. Innocenti, M. Candel, F. Tan, G. van Breukelen, Statistical Methods in Medical Research, 30, 2021, 357–375.

PARALLEL SESSION WO15: PRECISION MEDICINE 2

WO15.1 Optimizing information borrowing for bayesian hierarchical model in subgroup analysis

Lu X., Lee J.J.*
The University of Texas MD Anderson Cancer Center ~ Houston ~ United States of America

Subgroups occur naturally in a large variety of data sets and data analysis. For example, in basket trials we would like to estimate the efficacy of a targeted therapy in different cancer types or a immunotherapy in different subtypes of lung cancer. Bayesian hierarchical model (BHM) has been widely used in synthesizing information across subgroups. The typical assumption of exchangeability is very restricted and often does not hold. Efforts have been made in clustering the subgroups first, then, assuming exchangeability within cluster and borrowing information across subgroups within the same cluster. The two-step procedure has two main challenges: (1) How to determine the number of clusters? And (2) How much information to borrow within each cluster? To address these two interconnected challenges, we propose two distribution-free overlapping indices, namely, the overlapping clustering index for identifying the optimal clustering result and the overlapping borrowing index for assigning proper borrowing strength to clusters. Accordingly, we develop a new method BHMOI (Bayesian hierarchical model with overlapping indices). BHMOI includes a novel weighted K-Means clustering algorithm to obtain optimal clustering results, and an innate way to dynamically determining the borrowing strength in each cluster. BHMOI can achieve efficient and robust information borrowing with desirable properties. Examples and simulation studies are provided to demonstrate the effectiveness of BHMOI in heterogeneity identification and dynamic information borrowing. Chen, N. & Lee, J. J. (2019), 'Bayesian hierarchical classification and information sharing for clinical trials with subgroups and binary outcomes', Biometrical Journal 61(5), 1219– 1231. Chen, N. & Lee, J. J. (2020), 'Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints', Statistical Methods in Medical Research 29(9), 2717–2732. PMID: 32178585. Xu, G., Zhu, H. & Lee, J. J. (2020), 'Borrowing strength and

WO15.2 Prioritising the outcome in bayesian profile regression: an application to osteoarthritis proteomic data

Bondi L.*, Tom B., Richardson S.
MRC Biostatistics Unit, University of Cambridge ~ Cambridge ~ United Kingdom

There is a huge unmet need in osteoarthritis (OA) with an estimated 8.5 million people affected in the UK. It is regarded as a highly heterogeneous disease and is purported to exist in different forms. As part of the STEpUP OA collaboration, an academic-industry partnership, our work explores the molecular pathways of OA and aims at identifying subpopulations of patients homogeneous for protein marker profiles and such that each cluster has a clinical meaning (outcome-guided clustering). We carry out Bayesian profile regression (model-based outcome-guided clustering approach) to identify clusters of protein marker profiles that are associated with clinically relevant outcomes, such as disease radiographic grade (low vs advanced). This clustering methodology can handle possibly inter-related explanatory variables and uses the information in both these explanatory variables (i.e. ~ 6000 synovial protein markers) and the outcome to produce model-based clustering structures, where the uncertainty associated with these clustering structures and the number of clusters is reflected. Given the high dimensionality of the protein space, computational challenges arise when scaling profile regression in this context. The focus of this work is on strategies for dimensionality reduction and variable selection, taking account biological knowledge. Moreover, the influence of the clinical outcome to drive the clustering structure is investigated. Based on a two-step strategy that combines variable selection using the coefficient of variation of proteins and a latent embedding of the selected protein space, we found that the outcome had a limited influence compared to the embedding in uncovering the clusters. Possible reasons for this could be due to the imbalance towards advanced radiographic grade in the data and the higher dimensionality of the embedding protein space compared to the outcome space. We discuss a strategy based on generalized Bayes to increase the importance of the clinical outcome to achieve the aim of discovering clinically relevant clusters.
To achieve clinically relevant clusters of osteoarthritis patients using synovial proteins, one must adopt a flexible strategy based on both statistical and biological considerations.

1. Molitor J, Papathomas M, Jerrett M, Richardson S. "Bayesian profile regression with an application to the National survey of children's health". Biostatistics. 2010. 11(3). 484–498. 2. Rigon T, Herring AH, Dunson DB. "A generalized Bayes framework for probabilistic clustering". Biometrika. 2023. 3. Liverani S, Hastie DI, Azizi L, Papathomas M, & Richardson S. "PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes". Journal of Statistical Software. 2015. 64(7), 1–30.

WO15.3 Bayesian sequential design for identifiying and ranking of subgroups based on biomarkers in sepsis's patients

Vinnat V.*, Chevret S.
Université Paris Cité ~ Paris ~ France

Sepsis, characterized by a generalized inflammatory response associated with a severe infection, is the most common cause of admission to intensive care units. According to the WHO, out of the 30 million people who develop sepsis, 11 million die each year. The APPROCCHS study, a randomized phase III clinical trial, demonstrated an absolute reduction of 6% in sepsis mortality with a combination of hydrocortisone and fludrocortisone. Questions have arisen regarding the positive or negative impact of a number of biomarkers on the benefits of corticosteroids in sepsis. Additionally, simultaneous consideration of multiple biomarkers with potential impact on treatment effect could identify precise effective subgroups, towards personalized medicine. The RECORDS study is currently being conducted to determine if different immune status and/or corticosteroid biological activity signatures influence responses to the combination of hydrocortisone and fludrocortisone in adults with sepsis. Not all biomarkers will potentially be measured for each patient, which must be taken into account in the analysis. Our aim was to identify and classify patient subgroups that respond best to the experimental treatment in this context. We proposed a Bayesian sequential scheme to evaluate the therapeutic intervention as well as identify and classify subgroups. A zero-inflated truncated Poisson model was used to measure the primary outcome. We used posterior distributions of ranking and the surface under the cumulative ranking curve (SUCRA) proposed by Salanti et al. to obtain a final ranking of the different subgroups studied. Different subsets of subgroups were selected depending on the availability of information for each biomarker, and rankings were performed for each subset. Intermediate analyses including the possibility of early trial termination for efficacy were implemented. Scheme performance was evaluated through simulations under various scenarios.
The proposed scheme showed satisfactory results in terms of unbiased estimation. Furthermore, we found that in terms of subgroup identification and ranking, the design implemented presents robust and suitable performance.
In the years to come, personalized medicine will become increasingly present. It is therefore essential to propose sequential schemes that allow for subgroup identification while controlling decision errors at a reasonable threshold. In this context, our scheme offers a promising balance. Georgia Salanti, AE Ades, and John PA Ioannidis. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. Journal of clinical epidemiology, 64(2):163–171, 2011. Min-Hsiao Tsai and Ting Hsiang Lin. Modeling data with a truncated and inflated poisson distribution. Statistical Methods & Applications, 26(3):383–401, 2017

WO15.4 Statistical inference for roc curves after the box-cox transformation and use of the R package `rocbc'

Bantis L.[1], Brewer B.[1], Nakas C.*[2], Reiser B.[3]
[1]Dept. of Biostatistics and Data Science, University of Kansas Medical Center ~ Kansas City, KS 66160 ~ United States of America, [2]University of Thessaly ~ Volos ~ Greece, [3]University of Haifa ~ Haifa ~ Israel

Receiver Operating Characteristic (ROC) curve analysis is widely used in evaluating the effectiveness of a diagnostic test/biomarker or classifier score. Parametric approaches for statistical inference on ROC curves based on a Box–Cox transformation to normality have been discussed in the literature. Several investigators have highlighted the difficulty of considering the variability of the estimated transformation parameter when carrying out further analyses. This variability is usually ignored and inference is based on considering the estimated transformation parameter as fixed and known. We describe the problem and offer implementation options for accurate analyses. We briefly review the literature discussing the use of the Box-Cox transformation for ROC curves and the methodology for accounting for the estimation of the Box-Cox transformation parameter in the context of ROC analysis. We detail its application to a dataset of SARS2 antibody levels and illustrate via the R package that we have developed (named 'rocbc') which carries out all relevant analyses and is available on CRAN.
The rocbc package can be useful when applying the Box-Cox transformation in ROC analysis studies.

[1] L.E. Bantis, C. T. Nakas, B. Reiser, Biometrical Journal, 63, 2021, 1241-1253.
[2] C. Nakas, L. Bantis, C. Gatsonis, ROC analysis for classification and prediction in practice, CRC, 2023.

Parallel Sessions | Wednesday 30 August 2023

WO15.5 Treatment effect estimation for time-to-event outcomes in overlapping subgroups based on shrinkage methods

Wolbers M.*[1], Vázquez Rabuñal M.[2], Rufibach K.[1], Sabanés Bové D.[1]
[1]Product Development Data and Statistical Sciences, Hoffmann-La Roche Ltd ~ Basel ~ Switzerland, [2]Seminar for Statistics, Department of Mathematics, ETH ~ Zürich ~ Switzerland

In randomized controlled trials, forest plots are frequently used to investigate the homogeneity of treatment effect estimates in pre-defined subgroups based on clinical, laboratory, genetic, or other baseline variables. However, interpretation of naive subgroup-specific treatment effect estimates requires great care due to the smaller sample size of subgroups and the large number of investigated subgroups [1]. In the literature, Bayesian analyses have been proposed to address these issues, but they often focus on disjoint subgroups while subgroups in forest plots are overlapping [2]. We first build a flexible model based on all available observations, including categorical covariates that identify the relevant subgroups and their interactions with the treatment group variable. Interaction terms are then penalized using lasso or ridge regression to shrink subgroup-specific estimates towards the population treatment effect, or alternatively, a Bayesian shrinkage prior (the horseshoe prior) is applied [3]. One advantage of the Bayesian approach is the ability to derive credible intervals for subgroup-specific estimates. In a second step, this model is marginalized to obtain treatment effect estimates (hazard ratios) for all subgroups. The non-collapsibility of the hazard ratio complicates marginalization and leads to marginalized survival curves for the treatment groups that are not proportional. To address this issue, we use the average hazard ratio corresponding to the odds-of-concordance to quantify the treatment effect [4]. We illustrate these methods using data from a randomized clinical trial in follicular lymphoma and compare them in an extensive simulation study. In all simulation scenarios, the overall mean-squared error (MSE) of all methods significantly improved compared to naive subgroup- specific treatment effect estimates. The method based on the horseshoe prior performed slightly better in terms of bias, MSE, and frequentist coverage of 95% credible intervals, compared to the other methods, in scenarios where only one of the subgrouping variables is associated with treatment effect heterogeneity. We are implementing all these methods in an R package that we plan to upload to CRAN. Our method provides more precise treatment effect estimation in overlapping subgroups than the overall average treatment effect or naive subgroup-specific effects, respectively, and is a good compromise between these two extremes.

[1] Alosh, M., Huque, M. F., Bretz, F., & D'Agostino Sr, R. B. (2017). Tutorial on statistical considerations on subgroup analysis in confirmatory clinical trials. Statistics in Medicine, 36(8), 1334-1360.
[2] Jones, H. E., Ohlssen, D. I., Neuenschwander, B., Racine, A., & Branson, M. (2011). Bayesian models for subgroup analysis in clinical trials. Clinical Trials, 8(2), 129-143.
[3] Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics 11(2): 5018-5051.
[4] Rauch, G., Brannath, W., Brückner, M., & Kieser, M. (2018). The Average Hazard Ratio–A Good Effect Measure for Time-to-event Endpoints when the Proportional Hazard Assumption is Violated?. Methods of Information in Medicine, 57(03), 089-100.

Parallel Sessions | Wednesday 30 August 2023

PARALLEL SESSION WO16: LONGITUDINAL ANALYSIS 4

WO16.1 Extended joint models under the bayesian approach using jmbayes2

Rizopoulos D.*, Miranda Afonso P., Papageorgiou G.
Erasmus University Medical Center ~ Rotterdam ~ Netherlands

Joint models for longitudinal and time-to-event data have become a popular tool in follow-up studies. They are used to model endogenous time-varying covariates for survival outcomes and to account for nonrandom dropout in the longitudinal outcomes. Despite the number of development in this field, these models have not yet found their rightful place in the applied researchers' toolbox. This was mainly due to the unavailability of user-friendly and versatile statistical software. Building upon previous experience, we developed the R statistical package JMbayes2 to fill this gap. JMbayes2 is a robust R package for longitudinal and time-to-event that enables the user to:
- To include multiple longitudinal outcomes with different probability distributions;
- To accommodate different event time processes, such as competing risks, multi-state or recurrent events;
- To link the longitudinal outcomes to the risk of the events of interest using various functional forms, such as underlying value or slope; and
- To derive individualized dynamic predictions from the fitted joint models.

The package was developed under the Bayesian paradigm. The Markov chain Monte Carlo algorithms are implemented in C++, not relying on an external sampler. This characteristic enables model fitting in a timely fashion, despite its complexity. Despite model fitting functions, the package also includes several tools for summarizing and visualizing results as well as performing model diagnostics. The availability of an easy-to-use statistical tool such as JMbayes2 is likely to be a helpful solution to ease longitudinal and time-to-event data analyses by applied researchers in their everyday practice without sacrificing model adequacy. It thereby brings new insights into disease progression and contributes to better monitoring and treatment strategies.

Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Chapman & Hall/CRC, Boca Raton.

WO16.2 Bayesian inference for joint models of longitudinal and survival data with dynamic risk prediction

Rustand D.*, Van Niekerk J., Krainski E., Rue H.
King Abdullah University of Science and Technology ~ Thuwal ~ Saudi Arabia

The R package INLAjoint is designed to handle a variety of longitudinal models, including mixed effects, proportional odds and zero-inflated models, as well as survival models such as frailty, mixture cure, competing risks, and multi-state models. These models can be assembled to form complex joint models with shared or correlated random effects, providing a flexible and efficient method for analyzing multivariate longitudinal and survival data. The INLA algorithm performs Bayesian inference through deterministic approximation, it avoids the long computation times and convergence issues encountered when fitting the most complex joint models with standard iterative algorithms. Indeed, simulation studies demonstrate that INLAjoint substantially reduces computation time and the variability of parameter estimates compared to alternative strategies such as Markov chain Monte Carlo, Monte Carlo expectation maximization or Newton-like algorithms [1,2]. A key application of joint models is the dynamic prediction of the risk of an event, such as death or disease progression, based on changes in the longitudinal outcome(s) over time. INLAjoint allows for the estimation of dynamic risk predictions and can incorporate changes in the longitudinal outcome to update future risk predictions. Overall, INLAjoint offers a flexible and efficient method for modeling joint longitudinal and survival outcomes, handling a range of models, and enabling dynamic risk predictions. These features make it a valuable tool for analyzing complex health data and may help towards personalized decision in medicine.

[1] Rustand, D., van Niekerk, J., Rue, H., Tournigand, C., Rondeau, V., & Briollais, L.(2023). Bayesian estimation of two-part joint models for a longitudinal semicontinuous biomarker and a terminal event with INLA: Interests for cancer clinical trial evaluation. Biometrical Journal, 2100322.
[2] Rustand, D., van Niekerk, J., Krainski, E. T., Rue, H., & Proust-Lima, C. (2023). Fast and flexible inference approach for joint models of multivariate longitudinal and survival data using Integrated Nested Laplace Approximations. arXiv:2203.06256.

WO16.3 Joint analysis of disease progression markers and death using individual temporal recalibration

Saulnier T.*[1], Foubert--Samier A.[2], Proust--Lima C.[1]
[1]Univ. Bordeaux, Bordeaux Population Health Research Center, Inserm U1219 ~ Bordeaux ~ France, [2]CHU Bordeaux, Service de Neurologie des Maladies Neurodégénératives, IMNc, CRMR AMS, NS-Park/FCRIN Network ~ Bordeaux ~ France

Establishing the natural history of a disease permits to better understand its progression over time and identify therapeutic directions. However, in complex diseases such as neurodegenerative diseases, its study often faces multiple challenges. When the disease is difficult to diagnose, uncertainty remains around the time of disease onset. Patients are potentially recruited in cohorts at different disease stages, making time in study no longer meaningful. Occurrence of clinical events, such as death, also interrupts follow-ups, inducing missing data potentially not at random [2]. The present work introduces a joint model combining a disease progression model [1] based on an individual temporal recalibration to describe markers progression according to the latent disease time and a survival model to assess the association with death. The methodology is motivated by the study of Multiple system atrophy (MSA), a rare neurodegenerative disease. The markers' progressions are described according to disease time using nonlinear mixed-effect models. Disease time is defined according to disease severity at inclusion and an individual random shift. The shape of markers' progressions are either defined as sigmoids or determined from the data using fractional polynomials. The risk of death is jointly modeled according to the markers' dynamics and the individual temporal shift. Estimation, made available in the R- package LTSM, is carried out in the Maximum Likelihood Framework using Quasi-Monte-Carlo approximations and Marquardt-Levenberg optimizer. Annual data from 663 patients from the French MSA cohort were analyzed over 10.8 years. MSA progression was described by the Unified MSA Rating Scale sumscores I (functional sphere) and II (motor sphere). UMSARS-I and II progressions spanned over 12 years; compared to non-dependent patients at inclusion, mean time gaps between moderately- dependent and helpless patients at inclusion were estimated at 2.56 (95%CI=2.36,2.76) and 5.84 (95%CI=4.92,6.77) years, respectively. Risk of death highly depended on markers' dynamics features and individual disease time (with higher risk for more advanced patients). We developed a joint model for describing disease progression which tackles the potential heterogeneity of patients' profiles. Applied to MSA, this approach has potential in many complex diseases.

[1] Li D, Iddi S, Thompson WK, Donohue MC, Alzheimer's Disease Neuroimaging Initiative. Bayesian latent time joint mixed effect models for multicohort longitudinal data. Stat Methods Med Res 2019; 28: 835–45.
[2] Saulnier T, Philipps V, Meissner WG, et al. Joint models for the longitudinal analysis of scales in the presence of informative dropout. Methods 2022; 203: 142–51.

Parallel Sessions | Wednesday 30 August 2023

WO16.4 A lambert function-based procedure to fit joint models for multivariate longitudinal and time-to-event data

Charvat H.*
Faculty of International Liberal Arts, Juntendo University ~ Tokyo ~ Japan

Joint models are usually advocated to analyze the relationship between repeated measurements of possibly time-varying covariates, such as biomarkers, and the occurrence of a health event. However, computational issues sometimes limit the practical implementation of such models. In the present work, we developed an R function for joint model estimation based on an improved method for approximating the likelihood. The particular class of models considered here consists in the combination of a linear mixed- effects sub-model for the multivariate longitudinal outcome and a flexible parametric hazard regression sub-model for the time-to-event outcome. These sub-models are related through multivariate normally distributed shared random effects acting in a time-independent manner on the hazard (this corresponds to the so-called "random effects parameterization"). Under these specifications, the individual-specific marginal contributions to the likelihood involve a multidimensional integral that has no closed-form solution. Based on a recently published result [1], we show how the mode and curvature of the log-integrand can be expressed analytically in terms of the observed data and the model parameters thanks to the Lambert W function, greatly simplifying the procedure for approximating the integral via adaptive Gauss- Hermite quadrature (AGHQ). As this Lambert function-based approach leads to very simple closed-form expressions for higher-order derivatives of the log-integrand, we also describe the implementation of a second-order asymptotic approximation of the integral based on the Laplace method. Maximization of the resulting approximated model likelihood can then be obtained through standard Newton-type optimization routines. The performance of the approach is assessed through a small simulation study analyzing the impact of the number of individuals and the number of repeated measurements on the parameter estimation for joint models including multiple shared random effects, and using the Lambert function-based AGHQ or Laplace approximations. In an illustration based on real data from patients with primary biliary cirrhosis, the association between the trajectories of three biomarkers and mortality is assessed by including the corresponding patient-specific random intercepts and slopes as predictors in the time-to-event sub-model. The Lambert function-based integral approximations described here constitute an efficient computational approach for fitting a class of relatively simple but useful joint models.

[1] Charvat H. Using the Lambert function to estimate shared frailty models with a normally distributed random intercept. Am Stat 2023;77(1):41-50

WO16.5 A novel platform for analyzing semi-continuous medical cost and survival data

Shojaei Shahrokhabadi M.*[1], Mirkamali S.J.[2], Chen D.(.[3]
[1]Biostatistics Research Group, Population Health Sciences Institute, Newcastle University ~ Newcastle upon Tyne ~ United Kingdom, [2]Department of Mathematics, Faculty of Sciences, Arak University ~ Arak ~ Iran, Islamic Republic of, [3]College of Health Solutions, Arizona State University ~ Phoenix, AZ 85004 ~ United States of America

Marginalized Two-part Joint Models (MTJMs) were initially developed to address some challenges associated with medical costs including right skewness, clumping at zero, and censoring due to death or incomplete follow-up. The MTJMs are perhaps one the most considered approaches when the primary interest is to gain insight into the complex relationships between the average medical costs amongst the entire population of both users and non-users as one outcome and the occurrence of death as another outcome. The initial MTJM was constructed upon the log-normal (LN) distribution with a constant variance parameter for analyzing positive values (i.e., non-zero) costs. However, practical studies show that the log-normality assumption could be violated. This paper extends the classical MTJM by considering a generalized gamma (GG) distribution which contains LN distribution as a limiting case. A series of simulation studies are conducted to examine the finite-sample properties of the proposed MTJM model with respect to the bias of the parameter estimates, coverage of probability coverage, and estimation efficiency. The simulation results show that when the response distribution is unknown or misspecified, the GG-based MTJM provides a potentially more robust estimator than the classical LN-based MTJM. The advantage of the new MTJM is also demonstrated by analyzing electronic health records (EHRs) data collected in Iran. Shahrokhabadi, Mohadeseh Shojaei, et al. "Marginalized Two-Part Joint Modeling of Longitudinal Semi-Continuous Responses and Survival Data: With Application to Medical Costs." Mathematics 9.20 (2021): 2603. Smith, Valerie A., et al. "A marginalized two-part model for longitudinal semicontinuous data." Statistical methods in medical research 26.4 (2017): 1949-1968. Smith, Valerie A., and John S. Preisser. "A marginalized two-part model with heterogeneous variance for semicontinuous data." Statistical Methods in Medical Research 28.5 (2019): 1412- 1426.

PARALLEL SESSION WO17: HIGH DIMENSIONAL DATA 3

WO17.1 Co-clustering matrix tri-factorization: spatial and features constraints

Capitoli G.*[3], Denti F.[1], Galimberti S.[3], Risso D.[2], Sottosanti A.[2]
[1]Department of Statistics, Università Cattolica del Sacro Cuore ~ Milano ~ Italy, [2]Department of Statistical Sciences, University of Padua ~ Padova ~ Italy, [3]Department of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy

Over the last ten years, Matrix-Assisted Laser Desorption Ionisation (MALDI) Mass Spectrometry Imaging (MSI) has been one of the key technologies for cancer biomarker discovery directly in-situ [1]. This technology partitions a tissue sample into a grid of pixels, and a mass spectrum is acquired for each pixel. A generic spectrum conveys the mass-to-charge (m/z) values representing analytes of interest along the x-axis and the corresponding intensities along the y-axis. Finally, a three-dimensional MSI dataset is obtained by arranging the mass spectra and the grid of coordinates for each pixel to be investigated in further analyses. In this work, we investigate a co-clustering statistical model that partitions the molecules' spatial expression profiles, considering the spatial dependencies between neighbouring pixels. We perform co-clustering Non-negative Matrix Tri-Factorization (NMTF) [2], to infer the latent block structure of the data and to induce two types of clustering: 1) of the molecules, using their expression across the tissue, and 2) of the image areas, using the coordinates of the pixels. Our proposed methodology is validated with a series of simulation experiments investigating different types of constraints for Non-negative Matrix Factorization (NMF). In particular, we evaluate constraints both to the cluster indicator matrices and as penalized factors. We evaluate the ability of the methods to recover the subgroups (pixels) and the variables (molecules) that drive the clustering. Comprehending biomolecule functions and their interactions in different areas of the tissue is of great scientific interest, as it might lead to a deeper understanding of several key biological mechanisms. However, adequate statistical tools that exploit the new spatial mapping information to reach more specific conclusions are still lacking. In this work, we investigate a parsimonious computational method to segment the pixels of images obtained with the MALDI- MSI technique into biologically meaningful clusters employing the FNMT co-clustering method.

[1] Rohner, T.C., Staab, D., Stoeckli, M. Maldi mass spectrometric imaging of biological tissue sections. Mech Ageing Dev 126(1), 2005, 177–185
[2] Wang, H., Nie, F., Huang, H., & Makedon, F. Fast nonnegative matrix tri-factorization for large-scale data co-clustering. In Twenty-Second International Joint Conference on Artificial Intelligence. 2011

Parallel Sessions | Wednesday 30 August 2023

W017.2 Procrustes analysis for spatial transcriptomics data

Corbetta D.*[1], Andreella A.[2], Finos L.[1], Risso D.[1]
[1]University of Padua ~ Padua ~ Italy, [2]Ca' Foscari University of Venice ~ Venice ~ Italy

Spatial transcriptomics is a recent genome-sequencing technique that provides information on the spatial organization of tissues while simultaneously obtaining gene expression data. The application of this technology could revolutionize medical research by providing insights into the genomic basis of brain diseases. However, analysis of brains of different subjects is challenging because they are not functionally aligned. This could lead to biased results in the analysis of samples of different subjects under different biological or pathological conditions. Alignment of samples is indeed a preliminary and unavoidable step. In this project, we apply the ProMises alignment [1] to spatial transcriptomics data to demonstrate that this approach removes a major source of technical variability. Procrustes analysis is a statistical shape analysis that aligns matrices in a common reference space using similarity transformations. It has been rephrased as a statistical model in the perturbation model, which defines matrices as random rotations of a common reference matrix plus an error. However, its solution is not unique, lacking interpretability since the aligned images lose their anatomical structure. To address this issue, Andreella and Finos (2022), proposed the Efficient ProMises model, a Bayesian extension of the perturbation model that assumes a von Mises-Fisher distribution as the prior distribution for the rotation parameter. This model is suitable for high-dimensional data and provides a unique solution. In this study, we applied Procrustes alignment to spatial transcriptomics data obtained from the dorso-lateral-prefrontal cortex of three independent neurotypical donors [2]. The results demonstrated that ProMises transformation significantly reduced the technical variability caused by misalignment. Indeed, the analysis conducted on aligned samples revealed less false positives in differentially expressed genes, highlighting the need for alignment methods in spatial transcriptomics data analysis. ProMises alignment is a promising approach for aligning spatial transcriptomics data, which can remove technical unwanted variability and improve the accuracy of downstream analysis. Our findings demonstrate that Procrustes-alignment algorithms can be applied to spatial transcriptomics data from different donors and provide a more reliable and comprehensive view of gene expression in the brain. Further research is required to explore the potential of this approach in other settings.

[1] Andreella, A., Finos, L. Procrustes Analysis for High-Dimensional Data. Psychometrika 87, 1422–1438 (2022).
[2] Maynard, K.R., Collado-Torres, L., Weber, L.M. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci 24, 425–436 (2021)

W017.3 Statistical integration of multi-omics and drug screening data from cell lines

El Bouhaddani S.*[1], Uh H.[1], Houwing--Duistermaat J.[2]
[1]UMC Utrecht ~ Utrecht ~ Netherlands, [2]Radboud University ~ Nijmegen ~ Netherlands

Collecting data on multiple molecular levels has become a fundamental aspect of modern biomedical research. Examples are clinical cell line studies, where multiple omics datasets are measured, as well as interventional data such as high-throughput (drug) screens. A holistic "integrative" approach can provide new insights into the underlying pathology and molecular interactions. Our motivational example is a study of multiple system atrophy (MSA) and Parkinson's disease (PD), both devastating neurodegenerative disorders. To date, no disease- modifying treatment is available, and the molecular causes and consequences of the diseases remain unclear. The study aims to better understand the molecular basis of MSA and PD and find potential disease-modifying drugs. To this end, transcriptomics, proteomics and 1600 FDA- approved drug screening data are measured in human brain cell lines under a disease-inducing and control environment. Challenges include high-dimensionality, non-overlapping datasets, strong correlations between measurements, and heterogeneity across the multi-omics and screening datasets. For an integrative analysis of these data, appropriate statistical data integration approaches are needed. We propose a novel statistical data integration workflow to integrate multi-omics and drug screen data. For the first part of the workflow, we develop Probabilistic OPLS discriminant analysis, POPLS-DA, to model the multi-omics datasets in terms of joint, omics-specific, and residual components. These components consist of weighted linear combinations of genes and proteins. The outcome is included in the model via the joint components to obtain genes and proteins that best distinguish cases and controls across all omics data. All parameters are simultaneously estimated with maximum likelihood using a memory-efficient EM algorithm. Based on the top genes and proteins, a gene-gene interaction network is built using String-DB. For the second part, we propose a 'direct drug neighbor' approach: we use DrugBank to obtain all interactors of validated drugs from the screen and intersect these with our list of top genes and proteins. We study the performance of POPLS-DA with simulations and apply our approach to transcriptomics, proteomics and screening data measured in the cell lines. The obtained integrated interaction network will highlight druggable subnetworks that can potentially be used in a novel therapy for MSA and PD.

Bouhaddani, S. el, Höllerhage, M., Uh, H.-W., Bickle, M., Moebius, C., Höglinger, G. U., & Houwing-Duistermaat, J. J. (2022). Statistical omics data integration identifies potential therapeutic targets for synucleinopathies. BioRxiv, 2022.08.10.503452. https://doi.org/10.1101/2022.08.10.503452

W017.4 Analysis of compositional microbiome data with bias correction using poisson framework

Musisi C.*, Thas O., Kodalci L.
Hasselt University ~ Hasselt ~ Belgium

Microbiome refers to the totality of micriobial organisms that live in a particular environment (e.g. the human gut, a plant or river). The human microbiome is known to play an important role in the general body functioning hence the interest in microbime studies. DNA sequencing technologies are nowadays used for measuring the microbial composition in a biological sample. After preprocessing, the data take the form of counts: for each taxon in a sample, a count is assumed to be proportional to the true abundance of a taxon in the microbiome. Microbiome data show some particular characteristics: sparse (many zeroes), compositional (counts are only related to relative abundances), and overdispersed. The counts cannot be well described by a Poisson or negative binomial distribution. In many microbiome studies, the aim is to identify differentially abundant taxa between two conditions (e.g. healthy versus diseased subjects). Many methods have been described in the literature. ANCOM-BC is a method that performs well in terms of both false discovery rate (FDR) control and sensitivity. It makes use of a linear model for the log-transformed counts, and it is unique in the way it deals with systematic differences of the sampling fractions in two groups that have to be compared. The sampling fraction refers to the relative proportion of microorganism sampled from a fixed volume of the ecosystem. Differences in sampling fractions may cause bias in the estimation of effect sizes. We propose an alternative method that is inspired by ANCOM-BC, but which does not log- transform the counts and is therefore expected to suffer less from the sparseness of the data. Our method is based on a log-linear quasi-Poisson regression model. The parameter estimation method accounts for the overdispersion by nonparametrically estimating the mean-variance relationship. In this way we do not rely on a distributional assumption for the counts, and valid inference becomes possible. The new method is evaluated in a simulation study. In the simulation study, our method controls the FDR quite well.

Lin, Huang, and Shyamal Das Peddada. "Analysis of compositions of microbiomes with bias correction." Nature communications 11, no. 1 (2020): 3514.

WO17.5 Quantifying uncertainty in deep generative synthesis of tabular medical data with bayesian inference

Tippmann P.*, Farhadyar K., Zöller D.
Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg ~ Freiburg ~ Germany

Medical data is essential for medical research and patient care. However, data protection regulations frequently limit access to such data. Synthetic data generation is a promising solution to allow researchers to create synthetic medical data maintaining the statistical characteristics of the original data while ensuring confidentiality. We focus on Variational Autoencoders (VAEs) to synthesize tabular medical data, since they model the probability distribution of the data and provide a low-dimensional space for further analysis. One significant issue that arises when using deep generative models such as VAEs is accounting for the uncertainty in the data generation process. As other Deep Neural Networks, they can be overly confident when dealing with anomalous or Out-of-Distribution data. This can cause unreliable model predictions or inflated probability estimates during downstream analysis of synthetic data. To address this problem, we utilize Bayesian inference methods for uncertainty quantification. Previous work used Bayesian methods for uncertainty estimation in VAEs mainly with image data. Tabular medical data, on the other hand, present unique challenges such as treatment of heterogeneous data types and distributions, as well as missing or anomalous values that can introduce bias. Our approach to synthetic data generation involves developing a VAE framework specifically designed to address the above challenges. We then utilize two Bayesian inference methods, model averaging and Markov-Chain Monte Carlo techniques, to quantify epistemic (knowledge- related) model uncertainty. To evaluate the quality of the generated data, we use robust metrics and demonstrate that our approach preserves essential properties of the original data. Through real and simulation data, we illustrate how our proposed approach enhances the faithfulness of downstream task performance, such as classification and regression, by producing more precise and dependable synthetic data. In summary, our research focuses on using Bayesian inference to quantify uncertainty in the synthesis of tabular medical data using VAEs. We address the difficulties and factors to consider in generating synthetic tabular data. Using both real and simulated medical datasets, we demonstrate how our framework and approach enhance the usefulness of the models. Our work is a step towards the development of more trustworthy deep generative models for medical applications.

[1] M. Glazunov and A. Zarras, "Do Bayesian variational autoencoders know what they don't know?," in Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, PMLR, Aug. 2022, pp. 718–727. Accessed: Mar. 31, 2023. [Online]. Available: https://proceedings.mlr.press/v180/glazunov22a.html
[2] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," Information Fusion, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/j.inffus.2021.05.008.

Poster Presentations

MONDAY 28 AUGUST 2023

POSTER SESSIONS 1

MP1	Mutlu Altuntas	*On detecting outliers in a circular-linear regression by cephalometric angles data*
MP2	Koichi Hashizume	*A comparison of dose-finding method for incorporating single-agent historical data into combination trial*
MP3	Adnan Karaibrahimoglu	*On the comparison of classical and wrapped cauchy circular regression in patients with gastric carcinoma*
MP4	Muneyori Okita	*Estimation of between-study variance in bayesian meta-analysis: borrowing information from excluded studies*
MP5	Marie-Karelle RIVIERE	*Decision-making frameworks using multiple correlated endpoints*
MP6	Giulia Capitoli	*Bivariate copula for left and right censoring: validation of a complex surrogate for a time-to-event endpoint*
MP7	Susana Diaz Coto	*Reducing the overfitting in the groc curve estimation*
MP8	Ryo Emoto	*Enhancing the efficiency of predictive marker analyses by using treatment crossover data*
MP9	Kazuharu Harada	*Modeling and estimation for hierarchical ordinal outcome*
MP10	Emir Sehovic	*Meta-analysis of cell-free micrornas for early breast cancer detection: causes of inconsistent findings*
MP11	Nivetha Sridharan	*Estimating dce-mri ktrans repeatability by meta-analysis of data from 12 clinical studies*
MP12	Markus Waser	*The role of quantitative eeg (qeeg) in clinical trials: a methodological review*
MP13	Federica Bellerba	*Mediation analysis of obesity and adiponectin in breast cancer risk: a nested-cohort study in the ibisii trial*
MP14	Kate Ellis	*Staggered difference-in-differences for estimating effects of an online consultation system on prescribing*
MP15	Teresa Greco	*Causal effect analysis on spasticity-plus outcomes in patients with multiple sclerosis*
MP16	Maria Gueltzow	*The role of labor market inequalities in explaining the gender gap in depression risk among older us adults.*
MP17	Yasutaka Hasegawa	*Development of ensemble propensity score matching method*
MP18	Shaun Hiu	*Instrumental variables: systematic review of strategies to justify validity in clinical studies of dementia*
MP19	Simon Newsome	*Applying causal inference to clinical trial data to improve our understanding of how car-t cell therapies work*
MP20	Christian Pipper	*Unbiased and efficient estimation of causal treatment effects in cross-over trials*
MP21	Giulia Pontali	*Combining datasets in a mendelian randomization design: metabolomics and parkinson's disease case study*
MP22	YIWEN XU	*Trial emulation for group policies: impact of online consultation system on emergency hospital utilisation*
MP23	Junxian Zhu	*Accounting for non-compliance when estimating treatment effects of ordinal outcome in rct*
MP24	Urko Aguirre	*Risk factors for the survival of colorectal cancer patients modeled through a competitive risk analysis.*
MP25	Aziliz Cottin	*Ms-cpfi: a model-agnostic interpretable counterfactual perturbation feature importance for multi-state models*
MP26	Elsa COZ	*Flexible log-hazard model for adverse events analysis*
MP27	Xiang Geng	*Sample size determination based on restricted mean time lost in the presence of competing risks*
MP28	Thao Le	*Handling missing disease information in diseases that need two visits to diagnose*
MP29	Megan McGovern	*Using multistate models to explore sociodemographic inequalities in the risk of adverse pregnancy outcomes*
MP30	Lauren Rengger	*How rheumatoid arthritis disease activity affects the risk of cardiovascular multimorbidity*
MP31	zhiyin yu	*Restricted mean time lost model for covariates with time-varying effects*
MP32	jiaoyang cai	*Covid-19 infection in children with acute lymphoblastic leukemia in china: mild clinical course*
MP33	nadia dardenne	*Determinants of vaccine intention against covid-19: a serial mediation approach*
MP34	Anikó Lovik	*Social isolation and mental health of older citizens during the covid-19 pandemic*
MP35	Stefania Mondello	*Covid-19 health interventions targeting migrant populations: a systematic review*
MP36	Jose M Quintana	*Short-term forecasting evaluation of covid-19 epidemic waves in the basque country*
MP37	Nicolas ROMAIN-SCELLE	*Usefulness of ecological socio-economic indicators in sars-cov-2 infection modeling: a french case study*
MP38	Ruqayya Azher	*Optimal stage-wise allocation ratios in multi-arm multi-stage designs*
MP39	Svetlana Cherlin	*A latent variable model for borrowing of information in a basket trial*
MP40	Chieh Chiang	*Tolerance interval-based hypothesis testing for the similarity between two independent populations*
MP41	Maria Vittoria Chiaruttini	*Sample size recalculation for a skewed outcome in two-stage three-arm sequential noninferiority clinical trial*
MP42	Noura Darwish	*Use of an adaptive design in a randomly assigned, open-label, balanced incomplete block design study*
MP43	Heather Gunn	*A comparison of sample size re-estimation methods for cluster-randomized trials*
MP44	Richard Holubkov	*Design of a cluster-randomized trial:telehealth-enabled physician adherence monitoring in intensive care units*
MP45	Athina Kranidi	*Statistical validation methods of sound recording medical device on respiratory clinical trials*
MP46	Satomi Okamura	*Non-inferiority trials with evidence of assay sensitivity using population adjustment method*
MP47	Francesca Orsini	*When covid-19 specific vaccines became an intercurrent event*
MP48	Jan Wiemer	*Experiences with a bayesian adaptive design for the evaluation of a biomarker-based treatment algorithm*
MP49	John Belcher	*Modelling the rate of change of gfr in the presence of competing risks*
MP50	Ricarda Graf	*Multivariate repeated measures analysis - testing and post-hoc procedures under non-normality*
MP51	Eleonore Herquelot	*The use of multilevel models to explain regional disparities: illustration on pneumococcal vaccination*
MP52	Sofia Kaisaridi	*A multivariate bayesian mixed-effect model (leaspy) to analyze the trajectory of cognitive decline in cadasil*
MP53	Kylie Lange	*Performance of mixed effects models for partially clustered trials*
MP54	Annalisa Orenti	*10-Years changes in lung function of cystic fibrosis patients in europe: different statistical methods at work*
MP55	Taesung Park	*Structural pathway analysis of longitudinal multinomial phenotypes*
MP56	Jules Pereira Macedo	*Estimation of a risk difference in a cluster randomized trial.*
MP57	Diana Trutschel	*Applying network analysis for a high-granulated description of patients`needs in maternity care from ehrs*
MP58	SARA ALBASINI	*Deep learning based pipeline for automatic classification of breast microcalcifications*
MP59	Lucas Anzelin	*Prediction of neurodevelopment at two years of age among very preterm infants in the île-de-france region*
MP60	Lasai Barreñada	*Understanding random forests from a statistician's point of view: a simulation study*
MP61	Benoit Courbon	*Holographic microscopy data analysis with deep learning for the detection of antimicrobial mechanism of action*
MP62	Eleonora Di Carluccio	*Personalized diagnosis in suspected myocardial infarction: the artemis study*
MP63	Sarah Friedrich	*Regularization approaches in clinical biostatistics: a review of methods and their applications*
MP64	Saebom Jeon	*Effects of irregular sleep behavior on physiological and perceived health in shift workers*
MP65	Marzieh Mahmudimanesh	*Using deep learning for time series regression models in health: a simulation study*
MP66	Autumn O'Donnell	*A systematic review on machine learning techniques for survival analysis in cancer*
MP67	Josline Adhiambo Otieno	*Using machine learning to predict multi-class functional outcomes and death 3 months after stroke in sweden*
MP68	Pascal Rink	*Post-selection confidence bounds for prediction performance*

MONDAY 28 AUGUST 2023

MP69	Emma Todd	*Are lifestyle factors the most important predictors of common mental disorders? A systematic review*
MP70	Mi Hong Yim	*Classification of metabolic syndrome using arterial pulse wave via machine learning approaches*
MP71	Omololu Aluko	*Advanced statistical methods of handling ordinal missing data*
MP72	Joydeep Basu	*Using early or baseline data in a trial with missingness in a continuous primary endpoint*
MP73	Anca Chis Ster	*Evaluating bias when estimating causal mediation estimands with non-adherence and missing data*
MP74	Daisy Gaunt	*Simulation study of four methods for sensitivity analysis of the mar assumption with an application to rcts*
MP75	Sophie Greenwood	*Expert elicitation methods for the evaluation of missing data assumptions: a scoping review*
MP76	Nina Haug	*Correcting for bias due to missing data in longitudinal studies on quality of cancer patients*
MP77	Kawabata Emily	*Accounting for data missing not at random: comparison of bayesian and monte carlo probabilistic bias analyses*
MP78	Markéta Janošová	*Methods of dealing with attrition bias due to non-random dropout in models with binary outcome*
MP79	Shahab Jolani	*Imputation of cross-classified multilevel models*
MP80	Melissa Middleton	*Combining multiple imputation and inverse probability weighting to handle missing data in longitudinal studies*
MP81	Lucinda Archer	*Instability of clinical prediction models caused by dichotomisation of a continuous outcome*
MP82	VASILIKI BARALOU	*Real-time detection of hiv outbreaks among people who inject drugs and modelling of subsequent epidemic states*
MP83	Paula Dhiman	*Studies developing prediction models are not considering sample size requirements: a systematic review*
MP84	Zoë Dunias	*A comparison of hyperparameter tuning procedures for clinical prediction models: a simulation study*
MP85	María Escorihuela	*Clinical utility curve: a new proposal to analyze the utility of predictive models*
MP86	Shan Gao	*Predicting central line-associated bloodstream infections in hospitalized patients: a systematic review*
MP87	Frissiano Honwana	*Extended joint models for longitudinal viral load and virologic failure in hiv at western cape (south africa)*
MP88	Alexandra Hunt	*Clinical prediction models for transition to psychosis in individuals meeting at risk mental state criteria*
MP89	Amir Jalali	*Risk prediction models for individual diagnosis: a prostate cancer case study*
MP90	Elvis Karanja	*Real time prediction of infectious disease outbreaks based on google trend data in africa*
MP91	Samuel Kilian	*Developing a prediction model for survival – a case study*
MP92	Mi Mi Ko	*Development of blood stasis questionnaire on gynecological diseases*
MP93	Hye Ah Lee	*Performance evaluation and improvement of the framingham diabetes risk model using community-based koges data*
MP94	Yan Li	*Development and validation of a prognostic model for institutionalisation in parksinsonism: ipd meta-analysis*
MP95	Dave Lunn	*A method for predicting clinical trial enrollment under restrictive constraints*
MP96	Francesca Maher	*An external validation of the kfre in ckd patients of south asian ethnicity*
MP97	Masako Nishikawa	*Population modeling for circannual rhythms of hba1c in type 2 diabetic patients using large registry data*
MP98	Begum Irmak On	*Current methodological challenges in prognostic modeling - use case multiple sclerosis*
MP99	Juliette ORTHOLAND	*Longitudinal and survival joint prediction: time reparameterization in amyotrophic lateral sclerosis context*
MP100	Doug Thompson	*Introducing the 'tetris' plot for visualising the instability in variable selection*
MP101	Junfeng Wang	*Transforming a published nomogram back to formulas: how to do this manually or with ai*
MP102	Zijing Yang	*Dynamic prediction based on conditional restricted mean survival time for right-censored data*
MP103	Evert Cleenders	*Segmented regression in the context of kidney function after transplantation*

MP104	Emily Granger	*Benchmarking an emulated trial against a real target trial: challenges and illustration in cystic fibrosis*
MP105	Anita Lindmark	*Extending interventional disparity (in)direct effects to investigate inequalities in adverse stroke outcome*
MP106	Innocent Mboya	*Age and time-trends of the association between body mass index and mortality: evidence from the odds study*
MP107	Laura Quinn	*Interobserver variability of recall decisions between mammogram readers in breast cancer screening*
MP108	Rebecca Rylance	*Assessing the external validity of the validate-swedeheart trial*
MP109	Mohadeseh Shojaei Shahrokhabadi	*Exploring the sample quality: the comparison of data from the mapme2 intervention with ncmp, england, 2021/22*
MP110	Naïla Aba	*High-dimensional selection in a matched case-control study: cardiotoxicity in childhood cancer survivors.*
MP111	Abdelmajid Djennad	*Estimating unobserved prevalence of opiate and crack use in england*
MP112	Camille Giampiccolo	*Impact of combined exposure to multiple air pollutants on breast cancer risk using bayesian profile regression*
MP113	Bas Kellerhuis	*Effect of bias, composition, and decision-making in expert panels on diagnostic accuracy estimates*
MP114	Chiara Casamassima	*Identification of chronic patients with acute chikungunya: an analytical method based on agreement coefficient*
MP115	Pawin Numthavaj	*Metastatic prostate cancer treatment: umbrella review of systematic reviews and meta-analyses*
MP116	Erica Ponzi	*Challenges in planning and conducting evidence based studies in a pandemic: are rcts really the gold standard?*
MP117	Roma Puronaitė	*Using nonlinear models to identify breakpoints in disease prevalence among patients with multimorbidity*
MP118	Myanca Rodrigues	*Comparing approaches to rank interventions in a network meta-analysis: patients with opioid dependence*
MP119	Hokeun Sun	*Group penalized exponential tilt model for differentially methylated genes in epigenetic association study*
MP120	Toshiro Tango	*Can we estimate a risk without observing the relevant number of cases ?*
MP121	Oludare Ariyo	*Bayesian bent-cable model for longitudinal and survival time with heterogenous random-effects*
MP122	Samia Ashrafi *	*Detection of multiple change points in survival analysis with narrowest significant pursuit technique*
MP123	Rouven Behnisch	*Comparison of methods to analyze time-to-event endpoints in trials with delayed treatment effects*
MP124	Moritz Fabian Danzer	*Adaptive group sequential designs for clinical trials with multiple time-to-event outcomes in markov models*
MP125	Julie Dudasova	*The effect of immune correlates on the precision of vaccine efficacy evaluation: demographic subgroups*
MP126	Hiroya Hashimoto	*Impact of censoring on estimation for restricted mean survival time in small sample size*
MP127	Jinheum Kim	*Risk factors and transitional probability of clinical events in korean ckd patients using the multistate model*
MP128	Jayoun Kim	*Bias reduction for semi-competing risks model with rare events: application to a chronic kidney disease*
MP129	Kelsi Kroon	*Prevalence-incidence mixture model for the risk of high-grade cervical lesion based on individual risk factors*
MP130	Shun-Fu Lee	*Comparison of total event analysis in three cardiovascular trials with composite outcomes*
MP131	Shirin Moghaddam	*Non-parametric bayesian imputation of right censored data in survival analysis*
MP132	Eni Musta	*Testing for sufficient follow-up to detect the cured proportion*
MP133	Ulrike Pötschger	*Approaches to deal with non-proportional hazards in paediatric oncology - the context matters.*
MP134	Alessandro Previtali	*Asking the right question when assessing os in trials that allow for cross over: considerations and case study*
MP135	Elena Tassistro	*Adverse events in trials with survival outcomes: from clinical questions to methods for statistical analysis*

***WINNER OF ISCB44 CONFERENCE AWARD FOR SCIENTISTS**

MP01 On detecting outliers in a circular-linear regression by cephalometric angles data

Altuntas M.*[1], Karaibrahimoglu A.[2]
[1] *Sinop University ~ Sinop ~ Turkey, 2Suleyman Demirel University ~ Isparta ~ Turkey*

Orthodontic problems are one of the important health problems that affect the oral and dental health and aesthetic appearance of people. These problems are revealed by examining cephalometric angles. The aim of this study is to analyze the relationship between nasolabial angle and nose prominence with the circular-linear regression model and to detect outlying observations. A total of sixty patients with orthodontic problems were included in the study. Radian measure of nasolabial angle as circular dependent variable and the nose prominence observations as linear explanatory variable were used in simple circular-linear regression with von Mises distributed error. The estimates of the unknown parameters of model that are called mu, beta, kappa and standard errors of the estimators were obtained by applying iteratively reweighted least square method. The mean direction of nasolabial angle was 1.911 radian (109.50 degree), and the correlation between the variables was r=0.130 in total patients. The parameters' estimates were calculated -0.575, 3.609 (in radians) and 32.170, respectively [1]. One of the basic problems in regression modelling is the presence of outliers in any dataset, such observations influence the statistical inferences. Therefore, it is quite important to detect and evaluate these observations and to reduce their effect on the model of interest. The methods utilized to detect outliers in circular data are different from those for linear data and, recently, have rarely been used in the literature. In this study, the use of Huberized and CovRATIO methods was preferred to detect outliers in cephalometric angle data [2]. It was determined that there were ten outliers in data according to both methods. Analyses results and graphs were obtained using R programming codes. It can be concluded that the analysis of the cephalometric angles should be performed by circular regression models to obtain more accurate results. Also, it was emphasized that in the regression analysis of a circular data, outlying observations in the data should be correctly determined since they affect the parameter estimates of the model.

[1] H.H. Mohammad, S.Z.Satari, W.N.S.W.Yusoof, Review on circular-linear regression models, Journal of Physics, 2021, doi: 10.1088/1742-6596/1988/1/012108.
[2] A. Abuzaid, I. Mohamed, A.G. Hussin, A. Rambli, COVRATIO Statistic for Simple Circular Regression Model Chiang Mai. J. Sci., 38(3), 2011, 321-330.

MP02 A comparison of dose-finding method for incorporating single-agent historical data into combination trial

Hashizume K.*[1], Tsuchida J.[2], Sozu T.[1]
[1]*Tokyo University of Science ~ Tokyo ~ Japan,* [2]*Doshisha University ~ Kyoto ~ Japan*

Phase I combination clinical trials of anticancer drugs start after the tolerability of each drug is evaluated in phase I single-agent (monotherapy) clinical trials. In either trial, one of the crucial objectives is to identify the maximum tolerated dose. Several dose-finding methods to incorporate data from single-agent trials (SA historical data) into subsequent combination trials have been proposed. However, the relative performance of those methods, especially with heterogenous data between single-agent and combination trials, is largely unknown. Therefore, this study aimed to find a robust and stable method for the identification of the maximum tolerated combination dose under various toxicity scenarios with both homogenous and heterogenous data. We examined the relationship between the amount of SA historical data incorporated and the performance of dose-finding methods (e.g., the percentages of correct selection of maximum tolerated dose combination) by conducting an extensive simulation study with many random toxicity scenarios with both homogeneity and heterogeneity. We compared the performance by multiple statistical models including a copula-based model and a Bayesian logistic regression model that naturally incorporated SA historical data. We also evaluated the difference in performance between the incorporation methods of the basic power prior and meta-analytic prior. Our simulations showed that the performance among the methods was almost comparable. In addition, we found that incorporating SA historical data when it is homogenous with the current data is beneficial, and incorporating heterogeneous data is also advantageous, provided heterogeneity is not extremely high.

[1] Hashizume, K., Tshuchida, J. and Sozu, T. (2021). Flexible use of copula-type model for dose- finding in drug combination clinical trials. Biometrics 78 (4), 1651–1661.
[2] Neuenschwander, B., Roychoudhury, S. and Schmidli, H. (2016). On the use of co-data in clinical trials. Statistics in Biopharmaceutical Research 8 (3), 345–354.

Poster Sessions

MP03 On the comparison of classical and wrapped cauchy circular regression in patients with gastric carcinoma

Karaibrahimoglu A.*[1], Altuntas M.[2]
[1]*Suleyman Demirel University ~ Isparta ~ Turkey,* [2]*Sinop University ~ Sinop ~ Turkey*

Introduction and Objective(s): The follow-up of patients is very important in healthcare, therefore some important dates, such as diagnosis, treatment, operation, etc, should be into consideration, especially in cancer patients. However, generally, the date or angle type of data is not analyzed circularly. The aim of the study is to compare the von Mises circular-circular regression model with the wrapped Cauchy robust method to regress diagnosis and operation months in patients with gastric carcinoma. Method(s) and Results: The diagnosis and operation months of a total of 331 patients with gastric carcinoma were included in the study. Moreover, a dataset with a smaller sample (n=71) consisting of several outliers was set up. von Mises circular-circular and wrapped Cauchy robust models were established for both sample sets. The estimated parameters alpha and beta, the concentration parameters kappa and rho, standard errors of the parameters, goodness-of-fit indices, and mean circular error (MCE) of the models were calculated for both sample sizes. The analyses were performed by R programming codes for the theoretical background of the wrapped Cauchy robust model. The mean directions of diagnosis and operation months were 5.289 (~May) and 4.132 (~April), and the correlation between the variables was r=0.243 in total patients [1]. The parameter estimates were 0.23 and 0.99 respectively in von Mises model whereas they were found as 1.18 and -1.74 in the wrapped Cauchy model. The MCE value for the classical model was 0.129, but 0.966 for the wrapped Cauchy model. Although the number of outliers was very large in the smaller sample, the standard errors for estimates were found as 0.030 and 0.049 in the wrapped Cauchy model whereas they were 0.010 and 0.002 for the classical model. Moreover, the goodness-of-fit values for both models were close to each other [2]. Conclusions: Robust regression models like the Wrapped Cauchy method are often used when there are many outliers in a dataset. So they give more accurate results and parameter estimates. In our study, although the classical model gave more accurate results, the wrapped Cauchy robust method should be used in heavy-tailed data.

[1] A. Karaibrahimoglu, S. Ayhan, M. Karaagac, M. Artac, Journal of Applied Statistics, 48:13-15,2021, 2931-2943.
[2] S. Kato, M.C. Jones, Bernoulli, 19(1), 2013, 154–171.

MP04 Estimation of between-study variance in bayesian meta-analysis: borrowing information from excluded studies

Okita M.*, Emoto R., Matsui S.
Department of Biostatistics, Nagoya University Graduate School of Medicine, ~ Nagoya ~ Japan

Instability in estimating between-study variance of treatment effects has been recognized as one of critical limitations in many meta-analyses integrating small numbers of studies. For this issue, it is natural to consider a Bayesian approach that incorporates external variance estimates from past meta-analyses in a similar disease field as prior information [1]. In this paper, we consider another Bayesian approach that borrows variance information from a set of studies in the same meta-analysis that are selected at an initial screening stage of systematic review, but finally excluded due to incompatibility to strict selection criteria. As a simple way to incorporate the information from such excluded studies, we develop Bayesian methods with modified power priors [2] in some variance models. Numerical evaluations on the performance of the proposed methods, including the application to real meta-analyses, will be provided. The proposed approach can work with various Bayesian priors other than power priors for borrowing information from excluded studies in the same meta-analysis, as well as incorporate external variance information from past meta-analyses in a similar disease field.

[1] Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. J Clin Epidemiol. 2015;68(1):52-60.
[2] Duan, Y., Ye, K., & Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. Environmetrics, 17(1), 95–106.

Poster Sessions

MP05 Decision-making frameworks using multiple correlated endpoints

Riviere M.*
Saryga ~ Paris ~ France

Nowadays, quantitative decision making is an essential component of early clinical development. Rather than focusing on statistical significance only, some methods have been proposed in the literature that led to three-outcome decision-making frameworks by including a consider zone. NOGO/Consider/GO decision zones are determined based on pre-specified target value (TV) and lower reference value (LRV). In this context, historical data can be incorporated as an informative prior using Bayesian analysis.
We extended the classical decision-making framework to cases with multiple correlated endpoints that can be of various types (e.g. co-primary, hierarchical, ...). We considered bivariate normal and bivariate binomial endpoints, for one arm or two arms trials, both in Bayesian and Frequentist analyses. Operating characteristics with different correlations and under different scenarios including LRV and TV are presented. Different options to determine decision thresholds are discussed. Abellan et al., Implementation and practical aspects of quantitative decisionmaking in clinical drug development, Submitted on 16 Jul 2022
Frewer et al., Decision-making in early clinical drug development. Pharmaceutical Statistics, 2016; 15: 255-263. Fisch et al., Bayesian Design of Proof-of-Concept Trials. Therapeutic Innovation & Regulatory Science, 2014; 49: 155-162. Lalonde et al., Model-based drug development. Clinical Pharmacology & Therapeutics, 2007; 82(1): 21-32. Quan et al., Applications of Bayesian analysis to proof-of-concept trial planning and decision making, Pharm Stat, 2020 Jul;19(4):468-481. Wasserstein et al., Moving to a World Beyond "p < 0.05". The American Statistician, 2019, 73:sup1, 1- 19
Sverdlov et al., Exact Bayesian Inference Comparing Binomial Proportions with Application to Proof- of-Concept Clinical Trials Ther Innov Regul Sci, 2015, Vol. 49(1) 163-174. Pulkstenis et al., A Bayesian paradigm for decision-making in proof-of-concept trials, J Biopharm Stat, 2017;27(3):442-456. Wiesenfarth and Calderazzo, Quantification of prior impact in terms of effective current sample size, Biometrics, 2020 Mar;76(1):326-336.

MP06 Bivariate copula for left and right censoring: validation of a complex surrogate for a time-to-event endpoint

Capitoli G.*[2], Galimberti S.[2], Valsecchi M.G.[2], Rotolo F.[1]
[1]*Sanofi R&D, Oncology Biostatistics, Biostatistics and Programming Department ~ Montpellier ~ France,* [2]*Department of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy*

Early endpoints correlated to the patient prognosis and the treatment effect on clinically relevant long- term endpoints are known as surrogate endpoints (SEs) of a true endpoint. In hematological malignances, Minimal Residual Disease (MRD) -that measures the concentration of residual leukemic cells in the blood or the bone marrow- is a promising early endpoint. However, despite the strong rationale, the MRD response rate showed mixed evidence as SE for time-to-event endpoints across different indications [1] and is accepted by FDA only for accelerated approval to date. As the MRD response rate results from a categorization (dichotomic or in 3 ordered categories), the raw MRD concentration is potentially a more informative endpoint but there is a lack of methods for surrogate validation that deal with quantitative variables with a complex distribution. Indeed, due to a high proportion of low values corresponding to MRD-negative samples, MRD concentration is a random variable with substantial left censoring. In addition to this apparent peak close to 0, its distribution is highly skewed. In the context of the meta-analytic validation of SEs [2], we propose a copula model accommodating left-censoring for the candidate SE and right-censoring for the clinical time-to-event endpoint (true endpoint). Specifically, we derived the likelihood of bivariate copulas considering 3 copula functions (i.e. Clayton, Hougaard and Plackett) and Gamma and Weibull distributions for the surrogate and the true endpoint, respectively. Results of the motivating clinical study on MRD suggest that very low MRD values are associated with relatively long Event Free Survival (EFS) values. The degree of correlation obtained is higher than the one from the analysis simplified to 3 ordinal categories. The use of copula models accounting for left-censoring allows to avoid the loss of information due to dichotomization or categorization. Such adaptation of copula models extends and naturally fits into the state-of-the-art statistical methodology for the validation of SEs.

[1] S. Galimberti et al. JNCI Cancer Spectrum, Volume 2, Issue 4, 2018.
[2] T. Burzykowski, G. Molenberghs, M. Buyse, H. Geys, D. Renard. Journal of the Royal Statistical Society Series C: Applied Statistics, Volume 50, Issue 4, 405–422, 2021.

MP07 Reducing the overfitting in the groc curve estimation

Diaz--Coto S.*, Martinez--Camblor P.
Geisel School of Medicine at Dartmouth, Dartmouth College ~ Hanover ~ United States of America

The generalized receiver-operating characteristic, gROC, curve has been proposed to assess the diagnosis ability of biomarkers when both larger and lower values are associated with higher probabilities of being positive. Its empirical estimation implies the selection of the best classification subsets among those satisfying particular conditions. Both strong and wake consistency have been already proved. However, using the same data to select the classification subsets and calculate its gROC curve may lead to an over-optimistic estimation of the performance of the diagnosis criteria on future samples. We studied the bias of the gROC curve estimator and propose how to reduce it. The empirical estimator of the gROC curve can be obtained when the underlying classifications subsets are computed satisfying the self-contained condition ('with restrictions'), and when they are free of restrictions ('without restrictions'). We explored through Montecarlo simulations the reported bias in both situations and proposed two cross-validation based algorithms to reduce the issue. These new procedures remove the bias produced in the ROC curve context by leaving out pairs instead of units. The algorithm for the 'with restrictions' gROC curve is mainly based on the underlying transformation associated with this curve [1] and the one for the 'without restrictions' gROC involves a parametric part [2]. Both improve the estimation of the actual diagnosis accuracy of the biomarker and get almost unbiased areas under the gROC curve, in most of the considered scenarios. However, the cross-validation based algorithms reported larger L1-errors than the standard empirical estimators and incremented the computational cost of the procedures. The practical application of the proposed algorithms is illustrated through the analysis of a real-world dataset. Flexible statistical techniques frequently imply the estimation of a number of parameters which increase the risk of reporting over-fitted results. The empirical gROC curve estimator has been shown to report optimistic conclusions. To overcome this issue, we proposed two cross-validation-based algorithms adapted for dealing with the 'with' and 'without restrictions' gROC curves. The procedures remove the overfitting and reduce the absolute bias.

[1] Martínez-Camblor P, Pérez-Fernández S, Díaz-Coto S (2021) The area under the generalized receiver-operating characteristic curve. Int J Biostat 18:29
[2] Rutter C, Gatsonis C (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 20(19):2865–2884

MP08 Enhancing the efficiency of predictive marker analyses by using treatment crossover data

Emoto R.*[1], Igeta M.[2], Matsui K.[1], Matsui S.[1]
[1]*Nagoya University Graduate School of Medicine ~ Nagoya ~ Japan,* [2]*Hyogo College of Medicine ~ Nishinomiya ~ Japan*

Although the analysis of predictive markers is a key step in understanding the molecular mechanisms associated with between-patient heterogeneity in treatment responses and in developing diagnostics for treatment selection, the issues of confounding and lack of efficiency (in assessing treatment-by- marker interactions) will make such an analysis difficult in practical clinical studies. One possible solution to these issues is to consider within-patient treatment comparison. However, there are few studies of predictive marker analysis using treatment crossover data, except some on treatment selection [1,2]. In this study, we extend the traditional framework for statistical inference for the main effect of treatment in a standard two-sequence, two-period crossover design to that for treatment-by-marker interactions. Based on several transformations of continuous outcome variables within patients, two methods of predictive marker analysis are developed, without modeling random effects of between- and within-patient variations in the outcome variables, one assuming no carryover effects and the other estimating these effects. In scenarios of various between- and within-patient variabilities, we observed substantial efficacy enhancement by the first method in testing marker predictiveness and in estimating predictive signatures, compared to the standard analysis in parallel-group, randomized clinical trials. The proposed statistical framework and methods for predictive marker analysis in crossover trials with limited sample sizes are expected to be useful in many situations involving predictive markers, such as proof-of-concept evaluation of personalized medicine in early phase clinical trials and development and validation of personalized medicine for rare diseases.

[1] Nguyen, C. T., Luckett, D. J., Kahkoska, A. R., Shearrer, G. E., Spruijt-Metz, D., Davis, J. N., and Kosorok, M. R. (2020). Estimating individualized treatment regimes from crossover designs. Biometrics 76, 778–788.
[2] Kulasekera, K. and Siriwardhana, C. (2022). Multi-response based personalized treatment selection with data from crossover designs for multiple treatments. Communications in Statistics- Simulation and Computation 51, 554–569.

MP09 Modeling and estimation for hierarchical ordinal outcome

Harada K.*[1], Kawano S.[2], Taguri M.[1]
[1]Tokyo Medical University ~ Tokyo, Japan ~ Japan, [2]Kyushu University ~ Fukuoka ~ Japan

In disease prediction by biomarkers, multivariate prediction models are often employed. When predicting the presence or absence of disease, logistic regression is a better option; however, in some cases, it is necessary to predict not only the presence but also the severity of disease. In such cases, the outcome to be predicted is defined as an ordinal scale: e.g., healthy, mild, moderate, and severe. One option is to apply a classical ordinal logit model to this outcome, but the classical model has the limitation that the regression coefficients are common across all outcome classes. On the other hand, it is natural to think that our outcome has two-level hierarchy: healthy/unhealthy and mild/intermediate/severe. It is impossible to determine in advance whether the same biomarkers are important in both levels or only in one. This motivates us to apply a more flexible ordinal model. One known expansion simply relaxes the common coefficient assumption (e.g. [1]), but the relaxed model is not well-defined as a probabilistic model since its isoprobabilistic curves can cross among different classes, causing a severe problem in interpretation and computation. To develop a more flexible yet interpretable method, we propose a multi-task learning approach for this problem. We use two separate models for the hierarchical ordinal outcome and estimate the parameters by maximizing the penalized likelihood with the structural sparse penalty (e.g. [2]). In order to evaluate the proposed method, we compared it with existing methods in numerical experiments, considering four scenarios that differ in the two tasks. The proposed method consistently showed better or competitive performance in terms of the prediction error than the best existing methods in all scenarios, while every existing method showed poor performance in at least one scenario. The multi-task approach successfully captures the structure the hierarchical ordinal outcome, and it can be useful for constructing a prediction model.

[1] Peterson, B. and Harrell, F. E., J. R. Stat. Soc. Ser. C Appl. Stat., 39, 1990, 205–17.
[2] Simon, N., Friedman, J., Hastie, T. and Tibshirani, R, J. Comput. Graph. Stat., 22, 2013, 231–45.

MP10 Meta-analysis of cell-free micrornas for early breast cancer detection: causes of inconsistent findings

Sehovic E.*[1], Urru S.[2], Chiorino G.[1], Doebler P.[3]
[1]Cancer Genomics Lab, Fondazione Edo ed Elvo Tempia ~ Biella ~ Italy, [2]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences, and Public Health, University of Padova ~ Padova ~ Italy, [3]Department of Statistics, TU Dortmund University ~ Dortmund ~ Germany

A plethora of studies tried to identify cell-free circulating microRNAs (miRs) as a tool for early breast cancer (BC) detection. However, different study designs, laboratory protocols and biomarker modelling tools were applied [1]. A laboratory challenge which exacerbates this problem is the lack of a consistent normalizer for circulating miRs [2]. We aim to evaluate the diagnostic performance and sources of heterogeneity of thus far reported circulating miRs for early BC detection and discuss important steps for developing clinically viable tools. Fifty-six eligible research articles that reported diagnostic circulating miRs for detection of BC discovered by Real-Time Quantitative Reverse Transcription PCR were analysed. We ran a bivariate generalized linear mixed-effects model to obtain pooled sensitivity and specificity. The sample and study models' characteristics (ROC curve shape and derived perceived cost for not detecting a BC patient) were analysed to determine the potential preference of studies for sensitivity or specificity. Pooled sensitivity of 0.85 [0.81-0.88] and specificity of 0.83 [0.79-0.87] were obtained with a slight tendency of studies to prefer specificity. Subgroup analyses showed significant heterogeneity in the normalizer type, with endogenous normalizers having a higher diagnostic performance than exogenous ones. Further, a significantly better performance of multiple (sensitivity: 0.90 [0.86-0.93]; specificity: 0.86 [0.80-0.90]) vs single (sensitivity: 0.82 [0.77-0.86], specificity: 0.83 [0.78-0.87]) miR panels was obtained. Importantly, we observed a comparable pooled diagnostic performance between studies using serum (sensitivity: 0.87 [0.81-0.91]; specificity: 0.83 [0.78-0.87]) and plasma (sensitivity: 0.83 [0.77-0.87]; specificity: 0.85 [0.78-0.91]) as specimen type. The diagnostic ability of circulating miRs to detect early BC was reaffirmed, with no significant difference in diagnostic performance between plasma and serum studies and a superior diagnostic performance of panel to individual miRs. To overcome standardization and reproducibility issues in studies dealing with diagnostic circulating miRs, authors should utilize radio-based methods instead of using normalizers or applying global normalization [2], clearly report model methodology including cut- offs, strive to create parsimonious models, include independent validation cohorts as well as be aware that modelling and study-design decisions affect sensitivity or specificity preference.

[1] Aggarwal, V., Priyanka, K., & Tuli, H. S. Molecular diagnosis & therapy, 24(2), 2020, 153-173.
[2] Deng, Y., Zhu, Y., Wang, H., Khadka, V. S., Hu, L., Ai, J., Dou, Y., Li, Y., Dai, S., Mason, C.E., Wang, Y., Jia, W., Zhang, J., Huang, G., Jiang, B. Analytical chemistry, 91(10), 2019, 6746-6753.

Poster Sessions

MP11 Estimating dce-mri ktrans repeatability by meta-analysis of data from 12 clinical studies

Sridharan N.*, O"Connor J., Porta N.
The Institute of Cancer Research ~ London ~ United Kingdom

Functional imaging methods such as dynamic contrast enhanced magnetic resonance imaging (DCE- MRI) are used to assess response to cancer therapy in early phase trials. The most common measurement is median Ktrans, an imaging biomarker that quantifies a composite of blood flow and vessel permeability. Measurement repeatability enables the magnitude of true change in individual lesions to be identified and is usually estimated in test-retest studies [1]. The aim of this study is to estimate multisite repeatability analysis of Ktrans using individual patient data (IPD) meta-analysis of different test-retest studies identified in a previous systemic review that showed variation in the reporting of Ktrans repeatability (ISCB 2022). We contacted authors of the selected studies in the systematic review to provide IPD, and also obtained IPD from further test-retest DCE-MRI studies conducted at Royal Marsden Hospital and University of Manchester, UK. A multilevel mixed effects model was used to perform one-stage IPD meta-analysis[2] to estimate the overall within-patient variance and account for three levels of clustering (lesion, patient and study), and derive the within-subject coefficient of variation (wCV) on the original scale. Next, sub-group analysis was performed to identify potential modifiers including type of cancer, type of lesion (primary tumour, metastasis, nodal) and tumour location (head, thorax, abdominal, pelvis). 12 DCE-MRI studies were identified. Collectively, these included 276 patients and 408 lesions for analysis. The study cohorts had wCV values ranging from 10.7% to 32.6%. The overall estimate of wCV was 21.91% (95% CI: 20.41%, 23.51%). There were significant differences in the estimated wCV for disease type: 13.91% (95% CI: 11.81%, 16.39%) for ovarian cancer lesions, 21.97% (95% CI: 19.87%, 24.29%) for colorectal lesions, and 26.59% (95% CI: 20.67%, 34.33%) for NSLC. No differences were observed according to type of lesion or its location. Median Ktrans repeatability in DCE-MRI studies was estimated. Further analysis showed significant differences in wCV between different tumour types, most notably with Ktrans measured from ovarian cancer lesions being more repeatable than Ktrans from non-small cell lung cancer lesions and colorectal lesions.

[1] Raunig, D.L., et al., Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment (2015). Statistical Methods in Medical Research 24(1):27-67.
[2] Huang, E.P., et al., Meta-analysis of the technical performance of an imaging procedure: guidelines and statistical methodology (2015). Statistical Methods in Medical Research 24(1):141-74.

MP12 The role of quantitative eeg (qeeg) in clinical trials: a methodological review

Waser M.*
Biostatistics and Data Science, Staburo GmbH ~ Munich ~ Germany

Electroencephalography (EEG) facilitates non-invasive assessments of brain function with high time resolution. Observational studies have shown the potential of quantitative EEG (qEEG) measures as diagnostic biomarkers for neurological indications [1][2]. There are, however, no universal standards for qEEG application in interventional trials. This research reviews the current role of qEEG in clinical trials to provide reference for future consolidation efforts. A systematic search on ClinicalTrials.gov, the largest clinical trial registry, was conducted up to 09- Mar-2023. Search terms were "eeg" OR "electroencephalogram" OR "electroencephalography". Only planned, active or completed interventional trials in clinical phases 1–4 were included into the analysis; observational or terminated trials, as well as trials with unknown status, were excluded. The resulting trial summaries were manually reviewed to assess medical condition, intervention, EEG purpose, cognitive state during EEG recording, and qEEG endpoints. Outcomes were described by relative frequencies; no inferential analyses were conducted. The search resulted in a total of 663 included trials of which 20.8% were conducted in healthy subjects. Main medical conditions included epilepsy (13.4%), major depressive disorder (8.0%), surgery with anesthesia (7.4%), schizophrenia (4.9%), Alzheimer's disease (2.9%) and autism spectrum disorder (2.9%). The majority of trials (80.2%) used drug therapy as intervention; other interventions included magnetic, electric or light stimulation (10.7%), medical devices (2.7%), and neurofeedback training (2.3%). EEG was mainly used to demonstrate efficacy (68.8%) and for (anesthesia) monitoring (17.2%). Interestingly, EEG was conducted as part of the intervention in 6.8% of trials, e.g. in brain-computer interface applications. EEGs were recorded in resting state (20.7%), during stimulation paradigms (11.9%), during anesthesia (10.3%), during sleep (10.3%), or over longer periods with no specified state (9.0%). QEEG endpoints were mostly spectral power in relevant frequency bands (25.0%), spike-wave activity indicating seizures (15.5%), event-related potentials (12.4%) and sleep parameters (7.4%). Only few trials (3.2%) measured EEG complexity or connectivity. Review results suggest three major current roles of qEEG in interventional clinical trials: 1. Epileptic seizure detection, 2. anesthesia monitoring during surgery, and 3. efficacy assessment of interventions for neurological and mental and behavioral disorders, mostly via power spectral analysis.

[1] F. Ferreri, F. Miraglia, F. Vecchio, N. Manzo, M. Cotelli, E. Judica, PM Rossini: Electroencephalographic hallmarks of Alzheimer's disease. Int J Psychophysiol. 181, 2022, 85-94.
[2] F.S. de Aguiar Neto, J.L.G. Rosa: Depression biomarkers using non-invasive EEG: A review. Neurosci Biobehav Rev. 105, 2019, 83-93.

Poster Sessions

MP13 Mediation analysis of obesity and adiponectin in breast cancer risk: a nested-cohort study in the ibisii trial

Macis D.[1], Bellerba F.*[1], Valentina A.[1], Johansson H.[1], Guerrieri--Gonzaga A.[1], Lazzeroni M.[1], Sestak I.[2], Cuzick J.[2], De Censi A.[3], Bonanni B.[1], Gandini S.[1]
[1]European Institute of Oncology ~ Milan ~ Italy, [2]Queen Mary University ~ London ~ United Kingdom, [3]Ente Ospedaliero Ospedali Galliera ~ Genoa ~ Italy

Obesity is a risk factor for postmenopausal breast cancer (BC) and evidence suggests a role of adiponectin in the relationship between obesity and BC. We applied mediation analysis to investigate whether the effect of body mass index (BMI) on postmenopausal BC risk was mediated by adiponectin and/or other biomarkers in a cohort study nested in the IBIS-II Prevention Trial. We measured adiponectin, leptin, IGF-I, IGFBP-1, h-CRP, glycemia, insulin, HOMA-IR index, and SHBG in baseline and 12-month serum samples from 123 cases and 302 matched controls in the placebo arm of the IBIS-II Prevention trial. To break the matching between the cases and their controls and include the time dimension in the investigation of the association between the exposures and outcome (BC event), a weighted Cox regression analysis was employed. To obtain inferences for the full cohort, the women included in the analysis were up-weighted using the Borgan II weights. Weighted pseudo-likelihood was calculated from Cox regression models to estimate hazard ratios (HR) and 95% confidence intervals (95%CI) using robust standard errors. Hypotheses of causal relationships were graphed through directed acyclic graphs based on biological rationale. Single- and two-mediator models were employed using the approach proposed by Huang et al.[1], to assess if the total effect of BMI on BC risk was mediated by the 12-month adiponectin increase and any of the biomarkers. All models were adjusted for the Tyrer-Cuzick[2] score difference from the expected value in the population and lipid-lowering medications intake. We found no mediating effect in the multi-mediator analyses. The single-mediator analysis confirmed the lack of mediating effect of the adiponectin increase on the total effect of BMI on BC risk (Natural Indirect Effect: HR, 1.00; 95%CI, 0.98-1.02). However, in the multivariable Cox model, both 12-month adiponectin increase (HR, 0.60; 95%CI, 0.36-1.00, p=0.05) and baseline BMI were significantly associated with BC risk (HR, 1.05; 95%CI, 1.00-1.09, p=0.03). These results suggest that adiponectin plays an independent role in postmenopausal BC risk and does not mediate the effect of BMI. Increasing adiponectin may be a promising strategy for preventing postmenopausal BC. Tyrer, J., Duffy, S. W. & Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. Statistics in Medicine 23, 1111–1130 (2004).

[1] Huang, Y. T. & Yang, H. I. Causal mediation analysis of survival outcome with multiple mediators. Epidemiology 28, 370–378 (2017).

MP14 Staggered difference-in-differences for estimating effects of an online consultation system on prescribing

Ellis K.*[1], Keogh R.H.[1], Clarke G.M.[2], O"Neill S.[1]
[1]*London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom,* [2]*The Health Foundation ~ London ~ United Kingdom*

Adoption of online consultation (OC) systems by GP practices aims to improve patient access by enabling patients to make contact by submitting an online form. However, changes in healthcare delivery approaches require investigation of any unintended implications. We aimed to assess whether adoption of an OC system caused a change in antibiotic prescribing rates in English general practice between March 2019-February 2022. Health policy interventions (including OC system adoption) are often implemented at a cluster level and rolled out to different groups over time. Their evaluation requires careful consideration of which control groups and periods to use in the analysis. Recent literature highlights that the standard two-way fixed effects difference-in-differences (DiD) estimator using all units and periods gives a biased estimate of the average treatment effect in the treated (ATT) when roll-out is staggered and there is treatment effect heterogeneity over time. We use a staggered DiD method that has since been proposed by Callaway and Sant'Anna (2021) that explicitly avoids using inappropriate units and periods as controls[1]. We used Callaway and Sant'Anna's doubly robust method to estimate the average effect of adoption for each group of practices (defined by year of adoption) in each year. We aggregated these group- time ATTs (GTATTs) across all adopting practices, by group, and by time elapsed since adoption. We found strong evidence of a positive effect of adoption on antibiotic prescribing rates, albeit the magnitude of effect was relatively small (overall ATT = 1.7 items prescribed per 1,000 patients per month, 95% CI=(1.1, 2.4)). As time since adoption increases, the effect size increases, and effects vary by adoption year. We used a recently proposed method to assess the parallel trends assumption, which is intrinsic in DiD methods [2]. Using Callaway and Sant'Anna's approach we aggregated GTATTs into an overall ATT, thus avoiding the issues surrounding two-way fixed effects regressions using all units and periods when roll-out is staggered, and highlighted treatment effect heterogeneity across different dimensions. The results were all consistent with adoption of an OC system increasing antibiotic prescribing rates, but further research is needed to determine whether any increase is appropriate.

[1] B. Callaway, P. H. C. Sant'Anna, Difference-in-Differences with multiple time periods, Journal of Econometrics, 225, 2021, 200-230.
[2] A. Rambachan, J. Roth, A more credible approach to parallel trends, The Review of Economic Studies, 2023, https://doi.org/10.1093/restud/rdad018.

MP15 Causal effect analysis on spasticity-plus outcomes in patients with multiple sclerosis

Greco T.*[1], Poole E.[2], Alexander J.[2]
[1]*Jazz Pharmaceuticals, Inc., Gentium Srl ~ Villa Guardia ~ Italy,* [2]*Jazz Pharmaceuticals, Inc. ~ Philadelphia ~ United States of America*

The causal inference framework is becoming appealing for discovering relationships in epidemiology studies and clinical trials. Recent discussions between neurologists and experts in multiple sclerosis management resulted in a consensus to group interconnected symptoms within the spasticity-plus syndrome [1]. This study aimed to investigate the causal relationship between changes in spasticity severity as measured on the 0-10 Numerical Rating Scale (NRS-S) and improvements in spasticity- associated symptoms in patients randomized in a double-blind placebo-controlled trial to evaluate the efficacy of Sativex in subjects with symptoms of spasticity due to MS [2]. This study collected all data needed for this investigation. Data on a total of 335 participants, n=166 in Sativex and n=169 in placebo, were analyzed to determine the marginal Average Exposure Effect (AEE) of the change in NRS-S, measured as both clinically meaningful or initial response criteria (≥30% or ≥20% reduction in NRS-S), on spasticity-plus outcomes, i.e., spasms, pain, bladder function, fatigue, and sleep, at the end of the treatment period (14 weeks after randomization). Potential measured confounding factors on exposure were considered by including baseline characteristics (treatment, age, gender, body mass index, previous cannabis use, and expanded disability status scale in the structural causal model: Inverse Probability Weight using the Propensity Score method (IPW-PS). Exchangeability, conditional exchangeability, consistency, and positivity assumptions were assessed. Logistic regression model was fitted to the pseudo-population recreated by the IPW-PS to regress the exposure on spasticity-plus outcomes by adjusting for baseline covariates. Box-plot investigations delineated homogeneous allocation of baseline characteristics by exposure groups. Distributions of quintile categorization of predicted PS values overlapped for each spasticity- plus outcome. The estimates for the AEE indicate that reaching a response (both clinically meaningful and initial) is effective in obtaining an average improvement in spasticity-plus symptoms at the end of the treatment period (all p-values <0.05). Causal Inference framework may be a powerful method to investigate dependencies and generate new hypotheses in clinical trial settings. This study outlined that a reduction in NRS spasticity has a statistically significant causal effect on spasticity-related symptoms. patients with resistant spasticity: Analysis in relation to the newly described 'spasticity-plus syndrome'. Eur J Neurol. 2022; 29(9):2744-2753.

[2] Collin C, Ehler E, Waberzinek G, et al. A double-blind, randomized, placebo-controlled, parallel- group study of Sativex, in subjects with symptoms of spasticity due to multiple sclerosis. Neurol Res. 2010;32(5):451-9.

MP16 The role of labor market inequalities in explaining the gender gap in depression risk among older us adults.

Gueltzow M.*[1], Bijlsma M.J.[2], Van Lenthe F.J.[3], Myrskylä M.[1]
[1]*Max Planck Institute for Demographic Research ~ Rostock ~ Germany,* [2]*Unit PharmacoTherapy, -Epidemiology, and - Economic (PTEE), Groningen Research Institute of Pharmacy, University of Groningen ~ Groningen ~ Netherlands,* [3]*Department of Public Health, Erasmus MC, University Medical Center Rotterdam ~ Rotterdam ~ Netherlands*

In the US, women are twice as likely to suffer from depression as men. We aim to investigate to what extent gender inequality in the labor market contributes to the gender gap in depression risk in older adults. We analyze data from 35,699 US adults aged 50-80 years that participated in the Health and Retirement Study. The gender gap is calculated as the difference in prevalence in elevated depressive symptoms (score >= 3 on the 8-item Center for Epidemiological Studies Depression Scale) between women and men. We employ a dynamic causal decomposition and simulate the life course of a synthetic cohort from ages 50-80 with the longitudinal g-formula and introduce four nested interventions by assigning women the same probabilities of A) being in an employment category, B) occupation class, C) current income and D) prior income group as men, conditional on women's health and family status until age 70. The gender gap in depression risk is 2.9%-points at ages 50-51 which increases to 7.6%-points at ages 70-71. Intervention A decreases the gender gap over ages 50-71 by 1.2%-points (95%CI for change: -2.81 to 0.4), intervention D by 1.64%-points (95%CI for change: - 3.28 to -0.15) or 32% (95%CI: 1.39 to 62.83), and the effects of interventions B and C are in between those of A and D. The impact is particularly large for Hispanics and low educated groups. Gender inequalities at the labor market substantially explain the gender gap in depression risk in older US adults. Reducing these inequalities has the potential to narrow the gender gap in depression.

MP17 Development of ensemble propensity score matching method

Hasegawa Y.*[1], Yui S.[1], Ban H.[1], Negishi S.[2], Kikuchi T.[2]
[1]*Hitachi, Ltd. Research & Development Group ~ Tokyo ~ Japan,* [2]*Hitachi Health Insurance Society ~ Tokyo ~ Japan*

Japanese health insurance societies provide various health services such as health checkup and health guidance for disease prevention and cost saving, and are required to analyze the effects of these services to promote "Data Health". However, since it is necessary to analyze the effects of health services using observational data, it is essential to adjust covariates between intervention and non-intervention groups. Recently, propensity score matching [1] has been used as a method for adjusting covariates, but the challenge is how to adjust covariates with high precision. Therefore, we proposed an ensemble propensity score matching method that combines multiple propensity score estimation methods to adjust covariates between intervention and non-intervention groups with high precision. Although logistic regression is usually used to estimate propensity scores, it is not always the optimal propensity score estimation method. Therefore, we proposed an ensemble propensity score matching method that combines multiple propensity score estimation methods to match intervention and non- intervention groups so as to minimize standardized differences in covariates. Propensity scores were estimated using (a) Logistic regression, (b) LASSO regression, (c) Elastic net, (d) Gradient boosting decision tree [2], and (e) Neural network. To evaluate the proposed method, we compared the mean standardized differences of all covariates for matching with the proposed method and for matching with each of the five propensity scores from (a) to (e). This evaluation was conducted by one-to-one matching without replacement using 3,574 intervention subjects and 9,668 non-intervention subjects in the health guidance provided by the Hitachi Health Insurance Society. A total of 31 items were used as covariates, including basic information such as gender and age, laboratory values such as weight, blood pressure, glucose metabolism, blood lipids, and liver function, and questionnaire items such as smoking, exercise, eating, drinking, and sleeping. The evaluation results showed that the proposed method minimized the mean of standardized differences in 31 covariates. It also improved the mean of the standardized differences by 76.3% compared to the conventional logistic regression.
It confirmed that the proposed method can adjust covariates with high precision and can be used to analyze the effects of health services.

[1] Rosenbaum PR, Rubin DB, The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika 70, 1983, 41–55.
[2] Chen T, et.al., Xgboost: A scalable tree boosting system, In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 785–794.

MP18 Instrumental variables: systematic review of strategies to justify validity in clinical studies of dementia

Hiu S.*[1], Yong T.[2], Hasoon J.[3], Teare M.D.[1], Taylor J.[3]
[1]*Population Health Sciences Institute, Newcastle University ~ Newcastle upon Tyne ~ United Kingdom,* [2]*Cumbria, Northumberland, Tyne and Wear NHS Foundation Trust ~ Sunderland ~ United Kingdom,* [3]*Translational and Clinical Research Institute, Newcastle University ~ Newcastle upon Tyne ~ United Kingdom*

Causal inference methods in studies of dementia and neurodegenerative are becoming more frequent with the availability of large real-world datasets. Although causal effects are of significant clinical interest, they often rest on often-unverifiable assumptions. Instrumental variable (IV) analysis is appealing because it allows adjusting for unmeasured confounding, but relies on three "core" assumptions for an IV to be valid: Relevance, Exclusion Restriction, and Exchangeability. This study systematically reviews the strategies that clinical studies of dementia and neurodegenerative disease have adopted to justify the validity of their IV (PROSPERO: CRD42023392589). We searched PubMed, PsycINFO, and Web of Science for observational studies using instrumental variable analysis in an adult population. Eligible studies included patients diagnosed with dementia or neurodegenerative disease, an outcome of dementia or neurodegenerative disease, or investigated the causal effect of dementia or neurodegenerative disease on a set of pre-defined outcomes. We excluded studies using Mendelian randomisation (genetic instruments). Text data were extracted and coded into categorical descriptors for reporting. Verification of the extracted data and descriptors by co-authors is still ongoing. We identified 12 eligible studies. Regarding Relevance, all but one study reported test statistics supporting an association between the IV and the exposure. 58% of studies provided additional support to the assumption by citing prior quantitative/qualitative studies demonstrating an association or a proposed mechanism by which the IV influences the exposure. Three falsification test strategies were observed across five studies. Regarding Exclusion Restriction, only 50% provided subject-matter justification; mainly through implausibility of other mechanisms by nature of the disease, the exposure, or the IV itself. 33% of studies performed at least one falsification test to probe the plausibility of the assumption; five falsification strategies emerged from the data. Regarding Exchangeability, falsification testing of this assumption was the most common at 75% of studies conducting at least one falsification test across eight strategies, but only 33% provided subject-matter justification for why the IV was unconfounded with the outcome. Justifying the validity of IVs remains challenging, particularly for the Exclusion Restriction and Exchangeability assumptions. Close collaborations between clinicians and statisticians are encouraged to help address the three assumptions.

[1] Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol. 2018;15:1.
[2] Labrecque J, Swanson SA. Understanding the Assumptions Underlying Instrumental Variable Analyses: a Brief Review of Falsification Strategies and Related Tools. Curr Epidemiol Rep. 2018;5:214-20.

MP19 Applying causal inference to clinical trial data to improve our understanding of how car-t cell therapies work

Newsome S.*[1], Cao M.[2], Dunn R.[2], James D.[2], Ma W.[2], Rodosthenous T.[3], Schindl K.[4], Thivierge G.[4], Greenhouse J.[4], Jones J.[1]
[1]*Novartis Pharma AG ~ Basel ~ Switzerland,* [2]*Novartis Pharmaceutical Corporation ~ East Hanover, NJ ~ United States of America,* [3]*Novartis Pharmaceuticals UK Ltd ~ London ~ United Kingdom,* [4]*Carnegie Mellon University ~ Pittsburgh, PA ~ United States of America*

CAR-T cell therapy is a treatment for certain types of blood cancers. The starting material of the treatment is a patient's own white blood cells, which are collected to manufacture the CAR-T product, which is reinfused into the patient, where it can then multiply and attack the cancer cells. Clinical trials have demonstrated the overall efficacy and safety of CAR-T cell therapies, but these clinical trials also provide us with a wealth of data, which could help us to better understand the mechanisms of CAR-T therapy. These exploratory analyses present several challenges, as the clinical trials were generally not designed to answer these questions and as such when performing these analyses, we need to be cognizant of the potential for confounding and other sources of bias, which could impact the validity of such exploratory analyses. Our ongoing work uses tools from causal inference to examine and account for potential sources of bias that may influence the target effect of interest. The issue of confounding is particularly important in CAR-T therapy, since a patient's own white blood cells serve as the starting material to manufacture the CAR-T product. This means for example that the patient characteristics at the time of blood draw may confound the relationships between the final manufactured product and patient outcomes. A crucial component of this work is the use of directed acyclic graphs (DAGs) to help visualize and understand the complexity of the causal relationships within the data, and to help guide decisions on appropriate analyses targeting causal effects. We also explore the use of causal discovery as a method to help understand the underlying causal structures in the data. This work is part of the collaboration between Novartis and Carnegie Mellon University to understand CAR-T therapy through the lens of principled, flexible, and modern statistical tools. Through this work we demonstrate how causal inference is a powerful tool to harness clinical trial data, in order to understand the mechanisms of CAR-T therapy more deeply.

MP20 Unbiased and efficient estimation of causal treatment effects in cross-over trials

Halkjaer J.[1], Scheike T.[1], Pipper C.*[2]
[1]*Dept of Public Health, University of Copenhapgen ~ Copenhagen ~ Denmark,* [2]*Novo Nordisk A/S ~ Soeborg ~ Denmark*

We introduce causal inference reasoning to cross-over trials, with a focus on Thorough QT (TQT) studies.
For such trials, we propose different sets of assumptions and consider their impact on the modelling strategy and estimation procedure. We show that unbiased estimates of a causal treatment effect are obtained by a g-computation approach in combination with weighted least squares predictions from a working regression model. Only a few natural requirements on the working regression and weighting matrix are needed for the result to hold. It follows that a large class of Gaussian linear mixed working models lead to unbiased estimates of a causal treatment effect, even if they do not capture the true data generating mechanism. We compare a range of working regression models in a simulation study where data are simulated from a complex data generating mechanism with input parameters estimated on a real TQT data set. In this setting, we find that for all practical purposes working models adjusting for baseline QTc measurements have comparable performance. Specifically, this is observed for working models that are by default too simplistic to capture the true data generating mechanism. Cross-over trials and particularly TQT studies can be analysed efficiently using simple working regression models without biasing the estimates for the causal parameters of interest.

MP21 Combining datasets in a mendelian randomization design: metabolomics and parkinson's disease case study

Pontali G.*[1], Filosi M.[1], Borsche M.[2], König I.R.[3], Paglia G.[4], Del Greco M F.[1]
[1]*Institute for Biomedicine, Eurac Research ~ Bolzano ~ Italy,* [2]*Institute of Neurogenetics, University of Lübeck and University Hospital Schleswig-Holstein ~ Lübeck ~ Germany,* [3]*Institute of Medical Biometry and Statistics, University of Lübeck ~ Lübeck ~ Germany,* [4]*School of Medicine and Surgery, University of Milano-Bicocca ~ Milano ~ Italy*

Mendelian randomization (MR) is a statistical method that allows to investigate causal pathways from modifiable exposures to disease outcomes, using genetic variants within the instrumental variable setting. Thanks to the increasing availability of summary genetic data from huge meta-analyses, two- sample MR studies are widely performed to infer causal hypotheses. Multiple data sources on the exposure could be available, characterized by the presence of many differences in study design (e.g. on the measurement procedure; the type of blood sample; the study ancestry; the data transformation). The purpose of this work is to combine different data sources within an MR framework. We explore empirically a new workflow based on a discovery-replication-validation design, comparing it with the most common one, where only one MR analysis is performed using the largest dataset available. The robustness of the obtained causal evidence is evaluated. Following a hypotheses-free approach, a metabolome-wide MR is carried out including different genetic data sets on around 186 targeted metabolites, exposures measured with the same analytical technique (LC-MS Biocrates kit). After a literature review, genetic association studies with at least 5,000 participants are used: three genome-wide association studies (GWAS) with European ancestry – Draisma[1] and CHRIS (not published) on serum blood samples from the general population, and Lotta[2] on plasma samples from blood donors; and one whole-exome sequencing study (WES) – CHRIS[3]. The largest dataset is used for the Parkinson's disease outcome[4]. In the classical design, the instruments are selected once on the meta-analysis of Draisma-CHRIS. With the proposed design, that selection is repeated using each dataset independently. After the instruments' selection, different metabolites are analysed. The results show that the repetition of the analysis on different data allows an exposure's optimization by checking findings' consistency, leading to the identification of the strongest causal pathway, and opening towards new possible biological mechanisms. Reliable causal associations were pointed out with our design proposal when multiple datasets are available. The instruments' selection from WES allows exploring pros and cons of using them instead of those obtained from GWAS, making an innovative contribution to the instruments' choice in MR.

[1] Draisma HHM, Pool R, Kobl M, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015;6:7208
[2] Lotta LA, Pietzner M, Stewart ID, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. Nat Genet. 2021;53(1):54-64
[3] König E, Rainer J, Hernandes VV, et al. Whole Exome Sequencing Enhanced Imputation Identifies 85 Metabolite Associations in the Alpine CHRIS Cohort. Metabolites. 2022;12(7):604
[4] Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet I. 2019;18(12):1091-1102

MP22 Trial emulation for group policies: impact of online consultation system on emergency hospital utilisation

Xu Y.*[3], Keogh R.[3], Clarke G.[1], Diaz--Ordaz K.[2]
[1]*Improvement Analytics Unit, The Health Foundation ~ London ~ United Kingdom,* [2]*Department of Statistical Science, UCL London ~ United Kingdom,* [3]*Department of Medical Statistics, London School of Hygiene and Tropical Medicine ~ London United Kingdom*

Policymakers are interested in evaluating the causal effects of health policies in primary or secondary care settings. Such policies are often introduced without having first conducted a randomised study and their evaluation typically relies on observational data. Our aim is to use the target trial emulation framework to estimate the causal effect of an online consultation system (eConsult) on emergency care utilisation in England. Online consultation systems have been introduced in general practices to improve patient access and enable efficient triage. The trial emulation framework has been successfully used in many studies to estimate individual-level treatment effects. However, it has not yet been widely adopted for policy evaluation. Policies are typically implemented at the group level, which introduces challenges as the data structure and the causal estimands of interest are hierarchical. We emulate a clustered encouragement trial, using data on 232 practices that adopted eConsult and 812 control practices. Follow-up was for 12 months from January 2019. Our outcome is the monthly rate of emergency department (ED) visits per 1000 patients per practice. We conduct an intention-to- treat analysis to estimate the average treatment effect of initiating eConsult, and a per-protocol analysis to estimate the longitudinal effect of sustained use of eConsult versus non-use. We implement several estimation methods, including targeted maximum likelihood estimation, inverse probability treatment weighting with marginal structural models and g-computation. To attenuate biases due to model misspecification, we explore the use of SuperLearner. We account for clustering by using cluster-level analysis of aggregated data combined with minimum-variance weights in the variance estimators. We find no significant evidence that the adoption of eConsult for up to 12 months leads to a change in ED visits at the practice level. Minimum-variance weight has minimal impact on point estimates and variances. Our work provides a practical example of implementing the trial emulation framework and a range of causal methods for policy evaluation. Limitations of our findings are that usage of eConsult among the practices adopting the system during 2019 was low. Future work could investigate later periods and take into account usage levels.

[1] Hernán M A, Robins J M. Using big data to emulate a target trial when a randomized trial is not available[J]. American journal of epidemiology, 2016, 183(8): 758-764.
[2] van der Laan M J, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions[J]. The international journal of biostatistics, 2012, 8(1).

MP23 Accounting for non-compliance when estimating treatment effects of ordinal outcome in rct

Zhu J.*, Li J., Tai B.C.
National University of Singapore ~ Singapore ~ Singapore

In randomized clinical trials (RCTs) with non-compliance, evaluating the causal effects of interventions would lead to a more precise estimation of treatment effect. The intention-to-treat (ITT) method is commonly used in RCTs even though there may be substantial non-compliance. However, it may lead to a dilution in estimating treatment effect. While there is a large body of literature addressing the issue of non-compliance for continuous or time-to-event outcome, this issue is seldom discussed for ordinal outcomes. In this paper, we propose an extension of the inverse probability weighting method for handling non-compliance involving an ordinal outcome by fully utilizing the information of non-compliance and defining it as a categorical variable to describe the extent of non-compliance. This is in contrast to the usual convention where compliance is regarded as a binary variable. We compare our proposed method for estimating causal log odds ratio with four commonly used methods (ITT, per protocol, instrumental variable and inverse probability weighting) in simulation studies. The results demonstrate that the proposed method performs well in terms of bias, variance, coverage, power and Type I error under various scenarios, especially when there is selection bias. The methods are also compared using data from the JOBS II intervention trials to assess the causal effects. In conclusion, we recommend using IPW methods for ordinal outcome in the presence of non- compliance. While there exists selection bias, our proposed method performs better than ITT, PP and IV method. The proposed method has similar performance with IPW method, but provides more useful information than IPW method. Bang, H. and Davis, C. E. (2007). On estimating treatment eects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? Statistics in Medicine 26(5), 954{964. Boatman, J. A., Vock, D. M., Koopmeiners, J. S. and Donny, E. C. (2018). Estimating causal eects from a randomized clinical trial when noncompliance is measured with error. Biostatistics 19(1), 103{118.

MP24 Risk factors for the survival of colorectal cancer patients modeled through a competitive risk analysis.

Aguirre Larracoechea U.*[2], Gascon Perez M.[2], Ruiz Castro J.E.[1], Quintana Lopez J.M.[2], Legarreta Olabarrieta M.J.[2], Portillo Villares I.[3], Audicana Uriarte C.[4], Valverde Garcia M.A.[2], Muñoz Fernandez C.[2]
[1]*Departament of Statistics and Operational Research. University of Granada ~ Granada ~ Spain,* [2]*Research Unit, Osakidetza Basque Health Service, Barrualde-Galdakao Integrated Health Organisation, Galdakao-Usansolo Hospital. Kronikgune Institute for Health Services Research. Network for Research on Chronicity, Primary Care, and Health Promotion (RICA,* [3]*Colorectal Cancer Prevention Programme of the Basque Country. Basque Health Service-Osakidetza ~ Bilbao ~ Spain,* [4]*Basque Health Department ~ Vitoria-Gasteiz ~ Spain*

Colorectal cancer (CRC) is currently among the most frequent cancers in both women and men. The development of clinical prediction rules to predict adverse events (e.g, mortality) after surgical treatment of patients with colorectal cancer is important. Survival analysis is the modelling of time until the occurrence of mortality in order to assess the effects of several effects on survival time. Nevertheless, there are situations in which a subject may encounter one of several distinct categories of occurrences. In this situation, the competing-risks model is the optimal method for analysing survival time. Competing-risk analysis, which complements traditional survival analysis, helps evaluate the risk of a specific endpoint when accompanied by the competing events. The study aimed to identify the risk factors that predict 5-year specific mortality due to colorectal cancer or mortality from other causes. A total of 898 patients were recruited from public hospitals belonging to the Basque Health Service (Osakidetza) with a diagnosis of CRC and who underwent surgery between June 2010 and December 2012. They were followed up at different measure times up to five years after the intervention. A competing risks analysis was performed to predict colorectal cancer-specific and other causes- mortality, assessing a survival model for each cause. The findings of our study showed that 64% of the participants were men with an overall mean age of 68 years. An overall of 254 deaths were observed, of which 173 were due to colorectal cancer and 81 to other causes. The results for colorectal cancer-specific survival showed the following risk factors: alcoholism, age, ASA risk, outcome of surgery, invasion of adjacent organs, pTNM scale, leukocyte level and presence of anemia. As for the model for mortality from other causes, alcoholism, age, outcome of surgery and anemia were maintained as predictors, adding the variables of the Charlson Comorbidity Index, treatment with chemotherapy and radiotherapy and the PLR (platelet-lymphocyte ratio). Both models obtained a good predictive capacity with a c-index of 0.785 and 0.82 respectively. There are differences in risk predictors for colorectal cancer specific and other causes mortality.

[1] J. Beyersmann, A. Allignol, M. Schumacher, Competing Risks and Multistate Models with R, Springer 2012.
[2]D.G. Kleinbaum, M. Klein, Survival Analysis, Springer 2005.

MP25 Ms-cpfi: a model-agnostic interpretable counterfactual perturbation feature importance for multi-state models

Cottin A.*[1], Zulian M.[1], Guilloux A.[2], Katsahian S.[2]
[1]*Healthcare and Life Sciences Research, Dassault Systemes ~ Paris ~ France,* [2]*HeKa team, INRIA ~ Paris ~ France*

In healthcare, clinicians are interested in modeling the evolution of patients' disease through different states by using a multi-state process [2]. Statistical approaches as the extensions of the Cox model were the most popular methods; they are easily interpretable as they assume linear relationships between features and clinical risks. Unfortunately, this relationship can be more complex in some clinical situations. Machine learning algorithms have received a great attention in the recent years for addressing this limitation. The lack of interpretability of these models motivated the development of post-hoc (ie. model-agnostic) interpretability methods. This issue is even more critical in a clinical setting as explaining predictions contributes making the model into a medical decision support system. While methods for interpreting a survival black-block model exist, there is no method for a black-box multi-state model. The main component of interpretability for black box models is to assign a statistic of importance for each feature. The most intuitive method is Permutation Feature Importance [1]. Unfortunately, it faces some limitations and should be adapted for a multi-state model. We propose a novel model-agnostic interpretability algorithm called \textit{Multi-State Counterfactual Perturbation Feature Importance} (MS-CPFI) that compute variable importance for each transition of a multi-state model. It ranks features by their importance and provides a list of significant protective or risk factors for each transition. It computes a prediction-based variable importance score, and use a new counterfactual perturbation method that allow interpreting features effects while capturing the non-linearities. Experimental results on simulations display that MS-CPFI increases model interpretability in the case of non-linearities. Results on the METABRIC cohort confirm that MS-CPFI can detect clinically important features and can provide information on the disease progression by displaying features that are protective factors versus features that are risk factors for each stage of the disease. Explaining predictions in deep neural networks is challenging but essential in clinical applications where interpretability is evaluated to support medical decisions. Development of MS-CPFI could strongly improve patient prognosis and understanding of disease progression.

[1] Leo Breiman. Random forests. Machine learning, 45(1):5{32, 2001.
[2] Anthony J Webster. Multi-stage models for the failure of complex systems, cascading disasters, and the onset of disease. PloS one, 14(5):e0216422, 2019.

MP26 Flexible log-hazard model for adverse events analysis

Coz E.*, Fauvernier M., Maucort--Boulch D.
Université de Lyon, Hospices Civils de Lyon, Pôle Santé Publique, Service de Biostatistique et Bioinformatique, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé ~ Lyon ~ France

Over the last few years, several methodological gaps have been reported regarding the analysis and reporting of adverse events (AE) in clinical trials. Recent recommendations include the presentation of both absolute and relative risks, the use of survival models and to account for competing risks. In this study, we evaluate the relevance of a flexible penalized hazard models [1] (FPHM) for the analysis of adverse events in a competing risk setting to describe the occurrence of AEs and to provide patient comparisons regarding their toxicities. We applied the approach on simulated datasets that mimic real-life safety data observed in oncology. The analysis were compared with those provided by more popular models (i.e. Cox and Fine-Gray models and logistic regression). The simulations considered proportional and non-proportional hazards regarding a covariate of interest, as well as non-informative censoring. We used the model to described dermatologic adverse events (DAEs) in cancer patients treated with immunotherapy according to a well-known cancer prognostic factor: the neutrophil to lymphocyte ratio (NLR) [2]. The FPHM was able to describe the hazards correctly, even with medium sample size and number of events. Predictions of the cumulative probabilities of AEs were similar or better than those stemming from other models, particularly in the non-proportional scenario. Hazard ratios showed an important variability, particularly with low numbers of events. However, they described a non-constant effect of the treatment while those from the Cox model are an average effect over the follow-up duration. In real-data application, the model described non-linear and non-proportional effect of the NLR on the hazards of death or treatment discontinuation. Regarding adverse events, both DAEs hazards and cumulative probabilities seems higher for patients with lower NLR. The FPHM seems relevant to describe the occurrence of AEs over time and compare non-recurrent AEs (e.g. those leading to treatment discontinuation). Combined with cumulative probabilities, the approach offers an accurate interpretation of the effects of the variables in the tricky competing risks setting.

[1] M. Fauvernier, L. Roche, Z. Uhry, L. Tron, N. Bossard, L. Remontet, and the Challenges in the Estimation of Net Survival Working Survival Group, Multi-dimensional penalized hazard model with continuous covariates: applications for studying trends and social inequalities in cancer survival, J. R. Stat. Soc. Ser. C Appl. Stat. 68 (2019) 1233–1257.
[2] P.Y. Lee, K.Q.X. Oen, G.R.S. Lim, J.L. Hartono, M. Muthiah, D.Q. Huang, F.S.W. Teo, A.Y. Li, A. Mak, N.S. Chandran, C.L. Tan, P. Yang, E.S. Tai, K.W.P. Ng, J. Vijayan, Y.C. Chan, L.L. Tan, M.B.-H. Lee, H.R. Chua, W.Z. Hong, E.S. Yap, D.K. Lim, Y.S. Yuen, Y.H. Chan, F. Aminkeng, A.S.C. Wong, Y. Huang, S.H. Tay, Neutrophil-to-Lymphocyte Ratio Predicts Development of Immune-Related Adverse Events and Outcomes from Immune Checkpoint Blockade: A Case-Control Study, Cancers. 13 (2021) 1308.

MP27 Sample size determination based on restricted mean time lost in the presence of competing risks

Geng X.*, Li Z., Zhang C., Chen Z.
Southern Medical University ~ Guangzhou ~ China

The calculation of sample size is important in the design of clinical trials, and the presence of competing risks makes the design and statistical analysis of clinical trials with time-to-event endpoints more cumbersome. Restricted mean time Lost (RMTL) is not affected by the proportional hazards assumption, and the indicator itself is easy to understand and interpret, so it can be used as an alternative method for experimental design and analysis based on the cause-specific hazard (CSH) and subdistribution hazard (SDH) models. The sample size calculation methods in published RMTL studies did not take into account the relevant factors of clinical trials, such as accrual period and censoring, so we propose a new RMTL-based sample size determination method from the perspective of clinical trials. The Monte Carlo simulation is used to correct the impact of censoring on the population variance under the Weibull distribution, duration of accrual and follow-up were set, and the sample size was calculated by considering the experimental conditions such as equal allocation, uniform accrual, uniform censoring, and administrative censoring. The simulation results show that the sample size calculation results of RMTL based on Weibull distribution can approximately achieve the predefined power level and are relatively stable, and are similar to or even better than the sample size and power calculated based on subdistribution hazard ratio (SHR). Even if the event time does not obey the Weibull distribution, the RMTL sample size calculation method based on the Weibull distribution still performs well. The result of example analysis also validates the performance of sample size determination based on RMTL.
From the perspective of the results of this study, clinical interpretation, application conditions and statistical performance, we recommend that when designing clinical trials in the presence of competing risks, the RMTL indicator can be used for sample size calculation and subsequent effect size measurement.

[1] Lyu J, Hou Y, Chen Z. The use of restricted mean time lost under competing risks data. BMC Med Res Methodol. 2020;20(1):197.
[2]Wu H, Yuan H, Yang Z, Hou Y, Chen Z. Implementation of an Alternative Method for Assessing Competing Risks: Restricted Mean Time Lost. American Journal of Epidemiology. 2022;191(1):163- 172.

Poster Sessions

MP28 Handling missing disease information in diseases that need two visits to diagnose

Le Thi Phuong T.*[1], Wolfe R.[1], Heritier S.[1], Geskus R.[2]
1School of Public Health and Preventive Medicine, Monash University ~ Melbourne ~ Australia, 2Oxford University Clinical Research Unit ~ Ho Chi Minh city ~ Viet nam

In studies with interval-censored time to disease onset, participants who die without an observed disease will typically have unknown disease status after their last visit. Failing to properly account for the probability of developing the disease between the last observed disease-free visit and death can result in a biased estimate of the effect of a variable on disease risk [1,2]. We consider an additional scenario where two consecutive positive tests are needed for diagnosis, such as with persistent physical disability endpoint; In that case, disease status can also be unknown at the last visit preceding death. We show that this impacts the choice of censoring time for those who die apparently disease-free. We investigate two classes of model that quantify the effect of risk factors on disease outcome and four censoring strategies. The first is a Cox proportional hazards model with death as a competing risk. The second is an illness death model that treats disease as a possible intermediate state. For censoring strategies, participants without observed disease are censored at: death (Cox model only), the last visit, the last visit with a negative test, or the second last visit. We evaluate the performance investigated approaches on simulated data with a binary risk factor under different settings for mortality rates, risk factor effects, length of visit intervals, and frequency of false positive tests. We found that the illness death model with the second last visit as the last disease-free observation shows the best performance in all simulation settings. Other methods show bias that varies in magnitude and direction depending on the differential mortality in diseased subjects, the gap between visits, and the choice of censoring time. We illustrate these approaches in a study of 19,114 elderly participants estimating the effect of diabetes on persistent physical disability. We reconfirm the recommendation [1,2] that the illness death model should be used to analyse data that is susceptible to missing disease information due to death. In diseases requiring two consecutive positive tests for diagnosis, the second last visit should be used as the last disease-free observation.

[1] B. Nadine, et al., Journal of Clinical Epidemiology, 105, 2019, 68-79.
[2] L. Karen, et al., International Journal of Epidemiology, 42.4, 2013,1177-1186.

MP29 Using multistate models to explore sociodemographic inequalities in the risk of adverse pregnancy outcomes

Mcgovern M.*, Seaton S., Smith L.
University of Leicester ~ Leicester ~ United Kingdom

During pregnancy and the first month after birth, babies are at the highest risk of multiple adverse outcomes. Current methodologies to understand the impact of sociodemographic factors on the risk of these outcomes explore these endpoints independently but do not account for the changing risk and impact of risk factors. Conceptualising pregnancy and the first month after birth as a multistate process with multiple sets of competing risks offers the opportunity to analyse the time-to-event of a number of end points, allowing us to simultaneously capture the risk of different outcomes and assess the varying impact of sociodemographic factors. A flexible parametric multistate model describing the period from 22 weeks of pregnancy to the first 28 days after birth was built adjusting for sociodemographic and clinical factors, using national UK surveillance data of over 500,000 births in 2020. The model comprised transient states (states you can leave e.g., in utero) and absorbing states (states you cannot leave e.g., death) and a baby's pathway through the model was dependent on their outcome. Preliminary results show that the risk of death varied throughout pregnancy and after birth, and also with sociodemographic and clinical factors. For example, babies in the most deprived group were at increased risk compared to those in the least deprived group for antepartum stillbirth (death before labour begins) (HR: 1.62, 95% CI: [1.40, 1.86], p-value<0.001) and death during the early neonatal period (first 7 days following birth) (HR: 1.47, 95% CI: [1.14,1.89], p-value<0.01) but there was no significant association with the risk of an intrapartum stillbirth (death during labour) or death during the late neonatal period (between 7-28 days following birth). Further analysis to investigate time-dependent effects is ongoing; it is expected that the effect of gestational age and birthweight are likely to vary within particular transitions. This application of multistate models in exploring adverse pregnancy outcomes is innovative and has not been previously fully exploited. The new information gained from these analyses is useful for both clinicians and policy makers in identifying those babies most at risk and developing targeted interventions to improve outcomes.

MP30 How rheumatoid arthritis disease activity affects the risk of cardiovascular multimorbidity

Rengger L.*[1], Lewin A.[1], Macgregor A.[2], Dainty J.[2]
[1]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [2]University of East Anglia ~ Norwich ~ United Kingdom

The positive relationship between Rheumatoid Arthritis (RA) and Cardiovascular Diseases (CVDs) is well evidenced[1]. It is hypothesised that chronic inflammation, intrinsic to RA, is concomitant with increased risk of CVD. Less widely understood is the dose-response relationship between the level of inflammation and CVD risk, and how increased disease activity accelerates an individual's trajectory to multiple cardiovascular morbidities. The Norfolk Arthritis Register (NOAR) cohort study began recruiting in 1989. Nurse-led clinical assessments followed individuals with RA longitudinally. Markers of RA severity were repeatedly taken including swollen and tender joint counts and erythrocyte sedimentation rate (ESR), a biochemical measure of systemic inflammation; these three measures were used to calculate a validated time- updating disease activity score (DAS28). Comorbidities, including CVDs: angina, heart attack, heart failure, hypertension and stroke, were also recorded. Assuming the underlying process is Markovian, a time-inhomogeneous multistate model was fitted with 4 states: 0 CVDs, 1 CVD, 2+ CVDs and dead. Hazard ratios between DAS28 and CVD state transitions were estimated, adjusting for age and gender. A total of 10,031 clinical assessments from 2,171 patients were included (66.7% female), followed over a median of 5 years, with mean age at RA onset and study entry 53.2 and 58.5 years respectively. After adjustment, there is strong evidence that increased disease severity is associated to faster transitions from no CVDs to 1 CVD (HR(95%CI)=1.19(1.08,1.31), p<0.001); a unit increase in DAS28 raises the risk of a first CVD event by approximately 19%. There is significant evidence that after experiencing 1 CVD, higher disease activity is associated to faster transition rates to CVD multimorbidity (2+CVDs); (HR(95%CI)=1.36(1.12,1.66), p=0.002). Associations persist when adjusting for potential confounders baseline BMI and smoking. There is strong evidence of a dose-response relationship between RA disease activity and CVD risk. Higher disease severity is associated to a faster trajectory from no CVDs to cardiovascular multimorbidity. Conclusions are consistent with other studies suggesting inflammation drives elevated CVD risk[1]. Inferences demonstrate the importance of managing traditional CVD risk factors amongst patients with RA, concurrently supporting present guidelines which focus on treating RA to low disease activity or remission[2].

[1] D. H. Solomon et al, "Disease activity in rheumatoid arthritis and the risk of cardiovascular events.", Arthritis & Rheumatology, Volume 67.6 (2015): pp 1449-1455.
[2] L. Fraenkel et al, "2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis.", Arthritis & Rheumatology, Volume 73.7 (2021): pp 1108-1123.

MP31 Restricted mean time lost model for covariates with time-varying effects

Yu Z.*[1], Li Z.[1], Zhang C.[1], Hou Y.[2], Chen Z.[1]
[1]Southern Medical University ~ Guangzhou ~ China, [2]Jinan University ~ Guangzhou ~ China

Elderly patients with breast cancer often die from other diseases, and for studies that focus on breast cancer, a competing risks model is more appropriate. In this paper, the effects of prognostic factors were quantified by an absolute indicator, the difference in restricted mean time lost (RMTL), which is more intuitive relative to the hazard ratio. In long-term follow-up, prognostic factors of breast cancer tend to have time-varying effects, but existing RMTL regression models cannot yet handle these covariates with time-varying effects. So, we aim to develop a new model. We proposed a time-varying effect RMTL regression to handle this, which can explore the between- group cumulative difference in mean life lost over a period of time and obtain the real-time effect by the speed of accumulation, as well as personalized predictions on a time scale. A simulation validated the accuracy of the coefficient estimates in the proposed regression. Applying this model to an elderly early-stage breast cancer cohort, it was found that 1) the protective effects of positive estrogen receptor and chemotherapy decreased over time; 2) the protective effect of breast-conserving surgery increased over time; and 3) the deleterious effects of stage T2, stage N2, and histologic grade II cancer increased over time. Moreover, from the view of prediction, the mean C-index in external validation reached 0.78. Time-varying effect RMTL regression can analyze both time-varying cumulative effects and real-time effects of covariates, which can provide a more comprehensive prognosis and better prediction.

1. Andersen PK. Decomposition of number of life years lost according to causes of death. Stat Med 2013; 32: 5278–5285.
2 .Conner SC, Trinquart L. Estimation and modeling of the restricted mean time lost in the presence of competing risks. Stat Med 2021; 40: 2177–2196.

MP32 Covid-19 infection in children with acute lymphoblastic leukemia in china: mild clinical course

Cai J.*[1], Li C.[2]
[1]Shanghai Children's Medical Center ~ Shanghai ~ China, [2]Hong Kong Children's Hospital, The Chinese University of Hong Kong, Hong Kong SAR ~ Hong Kong ~ China

Children treated for cancer are at increased risk to develop severe illness after COVID-19 infection. Since December 2022, there was a surge of COVID-19 infection across the country in China due to a change of isolation policy. This study aimed to report the clinical course of children with acute lymphoblastic leukaemia (ALL) treated by uniform chemotherapy protocols after COVID-19 infection in China. Centers participating in the Chinese Children Cancer Group (CCCG) ALL studies were invited to participate in the study. Eligible subjects were children treated by CCCG ALL studies with a diagnosis of COVID-19 infection, either by RT-PCR or Rapid Antigen Test. The demographic data, chemotherapy details, clinical features and clinical course of COVID-19 infection were collected by healthcare workers and entered into electronic system. From December 1, 2022 to January 31, 2023, 675 children with ALL were reported to have COVID-19 infection. The median age was 5.1 years (range 0.5- 17.1). Of these, 184 patients (27.3%) had received at least two doses of COVID-19 vaccine prior to infection. Majority of patients were still on chemotherapy and 94 patients (13.9%) had completed chemotherapy within one year. The commonest symptoms were fever (78.8%) and cough (57.2%). After the infection, 20.7% did not require any treatment. Only 35 patients (6.8%) were hospitalized, 11 patients required oxygen supplement and five patients needed intensive care. One patient with refractory leukaemia died after COVID-19 infections. This wave of COVID-19 infection due to Omicron variant in China caused mild illness in great majority of children treated with chemotherapy for ALL.

[1] Baruchel A, Bertrand Y, Boissel N, et al; SFCE Leukemia Committee. COVID-19 and acute lymphoblastic leukemias of children and adolescents: First recommendations of the Leukemia committee of the French Society for the fight against Cancers and Leukemias in children and adolescents (SFCE). Bull Cancer. 2020;107:629-632.
[2] Rouger-Gaudichon J, Bertrand Y, Boissel N, et al; Leukemia Committee of the SFCE. COVID19 and acute lymphoblastic leukemias of children and adolescents: Updated recommendations (Version 2) of the Leukemia Committee of the French Society for the fight against Cancers and leukemias in children and adolescents (SFCE). Bull Cancer. 2021;108:490-500.

MP33 Determinants of vaccine intention against covid-19: a serial mediation approach

Dardenne N.*, Paridans M., Pétré B., Guillaume M., Donneau A.
University of Liège ~ Liège ~ Belgium

Mediation analyses are increasingly used in the health field to study behaviour change, with serial mediation analysis being appropriate when strong causal relationships exist between mediators. The aim of this exploratory study is to develop a serial mediation model dealing with latent variables to assess direct and indirect effects of the 6 Health Belief Model (HBM)(1) determinants; perceived susceptibility, severity, benefits, barriers, self-efficacy and cue to action; on COVID-19 vaccine intention. A questionnaire on vaccine intention was administered to staff and students at the University of Liège (Belgium) in the anti-SARS-CoV-2 (COVID-19) seroprevalence study (SARSSURV)(2) from April 2021 to June 2021. The sample consisted of 1256 participants. To evaluate direct and indirect effects of the HBM latent variables on vaccine intention (score 0-100), serial mediation models for each latent variable permutation were assessed by means Structural Equation Modeling (SEM) using Partial Least Squares Path Modeling (PLS-PM) with the R package SeminR(3). Bayesian information criterion (BIC) was used to compare models. Internal consistency reliability were evaluated using Rhoc and Cronbach's alpha while convergent and discriminant validity by average variance extract (AVE) and heterotrait-monotrait ratio (HTMT). Significance of effect was tested by boostrapping. Sociodemographic variables, health literacy, psychological profile, body mass index, chronic disease and previous COVID-19 infection were included in the models as confounding factors. After running all permutation chains, the final causal chain, with the lowest BIC value, was cue to action (exposure) – severity – self-efficacy – barriers – susceptibility – Vaccine intention (outcome) (R^2 = 0.527). This highlighted a significant indirect effect between cue to action and vaccine intention through the other HBM determinants. A direct significant effect was also found between cue to action and vaccine intention. Values for the Cronbach's alpha, Rhoc AVE and HTMT met the validity and reliability criteria's. The determinant benefits were removed due to no significant path and weak reliability. Non-significant confounding factors were removed. SEM using PLS-PM seems to highlight a serial mediation model to explain relationships between HBM determinants and their effects on vaccine intention. However, additional investigations about the mediation/moderation effects between these determinants must be carried out.

1. Champion VL, Skinner CS, Glanz K, Rimer BK, Viswanath K. Health behavior and health education. Theory, Res Pract. 2008;45–65.
2. Donneau AF, Guillaume M, Bours V, Dandoy M, Darcis G, Desmecht D, et al. University population- based prospective cohort study of SARS-CoV-2 infection and immunity (SARSSURV-ULiège): a study protocol. BMJ Open [Internet]. 2022 Jan 1 [cited 2023 Feb 22];12(1):e055721. Available from: https://bmjopen.bmj.com/content/12/1/e055721
3. Hair JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S. The SEMinR Package. 2021 [cited ;49–74. Available from: https://link.springer.com/chapter/10.1007/978-3-030-80519-7_3

MP34 Social isolation and mental health of older citizens during the covid-19 pandemic

Ljunggren M.[1], Magnúsdóttir I.[2], González--Hijón J.[1], Valdimarsdóttir U.[2], Fang F.[1], Lovik A.*[1]
[1]Karolinska Institutet ~ Solna ~ Sweden, [2]University of Iceland ~ Reykjavík ~ Iceland

Older age is a risk factor for COVID-19-related morbidity and mortality. In many countries, stricter rules or recommendations applied to older citizens, which may have resulted in worse mental health in this population. The aim of the study is to describe the mental health impact of the COVID-19 pandemic on older citizens (60-94 years old) in Iceland and Sweden, how mental health is associated with the type and frequency of social contact, and how these relationships changed during the course of the pandemic. Methods. We included 16 611 individuals from the COVIDMENT consortium: 9027 from the Icelandic C-19 Resilience cohort and 7584 from the Swedish Omtanke2020 Study. Both had 3 data collections between April 2020 and February 2022. We measured depressive symptoms with the 9-item the Patient Health Questionnaire and anxiety symptoms with the 7-item Generalized Anxiety Disorder Scale using a cut-off at 10. The type of social contact was either in-person or virtual (phone, social media) and frequency was assessed with a four-point Likert scale. Moreover, demographic, lifestyle and health information was adjusted for in the analyses. All analyses were stratified for age group as well as working status. We estimated the prevalence of mental health symptoms using modified Poisson regression (baseline) and generalised estimating equations (longitudinal analysis). Data from Iceland and Sweden was combined using a meta-analytic approach. *Results*. In the pooled analysis, we found a dose-response relationship over the entire study period between the frequency of contact and the prevalence of depressive symptoms both for virtual (low: 19% (95%CI: 16%-22%), high: 12% (95 CI: 11%-14%)) and in-person (low: 18% (95%CI: 14%-22%), high: 11% (95%CI: 9%-13%)) contact. This same trend was seen with anxiety (virtual: low: 11% (95%CI: 8%-13%), high: 7% (95%CI: 5%-8%), in-person: low: 10% (95%CI: 8%-13%), high: 6% (95%CI: 5%-7%)). Similar patterns were visible for both in-person and virtual contact, and for the two countries separately. Lower frequency of social contact was associated with a higher prevalence of mental health symptoms in older adults over a 22-month study period.

MP35 Covid-19 health interventions targeting migrant populations: a systematic review

Mondello S.*[1], Cernigliaro A.[6], Milli C.[2], Kobeissy F.[1], Silvestri C.[2], D"Amato S.[3], Di Napoli A.[4], Cruciani F.[7], Giorgi Rossi P.[5], Petrelli A.[4], Scondotto S.[6]
[1]University of Messina ~ Messina ~ Italy, [2]Agenzia regionale di sanità della Toscana ~ Firenze ~ Italy, [3]Ministry of Health ~ Rome ~ Italy, [4]INMP - National Institute for Health, Migration and Poverty ~ Rome ~ Italy, [5]Epidemiology Unit, Azienda USL – IRCCS di Reggio Emilia ~ Reggio Emilia ~ Italy, [6]Regional Health Authority of Sicily ~ Palermo ~ Italy, [7]Dipartimento Epidemiologia del S.S.R., Regione Lazio ~ Rome ~ Italy

The COVID-19 pandemic has amplified health disparities, particularly in populations with substantial vulnerabilities, such as international migrants and refugees. We thus performed a systematic review of studies evaluating health interventions for COVID-19 tailored to and targeting migrant populations to provide a rigorous and exhaustive evidence synthesis of their effectiveness and define best practices and policies capable of improving outcomes. Data sources included MEDLINE, EMBASE, LOVE Platform COVID-19 Evidence, and Cochrane Central Register of Controlled Trials (CENTRAL). Two authors independently performed study selection, data extraction, and quality assessment. The strength of evidence was determined using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE). Data were reported according to methodological guidelines for systematic review provided by the Cochrane Collaboration and the PRISMA statement. The search strategy retrieved 1941 unique citations. Three eligible studies were included. One study employed agent-based models of COVID-19 outbreaks in a refugee camp setting to evaluate non-pharmaceutical interventions, namely sectoring (i.e., dividing the camp into subunits with separate food lines and Services), face masks, remove-and-isolate (i.e., identifying and isolating infectious individuals and their families), and lockdown [1]. The Moria refugee camp, including 18,700 individuals, was used as the model population. The second study compared a 42-day prophylaxis regimen of either oral hydroxychloroquine, oral ivermectin, povidone-iodine throat spray, oral zinc/vitamin C combination, or oral vitamin C, administered to 3037 healthy migrant workers quarantined in a multi-story dormitory in Singapore [2]. The third study evaluated the effectiveness of a community-centered, culturally-tailored, theory-informed vaccination strategy, including "mobilization, vaccination, and activation" components, in increasing the uptake of COVID-19 vaccination among the underserved Latinx population [3]. This program was implemented in a large immigrant community in San Francisco, California, with an estimated population of 72,380 individuals (33.4% Latinx). The certainty of the evidence was very low across all three studies. There is insufficient evidence to draw reliable and generalizable conclusions regarding the potential effectiveness of interventions for COVID-19 targeting migrant populations. This review was conducted within a program funded by the Italian Ministry of Health aimed at defining and implementing interventions to control the COVID-19 pandemic in Italy (CCM).

[1] R.T. Gilman, S. Mahroof-Shaffi, C. Harkensee, et al. BMJ global health, 2020;5, e003727.
[2] R.C.S. Seet, A.M.L. Quek, D.S.Q., Ooi, et al. IJID: official publication of the International Society for Infectious Diseases, 2021,106, 314-322.
[3] C. Marquez, A.D. Kerkhoff, J. Naso, et al. PloS one, 2021,16, e0257111.

MP36 Short-term forecasting evaluation of covid-19 epidemic waves in the basque country

Larrea N.[1], Alvarez O.[2], Lee D.[2], Arostegui I.[3], Millan E.[4], Barrio I.[3], Quintana J.M.*[1]
[1]Research Unit Hospital Galdakao-Usansolo ~ Galdakao ~ Spain, [2]BCAM ~ Bilbao ~ Spain, [3]Department of Mathematics, UPV/EHU ~ Leioa ~ Spain, [4]Healthcare Services Sub-directorate, Osakidetza-Basque Health Service ~ Vitoria ~ Spain

One of the objectives during the pandemic has been to provide clinicians and healthcare managers with daily information on the evolution of the pandemic. The aim of this study is to evaluate whether the forecasting modelling approach used in our area to provide short-term forecasts was accurate enough to detect different phases of the epidemic and changes due to exogenous factors. Cohort study which included all the population of the Basque Country older than 18 years with a SARS-CoV-2 positive test between 14-3-2020 and 18-3-2022. Basic data on the evolution of the pandemic (death, number of hospitalisations and ICU admission due to COVID-19), and variables such as age, sex and place of residence were collected. We used smoothing methods (penalized regression splines) to consider daily counts of COVID-19 hospitalizations and deaths, and in order to deal with overdispersion we use the Negative Binomial distribution for the counts. To evaluate the model performance, the error of the forecasts was estimated by RMSE (root mean square error), absolute error and relative error in the 2-day and 5-day forecasts at two one-month time intervals, in the quarantine period (14-3,13-4 2020) and when the omicron variant wave was beginning (29-12- 21,28-1- 22). The median absolute error of the model for death during the quarantine period was 4.13 and 8.20 for the 2- and 5-day prediction respectively, and the median relative error was 11% and 28%. For hospitalizations, these errors increase to 18.45 and 29.05, and 25% and 38% respectively. However, when the Omicron variant appears, the median absolute error was 9.30 and 13.20, and the relative error 40% and 60% in the case of death, while in the prediction of hospitalizations, these errors increase to 69.53 and 114.37, and 75% and 121% respectively. In all cases, always underestimated. The results obtained indicate that a simple model for short-term prediction fits well when the epidemic curve is stable. However, when there is a change in trend over a minor time interval, the model is not able to predict it, and that fits better to objective data such as death than for hospitalisation

Carballo A, Durban M, Kauermann G, Lee D-J. A general framework for prediction in penalized regression. Statistical Modelling. 2021;21(4):293-312. doi:10.1177/1471082X19896867

Poster Sessions

MP37 Usefulness of ecological socio-economic indicators in sars-cov-2 infection modeling: a french case study

Romain--Scelle N.*[1], Riche B.[2], Benet T.[3], Rabilloud M.[1]
[1]Université Lyon 1, CNRS UMR 5558 Laboratoire de Biométrie et Biologie Evolutive ~ Villeurbanne ~ France, [2]Hospices Civils de Lyon, Service de Biostatistique et Bioinformatique ~ Pierre-Bénite ~ France, [3]Santé Publique France, Direction des Régions ~ Lyon ~ France

Following its emergence in January 2020, worldwide SARS-CoV-2 diffusion occurred for a year with only non-pharmaceutical interventions (NPIs) available as mitigation tools. Based on the knowledge of ecological indicators predictive of infectious risk (influenza, tuberculosis) (1,2), we aimed to assess the predictive capability of 10 census-based indicators at the neighborhood level on the infection risk by SARS-CoV-2 in the French Auvergne-Rhône-Alpes region to improve targeting of NPIs. We collected and aggregated all counts of biologically confirmed cases of SARS-CoV-2 infection in the region ARA at the neighborhood level for 4 epidemic phases: low incidence (summer 2020), growth (September-October 2020), peak and decrease (November-December 2020) and stabilization (winter 20-21). Vaccination was in early rollout phase by the end of study period and ignored. 10 census-based ecological covariates for human mobility and socio-economic position (population density, education level, unemployment, immigrants, household composition, and usual mobility modes) were evaluated as predictors of case incidence using a Poisson regression with conditional autoregressive (CAR) spatial effects (3). Benefits of CAR effects and covariates on model fit were evaluated using pseudo-R^2 and Moran's I statistics. 438,992 infection cases over 5,410 neighborhoods among 7,917,997 inhabitants were analyzed. The association between covariates and case incidence was inconstant: the population density, the proportion of individuals working outside their residence town and the proportion of household without child are the only covariates with a consistent effect across time periods (RR 1.29 [1.20;1.39], 1.04 [1.02;1.07], 0.94 [0.91;0.97] respectively, growth phase). Spatial correlation was estimated at high levels for the last three periods. Spatial CAR effects were necessary to improve on the pseudo-R^2 and the Moran's I statistics compared to the null model (intercept only). The ecological covariates assessed were insufficient to adequately model the distribution of cases without accounting for the spatial organization of the epidemic. Association between covariates and incidence is inconstant across time and scale of analysis. Use of modeling strategies providing local estimates of parameters effect would be useful in this specific context (4).

1. Duarte R, Aguiar A, Pinto M, Furtado I, Tiberi S, Lönnroth K, et al. Different disease, same challenges: Social determinants of tuberculosis and COVID-19. Pulmonology. 2021 Jul 1;27(4):338– 44.
2. Grantz KH, Rane MS, Salje H, Glass GE, Schachterle SE, Cummings DAT. Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. Proc Natl Acad Sci. 2016 Nov 29;113(48):13839–44.
3. Lee D. CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. J Stat Softw. 2013;55(13):1–24.
4. Fotheringham AS, Sachdeva M. On the importance of thinking locally for statistics and society. Spat Stat. 2022 Aug 1;50:100601.

MP38 Optimal stage-wise allocation ratios in multi-arm multi-stage designs

Azher R.*, Wason J., Grayling M.
Newcastle university ~ United Kingdom

Multi-arm multi-stage trials (MAMS) are an efficient approach to simultaneously evaluating several experimental treatments against a shared control arm. In MAMS trials, multiple stages allow experimental treatments to be removed from the trial early, e.g., if they are unlikely to be significantly better than control. This can greatly reduce the required sample size compared to testing each experimental arm separately against the control arm in a series of single-stage two-arm trials. At the interim analyses in a MAMS trial, it may be of interest to adjust the allocation ratios in the subsequent stage(s), e.g., to try to maximise power. In a fixed-sample multi-arm trial, the optimal allocation ratio for maximising power is, under certain standard assumptions, K^0.5 : 1 in favour of control, where K is the number of experimental treatment arms. However, this is not the case for MAMS designs because treatments can be removed early. : In this work, we therefore seek to determine the optimal stage-wise allocation ratios in MAMS trials that can achieve maximal marginal, disjunctive, or conjunctive power. Using the TAILoR trial as a motivating example, we use analytical formulae to explore a range of allocation ratios, fixing the study's allowed maximal sample size and boundary 'type' (e.g., Pocock) for a fair comparison. Results indicate that the choice of stage-wise allocation ratios has a considerable impact on the various types of power. For two-stage designs for three experimental arms, the highest marginal power is observed for the allocation ratios 1.33:1 and 2:1 in stages 1 and 2 respectively. By contrast, the highest disjunctive power is achieved by 3:1 allocation in stage 1 and 2:1 in stage 2. Alteration of the stage-wise allocation ratios is also, as expected, associated with sizeable variation in the expected sample size. Our findings demonstrate that for a fixed maximal sample size researchers can alter allocation ratios to effectively increase power or reduce the expected sample size.

1. Neuhäuser, M., Mackowiak, M. M., & Ruxton, G. D. (2021). Unequal sample sizes according to the square-root allocation rule are useful when comparing several treatments with a control. Ethology, 127(12), 1094-1100.

2. Wason, J. M., & Jaki, T. (2012). Optimal design of multi-arm multi-stage trials. Statistics in medicine, 31(30), 4269-4279.

MP39 A latent variable model for borrowing of information in a basket trial

Cherlin S.*, Wason J.M.S.
Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

In clinical trials of immune-mediated inflammatory diseases, the outcome is often a combination of a dichotomised measure of a disease activity score and a binary indicator such as administration of rescue therapy. Patients are classified as responders if their disease activity score improves by a prespecified level, and they do not withdraw or do not require rescue medication. Due to the loss of information in dichotomising the data, the analyses of such trials do not make the maximal use of the data available. The augmented binary method [1] has been proposed to utlilise the continuous measures to improve the precision of the estimates. Here we extend the augmented binary method to basket trials. Basket trials are a new class of trials that evaluate a new treatment in several related conditions simultaneously. Originally developed for oncology, they are increasingly of interest for use in immune-mediated inflammatory diseases. Basket trials allow for more efficient analysis due to borrowing of information across subtrials, i.e., treatment effect in one subtrial may provide information on treatment effect in other subtrials. We develop methodology that extends the augmented binary method to basket trials with responder outcomes. We propose a Bayesian hierarchical latent variable model, which assumes that the discrete outcomes are manifestations of latent continuous measures, and jointly model the observed and the latent continuous variables. The observed discrete variable is related to the latent continuous variable by partitioning the latent variable space. Hierarchical modelling allows borrowing of information between the subtrials, with the amount of borrowing being determined by the prior distributions of the parameters. We investigate the operating characteristics of the method using simulated data and show that it results in narrower credible intervals for log odds ratios, in comparison to a standard logistic regression modelling. Numerical results suggest that our methodology can improve the precision of estimates in certain scenarios, in comparison to a standard approach, which could lead to increased statistical power. Further work would focus on investigating the optimal mechanism for borrowing information between the subtrials.

[1] J.M.S. Wason, S.R. Seaman, Statistics in Medicine, 32, 2013, 4639-4650.

MP40 Tolerance interval-based hypothesis testing for the similarity between two independent populations

Chiang C.*[1], Hsiao C.[2]
[1]Tamkang University ~ New Taipei ~ Taiwan, [2]National Health Research Institutes ~ Zhunan ~ Taiwan

In many fields, judging whether two responses are equivalent or similar is often of interest. However, how similar is similar? Testing the mean difference traditionally offers a statistical answer to the question, but this might be not enough when the two populations have heterogeneous variances. Furthermore, an additional test for assessing variances may lead to multiple adjustments for the error rates. In this study, we suggest testing whether a certain proportion of the test population is included within a certain proportion of the reference population. In doing so, accuracy and precision can be assessed simultaneously by tolerance interval-based hypothesis testing. Moreover, the proposed method allows that the accuracy of the test group is a poorer, but acceptable, with a better precision as compared with that of the reference group. An asymptotic distribution of the two-sided tolerance interval is derived to calculate statistical characteristics such as power, p-value, and required sample size. Suitable statistical properties are shown by simulation. This study suggests an alternative interval estimator-based hypothesis testing for comparing two populations. Based on the normal assumption, although the exact distribution of the tolerance bounds is still unclear, simulation shows that they follow a bivariate normal distribution asymptotically. The proposed method can apply to develop a hypothesis testing based on a prediction interval.

MP41 Sample size recalculation for a skewed outcome in two-stage three-arm sequential noninferiority clinical trial

Chiaruttini M.V.*[1], Azzolina D.[2], Gallocchio J.[1], Desideri A.[3], Gregori D.[1]
[1]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic and Vascular Sciences, University of Padova ~ Padova ~ Italy, [2]Department of Environmental and Preventive Science, University of Ferrara ~ Ferrara ~ Italy, [3]Cardiological Department, ULSS2 Marca Trevigiana ~ Treviso ~ Italy

The gold standard for the non-inferiority study is a three-arm design including a placebo in addition to the experimental group and the active comparator. Three-arm non-inferiority trials are challenging for the hypothesis formulation, and their design is often characterized by uncertainty in estimating the experimental treatment effect. Some methods have been proposed to optimize the recalculation of the sample size at interim analysis for Gaussian, Bernoulli, and Poisson outcomes but not for the continuous skew outcome. The present simulation study focuses on calculating the sample size of two-stage, three-arm sequential non-inferiority clinical trials with skewed endpoint distribution. In the case of asymmetrical outcomes simulated as gamma variables, since the variance is function of the mean, the homoskedasticity has been pointed out when a mean difference in treatment effects is assumed. Thus, for comparison purposes, we provide a resampling algorithm that maintains the normal assumption but accounts for different standard deviations to be set across the three arms. We found that if the discrepancy between the group variances is considered, we can achieve the desired power, avoiding the risk of overestimation (saving patients) or underestimation (saving power), even in case of deviation from normality. We provide a real data example from the COSTAMI trial [1]. Lastly, we developed a Web application for sample size/coverage probability estimation available at https://r- ubesp.dctv.unipd.it/shiny/reskout/. The proposed algorithm and the corresponding tool help to keep track of the properties of the design, as it provides an estimate of an overall probability of success/failure of the study, giving us the possibility to choose the best reliable set of parameters to optimize the sample size estimation procedure.

[1] A. Desideri et al. Eur. Heart J. 24, 2003, 1630–1639.

MP42 Use of an adaptive design in a randomly assigned, open-label, balanced incomplete block design study

Darwish N.*
Société des Produits Nestlé ~ Lausanne ~ Switzerland

Share the implementation of an adaptive design in a randomly assigned, open-label, balanced incomplete block design study with a nutrition intervention. In a randomized, open-label, balanced incomplete block adaptive design study, 24 participants were assigned to a sequence of two out of three products. Two dosages of a chocolate/milk drink and water were given to children after an overnight fasting and liver glycogen was measured in regular intervals after the ingestion of the drinks by NMR imaging technique. Given the lack of knowledge in the field investigated, a formal sample size calculation was not performed at the stage of protocol development and an interim analysis was planned after 9 subjects had completed both visits, to assess the conditional statistical power for eventual modifications of the study design, such as stopping the trial for either success or futility at interim or dropping one of the arms in the second stage of the trial. In order to control the inflation of Type I error rate, Pocock alpha-spending function was used. The adaptive design was implemented using the rpact package in R version 3.0.4 [1]. Furthermore, to control for the multiplicity of comparisons, a hierarchical order for hypotheses testing was put in place. The computed Pocock p-value boundary was 0.0280 at interim and final analysis. At interim, the available data provided enough evidence to demonstrate a significant difference at a 2.8% level only for one out of the two comparisons. The milk dose that showed statistical difference in liver glycogen compared with water was dropped and the conditional power was considered as sufficient on continuing the "unsuccessful" arm until 24 participants completed the study. We had an internal Data Monitoring Committee. The final analysis demonstrated the nutritional objective. The somehow unusual application of an adaptive design in a small sample size metabolic study was justified by a low recruitment rate which got additionally prolonged by the COVID-19 lockdown. Overall, the implementation of the adaptive design was efficient and led to a successful trial.

[1] G. Wassmer, F. Pahlke, rpact: Confirmatory Adaptive Clinical Trial Design and Analysis, 2021, https://CRAN.R-project.org/package=rpact.

MP43 A comparison of sample size re-estimation methods for cluster-randomized trials

Gunn H.*[1], Kravets S.[2], Mandrekar S.[1]
1Mayo Clinic ~ Rochester ~ United States of America, 2University of Illinois ~ Chicago ~ United States of America

Cluster-randomized trials are becoming more common due to their pragmatic implementation, but this comes at the cost of statistical complexity. When conducting an a priori power analysis for a parallel arm cluster-randomized trial, the intraclass correlation (ICC) is a necessary but often unknown parameter. Underestimating the ICC at the design stage can lead to an underpowered analysis after accrual is completed. Previous studies that have investigated an internal pilot procedure for cluster-randomized trials considered family as the cluster [1].
There are nuances and limitations when considering clinic as a cluster (e.g., often difficult to increase number of clusters but can increase cluster size, different average baseline ICC values). Using previous trial data to determine parameter values, we conducted a Monte Carlo simulation study to compare two sample size re-estimation procedures – an internal pilot study approach and the combination test – to the conventional design in the context of a cluster-randomized trial assuming patients are clustered within clinics. At the design stage, we solved for the average number of patients needed per clinic to achieve 0.90 power, allowing for unequal cluster sizes. The re-estimation procedures were used when accrual reached 50% [2]. We hypothesize that the two sample size re- estimation procedures will have greater power than the conventional procedure at the final analysis stage due to the re-estimation. Additionally, we hypothesize that the combination test will have more acceptable Type I error rates than the other two designs because the internal pilot design does not control for the sample size in the second phase being dependent on the variance estimates of the first stage of data collection, causing the Type I error rates to inflate. We varied the data-generating ICC, the initial ICC estimate at the design stage, the initial number of clusters, and the effect size. Simulation results will be presented. Nuisance parameters like the ICC, if not properly accounted for at the design stage, can have a large impact on the power of a study. This study is the first instance of applying the combination test to a cluster-randomized design in the context of sample size re-estimation.

[1] S. Lake, E. Kammann, N. Klar, R. Betensky. Statistics in Medicine, 21, 2002, 1337-50.
[2] S. van Schie, M. Moerbeek. Statistics in Medicine, 33, 2014, 3253-68.

Poster Sessions

MP44 Design of a cluster-randomized trial:telehealth-enabled physician adherence monitoring in intensive care units

Holubkov R.*[1], Grissom C.K.[2], Srivastava R.[2], Knighton A.J.[2], Wolfe D.[2], Jacobs J.R.[2], Peltan I.D.[2]
[1]University of Utah ~ Salt Lake City ~ United States of America, [2]Intermountain Healthcare ~ Salt Lake City ~ United States of America

We designed and implemented a Type II hybrid effectiveness-implementation cluster-randomized trial assessing whether telehealth-enabled real-time audit/feedback to physician adherence ("TEACH"), compared to usual audit/feedback ("control"), improves adherence to patient awakening/extubation protocols and increases ventilator-free patient days (VFDs). Important design facets include small number of clusters, multiple levels of clustering for the adherence outcome, and nonparametric analysis of VFDs in a clustered setting. Following six months of baseline-phase data collection, 12 hospitals were randomized 1:1 to TEACH or control, stratifying by patient volume. Following a five-month run-in phase in late 2022, intervention effect is presently being assessed for 33 months. We assess TEACH effect on adherence as relative odds of improvement from baseline to intervention phases (patient day being unit of analysis) at TEACH versus control hospitals. Analysis implements a mixed logistic model with fixed effects for phase and treatment arm, and random effects for hospital, patient, and calendar day within hospital. Due to small number of clusters, inference will use a t- distribution with degrees of freedom indexed to number of centers[1]. Simulation-based power estimation algorithms directly modeled center effect, with clustering substantially reducing effective within-center sample sizes (conservatively assuming ICC of 0.225 for patient "clusters"). Power is substantial assuming TEACH intervention improves adherence odds 1.33-fold.
VFDs have U-shaped distributions due to substantial mortality (death=0 VFDs) and many patients extubated early (=high number of VFDs). TEACH effect will be assessed as differences between treatment arms in VFD change from baseline to intervention phases, using a proportional odds model with fixed effects for study phase, treatment, prespecified baseline patient covariates, and a random hospital effect[2]. Due to potential model convergence issues with many distinct VFD values, the design includes a prespecified scheme to maximize VFD granularity independently of any treatment effect. Power is satisfactory assuming TEACH increases VFDs by 1 day among two-thirds of patients who survive. Recruitment has exceeded expected numbers as of 3/2023. The DSMB will review progress/safety in 4/2023, with interim efficacy analysis in late 2023. The TEACH trial design combines statistical rigor with prespecified flexibility of analyses, implementing contemporary design-based approaches supported by extensive simulation-based power and performance assessments.

[1] P. Li, D. T. Redden. Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. BMC Medical Research Methodology 15, 2015, 38.
[2] R.H.B. Christensen. ordinal—Regression Models for Ordinal Data. R package version 2022.11-16 https://CRAN.R-project.org/package=ordinal.

MP45 Statistical validation methods of sound recording medical device on respiratory clinical trials

Kranidi A.*, Pozzebon A., Grant T., Crispino G.
StatisticaMedica Ltd ~ Dublin ~ Ireland

Digital technologies are being used more often in clinical trials to measure a treatment's efficacy. Having the medical device fully validated and accepted by the regulatory authorities is therefore essential. A recording device that records and stores cough sounds over a 24-hour period is being developed. Trained analysts count the coughs from a file which is compressed by an algorithm to a recording with an approximately one-hour duration. Validating that the number of coughs in the compressed recording agree with the uncompressed cough count is required. Firstly, the agreement between compressed and uncompressed (24-hour long file) recordings is assessed. The Bland-Altman plot is chosen as a measure of agreement at various timepoints. The percentage difference of the uncompressed minus compressed counts is presented in the y-axis and the uncompressed counts in the x-axis. A bandwidth of bias (accuracy) ±1*standard deviation (precision), ±8%, is used as warning limits which is based on historical data for the device. Secondly, to assess the agreement between raters, a second application of the Bland-Altman (BA) approach is considered. While consistency of raters is the more traditional approach, the ICCs in this application are too high and are saturated. Given the narrow variability observed in the historical inter- rater reliability, the control charts bandwidth is set at 3*standard deviation, indicating that a bandwidth of ±50 cough counts difference should be used. In addition to the BA plot, the Intra-Class Coefficient (ICC) is calculated to identify the level of reliability of the results. A series of simulations utilizing historical data show a threshold of 99.65% is more appropriate for the sound device demonstrating the ineffectiveness of this measure in this application. The results showed that the cough counting process is in control. This project provides insight to the methods used in Statistical Process Control, e.g. control charts, BA plots and ICC in the specific context of devices with high consistency. At the same time, it proves how important is the utilization of historical data in this work.

[1] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet, 327:8476, 1986, 307-310.
[2] J. Smith, et al., Reliability of Manual Cough Counting Using the VitaloJAKTM, American Thoracic Society, TP44.TP044 ASSESSMENT AND TREATMENT OF COUGH AND CHRONIC DYSPNEA, American Journal of Respiratory and Critical Care Medicine, 203:A2361, 2021.

MP46 Non-inferiority trials with evidence of assay sensitivity using population adjustment method

Okamura S.*[1], Hida E.[2]
[1]Department of Medical Innovation, Osaka University Hospital ~ Osaka ~ Japan, [2]Graduate School of Medicine, Osaka University ~ Osaka ~ Japan

The choice of a non-inferiority (NI) margin and assurance of assay sensitivity are well-known issues in 2-arm NI trials. A 3-arm NI trial including both a placebo and a control treatment is strongly recommended to assess assay sensitivity. However, there are concerns about the ethics and feasibility of including a placebo; consequently, practical applications of the 3-arm NI trial have not progressed. Therefore, new methods are needed to quantitatively assess assay sensitivity of NI trials using the results of a 2-arm NI trial and historical trials. We propose one approach to confirm assay sensitivity in a 2-arm NI trial by using aggregate data from other historical trials. In addition, simulation studies evaluate the performance of the proposed approach and discuss its feasibility in terms of confirming assay sensitivity. To assess assay sensitivity of NI trials, the control treatment must be NI margin or more effective than a placebo (substantial superiority: Hida & Tango, 2018). However, there may be limited historical trials and information used to demonstrate this substantial superiority. For example, characteristics such as trial design and patient background may differ between historical trials with only summary statistics and a 2-arm NI trial with current individual patient data. If these differences in patient background between trials are ignored, the assay sensitivity of a NI trial cannot be properly assessed. To address this problem, we propose a method to assess assay sensitivity using partially a population adjustment method: the matching adjusted indirect comparison or the simulated treatment comparison. Furthermore, the performance of the proposed method is investigated in terms of actual Type I error rates and power by Monte Carlo simulations set up under various scenarios based on real clinical trials. The level of evidence for the proposed method may be lower than that for a 3-arm NI trial, owing to the use of external information and indirect comparison. However, the results of various simulations suggest that the performance of the proposed method is useful as one of the more feasible methods to assess assay sensitivity of the 2-arm NI trial. Hida E. and Tango T. Pharmaceutical Statistics. 2018. 17(5). 489-503.

MP47 When covid-19 specific vaccines became an intercurrent event

Orsini F.*
Clinical Epidemiologu and Biostatistic Unit, Murdoch Children's Research Institute, Melbourne Children's Trial Centre ~ Melbourne ~ Australia

The COVID-19 pandemic had a global impact on the conduct and analysis of trials. Complications arose in terms of both operational and health-related challenges. This resulted in unforeseen intercurrent events (ICE) that affected the existence and the interpretation of the measurements associated with the research question. The BRACE (BCG vaccination to Reduce the impact of COVID-19 in healthcare workers following Coronavirus Exposure) trial[1] was a phase III, two arm, multicentre, randomised placebo-controlled trial of BCG vaccine to reduce the incidence of COVID-19 in healthcare workers during the coronavirus pandemic. BRACE was stopped prematurely due to the rollout of COVID-19-specific vaccines, affecting the ability of the trial to determine the effectiveness of BCG vaccination in protecting against COVID-19. The aim of this work is to describe potential estimands of interest within the BRACE trial and the effect this has on the estimated efficacy of the BCG vaccination. The estimand of interest was the difference in the proportion of participants with symptomatic COVID- 19 by 6 months between the BCG and placebo arms, estimated using a time-to-event analysis via a flexible parametric survival model (Royston-Parmar model). In the primary analysis, the ICE of receiving a COVID-19 vaccine was handled using a hypothetical strategy, to assess the efficacy of BCG if a COVID-19 vaccine had not been found. Under this strategy, participants who received a COVID-19 vaccine had their data censored at the date of their first dose of a COVID-19 vaccine. As supplementary analysis, the same ICE was handled using a Treatment Policy Strategy, including follow-up after COVID-19-specific vaccine. The estimated treatment effect under the two strategies were similar, but the 95% confidence intervals (CI) were narrower in the supplementary analysis as expected given the additional follow-up information. The results of this study suggested that the ICE of receiving a COVID-19 vaccine did not alter the efficacy of the BCG vaccine. This case study highlights the importance of the estimand framework in guiding analysis planning, as it helps to specify the research question that the trial is designed to answer, leading to a better understanding of the treatment effect being estimated.

[1] Pittet LF, Messina NL, Gardiner K the BRACE trial Consortium Group, et al BCG vaccination to reduce the impact of COVID-19 in healthcare workers: Protocol for a randomised controlled trial (BRACE trial) BMJ Open 2021;11:e052101. doi: 10.1136/bmjopen-2021-052101

Poster Sessions

MP48 Experiences with a bayesian adaptive design for the evaluation of a biomarker-based treatment algorithm

Wiemer J.C.*[1], Gehrig S.[2], Johannes S.[1], Inlall D.[1], Atallah J.[3], Warren H.M.[4], Mansour M.K.[3]
[1]B·R·A·H·M·S GmbH, part of Thermo Fisher Scientific ~ Hennigsdorf ~ Germany, [2]estimact ~ Berlin ~ Germany, [3]Division of Infectious Diseases, Massachusetts General Hospital & Department of Medicine, Harvard Medical School ~ Boston ~ United States of America, [4]Division of Infectious Diseases, Massachusetts General Hospital ~ Boston ~ United States of America

Advantages of Bayesian statistics include intuitive uncertainty by probabilistic statements and an inherently adaptive nature, which prompts its use for adaptive clinical trials. We set up a study for the validation of a biomarker-based treatment algorithm with a Bayesian adaptive design. Here, we share our practical experiences in the application of the methodology, both computationally and statistically. Clinical study: The clinical trial "ProSAVE" (NCT04158804) aims to validate advantages of a biomarker-based algorithm for the medical treatment of patients with suspected pneumonia: reduction of antibiotic exposure (more short antibiotic treatments lasting less than four days, primary superiority endpoint) without increasing the risk of adverse events (composite adverse event endpoint, secondary non-inferiority endpoint), and also reducing the number of antibiotic prescriptions at hospital discharge. Statistical approach: The study was designed for adequate statistical power with binary endpoints and a maximum of 700 patients. Interim analyses were planned concerning futility with 200, 350, 550 and 650 patients (futility: <10% predictive probability of study success with 700 patients, including partial follow-up, non-binding stopping recommendation) and efficacy with 550 and 650 patients (efficacy: >90% predictive probability of study success with so far enrolled patients for three endpoints, binding stopping). Simulations for design characteristics were initially conducted by external consultancy using Amazon Web Services and later within a parallelized R simulation framework (dplyr-tidyverse-furrr) on local machines [1]. (R1) All necessary simulations for study design characteristics could be carried out on a laptop within the R simulation framework (Intel®Core™i7-10850H@2.70GHz, 32GB RAM). Internal simulations took about 3 weeks (implementation, testing, results generation). Highest computer requirements were for type-1 error calculations (4 hours per fixed set of study design parameters). (R2) Interim analyses triggered changes in study protocol (from superiority to non-inferiority, order of endpoints), and questions on multiple testing corrections and adequate type-1 error control [2]. (R3) Additional challenges were data generation assumptions and meaningful communication of numerous simulation results in figures and tables for cross-functional discussion and alignment. The R-laptop-based approach allowed specification of study design parameters, additional insights into study design, and avoidance of high costs for computer facilities and consultancy. We present and discuss the implemented solutions.

[1] Neilson, M: PSI Online Training Course "Simulation of Clinical Trials using Tidyverse", 14th March - 7th April 2022
[2] CPMP/EWP/482/99: Points to consider on switching between superiority and non-inferiority, 2000

MP49 Modelling the rate of change of gfr in the presence of competing risks

Belcher J.*, Solis--Trapala I., Sim J.
Keele University ~ Stoke on Trent ~ United Kingdom

Van Eijk et al. [1] observed that the prospective use of joint modeling of time-to-event and longitudinal outcomes remains negligible in clinical trial practice. Chesnaye et al. [2] recommended these approaches should be used more extensively in nephrology research. The BISTRO Trial was an open-label, two-arm pragmatic randomized trial of 437 subjects, designed to investigate whether use of bioimpedance spectroscopy helps guide fluid management by avoiding dialysis-related fluid volume depletion in incident haemodialysis patients. Primary outcome measures were time to anuria and rate of decline in residual kidney function measured by glomerular filtration rate (GFR). Time to anuria was not observable for participants who underwent a kidney transplant or died during the follow-up period. The occurrence of anuria, transplantation and the risk of death are associated with the rate of GFR, inducing non-ignorable missing values on this outcome. A competing risks joint model is proposed to make inferences on the GFR slope. The longitudinal submodel consisted of a linear mixed-effects segmented regression defined to capture the rate of change of GFR in years 1 and 2. Linear terms for time and trend changes and a dummy variable reflecting level change were defined, together with interaction terms reflecting treatment slope differences. This simple interrupted time- series design allows clinicians to easily assess changes in slopes and intercepts. The survival component took the form of a relative risk submodel for each possible competing event using a proportional hazards model with the log baseline hazard approximated using B-splines. Joint modeling and separate longitudinal analysis using linear mixed-effects segmented regression yielded similar estimates, but smaller standard errors were observed in the former.

Term	Linear mixed-effects model	SE	Joint model	SE
Year 1 Slope [control] *	−0.167	0.018	−0.167	0.012
Year 1 Slope difference	−0.014	0.025	−0.016	0.016
Year 2 Slope change [control] *	0.143	0.032	0.140	0.022
Year 2 Slope change difference	−0.052	0.043	−0.040	0.030

*reference category
The value of our joint modelling approach was modelling of the data generating mechanism tailored to provide ease of interpretation on the parameters of interest and improved precision, at the cost of model complexity.

[1] R.P.A. van Eijk, K.C.B. Roes, L.H. van den Berg et al. Journal of Clinical Epidemiology, 2022, 147, 32-39.
[2] N.C. Chesnaye, G. Tripepi, F.W. Dekker et al. Clinical Kidney Journal, 2020, 13, 143-149.

Poster Sessions

MP50 Multivariate repeated measures analysis - testing and post-hoc procedures under non-normality

Graf R.*[1], Zeldovich M.[2], Friedrich S.[1]
[1]Department of Mathematics, University of Augsburg ~ Augsburg ~ Germany, [2]Institute of Medical Psychology and Medical Sociology, University Medical Center Göttingen ~ Göttingen ~ Germany

Linear classification methods for repeated measures data are rarely discussed in the literature but have potential applications in psychology and sociology, where Linear Discriminant Analysis (LDA) is frequently applied. Despite longitudinal data collection in these fields, the application of LDA to data measured at a single time point predominate. Longitudinal data are characterised by complex correlations between time points and variables. Datasets in psychology and sociology rarely fulfill the normality assumption, so we are focussing on existing methods that are robust to deviations from normality. Repeated measures multivariate analysis of variance (MANOVA) methods for testing the statistical significance of differences in means based on resampling approaches are suitable for nonnormal data with unequal class covariances. After a significant group difference has been determined, interest focusses on post-hoc comparisons. Here, different classification algorithms for prediction can be applied. We compare the performance of linear classification algorithms in a simulation study based on parameter estimates of real datasets comprising Likert-type data and by evaluating different performance measures. The performance of standard repeated measures LDA, which depends on the multivariate normality assumption, is compared to multiple robust and nonparametric alternatives. Multivariate repeated measures data should be analysed using suitable methods if considerable correlations between measurement occasions exist instead of conducting a series of (independent) analyses for single time points.
Our method comparison study revealed that methods that include the removal of multivariate outliers before parameter estimation as well as a method using robust estimation of parameters by application of joint Generalized Estimating Equations showed better performance compared to the standard repeated measures LDA. The results of the longitudinal Support Vector Machine were not competitive.

Banks, J., Batty, G., Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., and Zaninotto, P. (2021). English Longitudinal Study of Ageing: Waves 0-9, 1998-2019 [data collection]. 36th Edition. UK Data Service. SN: 5050.

Brobbey, A. (2021). Classification Models for Multivariate Non-normal Repeated Measures Data. [Doctoral thesis, University of Calgary] https://prism.ucalgary.ca/handle/1880/112972?show=full

Brobbey, A., Wiebe, S., Nettel-Aguirre, A., Josephson, C. B., Williamson, T., Lix, L. M., and Sajobi, T. T. (2022). Repeated measures discriminant analysis using multivariate generalized estimation equations. Statistical Methods in Medical Research, 31(4):646–657.

Chen, S. and Bowman, F. D. (2011). A novel support vector classifier for longitudinal high-dimensional data and its application to neuroimaging data. Statistical Analysis and Data Mining: The ASA Data

Science Journal, 4:604–611.

Friedrich, S. and Pauly, M. (2018). MATS: Inference for potentially singular and heteroscedastic

MANOVA. Journal of Multivariate Analysis, 165:166–179.

Inan, G. (2015). JGEE: Joint Generalized Estimating Equation Solver. CRAN. https://CRAN.R- project.org/package=JGEE

Lix, L. and Sajobi, T. (2010). Discriminant Analysis for Repeated Measures Data: A Review. Frontiers in Psychology, 1:146.

Wahl, S., Boulesteix, A.-L., Zierer, A., Thorand, B., and Wiel, M. (2016). Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. BMC Medical Research Methodology, 16:144.

Zeldovich, M. (2018). Outcome Measurement in Russian Clinical Praxis: Clinical Outcome in Routine Evaluation – Outcome Measure (CORE-OM) [Doctoral thesis, Alpen-Adria University of Klagenfurt]. https://netlibrary.aau.at/obvuklhs/content/titleinfo/5370233

MP51 The use of multilevel models to explain regional disparities: illustration on pneumococcal vaccination

Herquelot E.*[1], Assi N.[1], Batisse A.[1], Goussiaume G.[2], Grenier B.[1], Raguideau F.[1]
EVA ~ Lyon ~ France, [2]Pfizer ~ Paris ~ France

The regional disparities on an epidemiologic criterion can be explained by the characteristics of the region (socio economic level, care network) and the case-mix of subjects in the region. This information is hierarchic and the subjects of the same region are correlated. The multilevel models are a special case of mixed model designed to deal with this problematic. The aim of this study is to present an illustration of multilevel models with two levels information (patients and region) in the context of regional disparities using 2018 data of diabetic adults from Covarisq study (1). A generalized hierarchical mixed linear model was performed on the probability of pneumococcal vaccination with a binomial distribution and a logit link. The characteristics at the patient-level and the characteristics at region-level were included as fixed effects and the region as a random effect. For continuous covariates, linear splines were implemented. A total of 2,374,070 diabetic patients were included in 2018. Among them, 1.8 % were up to date with their pneumococcal vaccination. The following patient-level variables were significantly associated with pneumococcal vaccination: sex (Odd-Ratios (OR) of 1.08), age (spline association) the presence of other vaccinations (including flu vaccination) (OR between 1.67 and 2.40), the number of consultations (spline association). The following regional-level variables were significantly associated with pneumococcal vaccination: the percentage of unemployment under 10% (OR of 1.31) and the percentage of persons under the poverty level under 20% (OR of 1.47). The general practitioner density, nurse density, percentage of white collar workers, percentage of patients with high education level were not significantly associated with rate of vaccination. After adjustment on covariates, the random effect on region was significant (variance at 0.06) with a significant predicted effect of over- and under-vaccination unexplained by covariates in several regions.
In this illustration, even if significant associations were found, the variability between department is partially explained by observed characteristics of patient and characteristics of region. The multilevel models are well suited to study the regional disparities.

1. Wyplosz B, Fernandes J, Sultan A, Roche N, Roubille F, Loubet P, et al. Pneumococcal and influenza vaccination coverage among at-risk adults: A 5-year French national observational study. Vaccine. 2022 Jul 7

MP52 A multivariate bayesian mixed-effect model (leaspy) to analyze the trajectory of cognitive decline in cadasil

Kaisaridi S.*[1], Chabriat H.[2], Tezenas Du Montcel S.[1]
[1]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013 ~ Paris ~ France, [2]Centre Neurovasculaire Translationnel-Centre de Référence CERVCO, FHU NeuroVasc, Hôpital Lariboisière, AP-HP, Université de Paris, INSERM, Unité Mixte de Recherche 1161 ~ Paris ~ France

Multivariate mixed effects models represent a promising tool to analyze longitudinal changes of multimodal data. Such models can be used to estimate long-term disease progression and to reconstruct individual trajectories accounting for the variability between patients but also between modalities. Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), the most frequent cerebral small artery disease is caused by stereotyped mutations of the NOTCH3 gene. The multifaceted clinical presentation of this disorder can be assessed using various measures. We aimed to analyze the evolution of 16 clinical scales or cognitive tests using this statistical approach in CADASIL. We analyzed data obtained during 2008 visits in 395 patients recruited at the French National Referral Centre CERVCO, using a multivariate bayesian mixed-effect model (Leaspy) to assess the temporal variability, resulting from different pace of progression and temporal offset, along with the spatial variability, influencing the variable sequence of events. Multivariate analysis allowed modelling disease progression as a 16-dimensional vector. This could account for the relationships between the dependent variables and showed variations in the acceleration rate and the starting deterioration timepoint, aligning with previous results of univariate analysis[1]. We were also able to identify different groups of evolution according to gender, education level, smoking and most importantly the mutation's position previously identified as an important determinant of disease severity only at cross-sectional level [2]. Leaspy modelling allowed to highlight distinct variations in various cognitive or clinical scores during the progression of CADASIL. The results highlight the heterogenous progression of decline in different cognitive performance. The disease trajectory is influenced by mutation location as well as gender, education, and smoking

[1] S. Brice, S. Reyes, A. Jabouley, C. Machado, C. Rogan, N. Gastellier, N. Alili, S. Guey, E. Jouvent, D. Hervé, S. Tezenas du Montcel, H. Chabriat, Neurology, 99, 2022, e1019-e1031.
[2] J. W. Rutten, B. J. Van Eijsden, M. Duering, E. Jouvent, C. Opherk, L. Pantoni, A. Federico, M. Dichgans, H. S. Markus, H. Chabriat, S. A. J. Lesnik Oberstein, Genetics in Medicine, 21, 2019, 676- 682

MP53 Performance of mixed effects models for partially clustered trials

Lange K.*[1], Sullivan T.[2], Kasza J.[3], Yelland L.[2]
[1]The University of Adelaide ~ Adelaide ~ Australia, [2]South Australian Health and Medical Research Institute ~ Adelaide ~ Australia, [3]Monash University ~ Melbourne ~ Australia

Methods for designing and analysing cluster randomised trials are well established. However, many clinical trials involve partially clustered data, where only some observations belong to a cluster. For example, neonatal trials may include infants from single or multiple births, while ophthalmology trial participants may need treatment for one or both eyes. Recently, we defined four types of partially clustered trial designs characterised by whether the clustering occurs pre- or post-randomisation and the method of randomisation for clustered observations [1]. However, the performance of analysis methods for such trials, including mixed effects models and generalised estimating equations (GEEs), have received limited attention and sample size formulas are only available for GEEs [2]. The aims of this study were to assess (1) the performance of mixed models versus GEEs for analysis of partially clustered trials, and (2) whether existing sample size formulas based on GEEs provide appropriate power for analysis via mixed models. A simulation study was conducted in R to evaluate the performance of mixed models versus GEEs for estimating the effects of treatment on a continuous outcome. We considered a maximum cluster size of 2 and simulated datasets with the sample size required to achieve 80% power according to GEE- based formulas. Simulation parameters were chosen to reflect the range of scenarios observed in practice (effect size 0-0.8, ICC 0.1-0.9, proportion of paired observations 0.015-0.7, pairs randomised using cluster, individual or balanced randomisation). Datasets were analysed using mixed effects models and GEEs with an independence or exchangeable working correlation structure. GEEs generally performed well with some exceptions when the ICC was high with individual or balanced randomisation. Performance of the mixed model was typically comparable to GEEs, though non-convergence and under-coverage occurred (maximum non-convergence rate 11%, minimum coverage rate 75%) in more extreme settings (e.g. few pairs, high ICCs). Calculating the target sample size using the exchangeable correlation GEE resulted in approximately 80-85% power for the mixed model across all scenarios. In many partially clustered trial settings, both GEEs and mixed effects models perform well. Existing sample size formulas based on GEEs may be appropriate for analysis via mixed models.

[1] K.M. Lange, J. Kasza, T.R. Sullivan, L.N. Yelland. Clinical Trials, 20(2), 2023, pp 99-110
[2] L.N. Yelland, T.R. Sullivan, D.J. Price, K.J. Lee. Statistics in Medicine, 36(8), 2017, pp 1227–1239

Poster Sessions

MP54 10-Years changes in lung function of cystic fibrosis patients in europe: different statistical methods at work

Orenti A.*[1], Adamoli A.[1], Kerem E.[2], Hatziagorou E.[3], Zolin A.[1], Ambrogi F.[1]
[1]University of Milan, Department of Clinical Sciences and Community Health, Laboratory of Medical Statistics, Biometry and Epidemiology "G. A. Maccacaro" ~ Milano ~ Italy, [2]Hadassah University Medical Centre, Hebrew University Hadassah Medical School, Department of Paediatrics and Centre for Cystic Fibrosis ~ Jerusalem ~ Israel, [3]Aristotle University of Thessaloniki, Paediatric Pulmonology and Cystic Fibrosis Unit, 3rd Paediatric Department, Hippokration Hospital ~ Thessaloniki ~ Greece

Cystic fibrosis (CF) is the most common severe autosomal recessive disease in Europe, with pulmonary insufficiency as the main cause of death. For prognosis, forced expiratory volume in 1 second percent of predicted (FEV1pp), is regarded as the best generally available measure for assessing CF lung disease. Since FEV1pp has a slightly asymmetric distribution, it is often summarized using median and quartiles. However, when fitting regression models, the results are usually provided in terms of means. The aim of the current study is to explore changes in FEV1pp during the last decade, comparing results obtained with different statistical regression methods including random effects. To estimate the difference in FEV1pp values over 2011 and 2021, data of 18756 people with CF, homozygote for F508del mutation and included in the European Cystic Fibrosis Society Patient Registry, are used. Three regression models including a random effect for patients and with FEV1pp as response variable are fitted using R software: the classical generalized estimating equations (GEE) model with Gaussian family, a linear quantile mixed model (LQMM) [1], a Generalized Additive Models for Location, Scale and Shape (GAMLSS) [2] with Normal family distribution. Two different setting are explored: in the first one the year of follow-up is included as a continuous variable, in the second one it is included using dummy variables. The results of the different models are comparable in terms of coefficient estimates. GAMLSS provides the narrower confidence interval then GEE and LQMM when year is included as a continuous variable, LQMM gives the narrower CI then GEE and GAMLSS when year is included as dummy variables. The main problem in fitting models in R software on our big dataset, is the long computational time. To obtain coefficient estimates and standard errors: 50 minutes for GAMLSS and almost 6 hours for LQMM. In conclusion, these models need to be additionally compared in detail for diagnostic measures., Further research is needed to fulfill the unmet need of providing robust regression coefficient estimates on mixed effects models on big datasets, also simulation studies mimic real world practice are necessary.

[1] M. Geraci, Linear quantile mixed models: The lqmm package for Laplace quantile regression. Journal of Statistical Software, 57(13), 2014,1-29.
[2] D.M. Stasinopoulos, R.A. Rigby, Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software, 23(7), 2007, 1-46.

MP55 Structural pathway analysis of longitudinal multinomial phenotypes

Kamruzzaman M.[1], Park T.*[2]
[1]Jagannath University ~ Dhaka ~ Bangladesh, [2]Seoul National University ~ Seoul ~ Korea, Republic of

Several statistical methods for pathway analysis have been developed to test the association between pathways and phenotypes of interest. Since pathways are highly correlated, thus a hierarchical structural component model (HisCoM) was developed to analyze all pathways in a single model and take into consideration their correlation. HisCoM was originally developed to analyze a single phenotype using only one measurement per individual. Later, it was extended to analyze multiple phenotypes (HisCoM-multi) and longitudinal phenotypes (HisCoM-GEE). These methods have been used to analyze continuous, counts, and binary phenotypes from cross-sectional, clustered, and longitudinal studies. In this study, we propose a hierarchical structural component model for pathway analysis of longitudinal multinomial phenotypes (HisCoM-RCateg). HisCoM-RCateg is proposed by combining the hierarchical structural component model and generalized estimating equations for correlated multinomial phenotypes. HisCoM-RCateg accounts for the biological hierarchy of all biomarkers and pathways into a single model. In the simulation, the proposed HisCoM-RCateg appeared to have high power than other existing methods and effectively controlled type I error for longitudinal multinomial phenotypes. To demonstrate the performance, we also applied HiscoM-RCateg to two distinct types of longitudinal omics data, namely the metabolite dataset and the metagenome dataset. HisCoM-RCateg has an advantage of taking into account the true biological hierarchical structure directly into the statistical model.

MP56 Estimation of a risk difference in a cluster randomized trial.

Pereira Macedo J.*[1], Agrinier N.[2], Minary L.[2], Kivits J.[3], Giraudeau B.[1]
[1]*Université de Tours, Université de Nantes, INSERM, SPHERE U1246 ~ Tours ~ France,* [2]*EA4360 APEMAC, Université de Lorraine, Université Paris Descartes ~ Nancy ~ France,* [3]*Université Paris Cité, ECEVE, UMR 1123, Inserm ~ Paris ~ France*

In cluster randomized trials (CRTs), clusters of individuals are randomized, rather than individuals themselves. CRT results are usually analysed using a conditional approach (with generalized linear mixed models) or a marginal one (using generalized estimating equations [GEE]). When the outcome is binary, a logit link function is classically used [1]. Hence, the results are expressed as a relative effect, with an odds ratio. Relative effects are not clearly understandable and lead to an over-optimistic appraisal of the results. The CONsolited Standards of Reporting Trials (CONSORT) statement [2] recommends reporting both relative and absolute effects, so for binary data, a risk difference (RD) is reported. Presently, we lack guidelines regarding the best way to estimate a risk difference in a cluster randomized trial. The objective was to assess the statistical properties of different methods used to estimate an adjusted risk difference from clustered data. We conducted a simulation study. We generated a binary outcome and considered a two parallel- group CRT with multiple covariates at both individual and cluster levels. Individual-level covariates were generated as confounding factors. We used a GEE to estimate the intervention effect. We considered a Gaussian distribution with an identity link function, thus directly estimating an RD. We also considered binomial and Poisson distributions with associated logit and log link functions to estimate relative intervention effects, then used the g-computation method to estimate an RD. We considered an exchangeable correlation matrix. All methods showed exemption from bias. Coverage rates ranged from 89.8% and 97.5%, with extreme values when the number of clusters was small. Ratios between the model mean and empirical standard error were close to 1, except in the situation of 5 clusters per group. The three methods had low bias and nearly identical coverage rates, which were close to the nominal value, except when the number of clusters was small. The method using the identity link with Gaussian distribution is probably to be preferred because of its ease of use.

[1] E. L. Turner et al., « Completeness of reporting and risks of overstating impact in cluster randomised trials: a systematic review », The Lancet Global Health, vol. 9, no 8, p. e1163-e1168, august 2021, doi: 10.1016/S2214-109X(21)00200-X.
[2] M. K. Campbell, G. Piaggio, D. R. Elbourne, et D. G. Altman, « Consort 2010 statement: extension to cluster randomised trials », BMJ, vol. 345, p. e5661, september 2012, doi: 10.1136/bmj.e5661.

MP57 Applying network analysis for a high-granulated description of patients` needs in maternity care from ehrs

Trutschel D.*[1], Eggenschwiler L.[1], Kuipers J.[2], Moffa G.[1], Simon M.[1]
[1]*University of Basel ~ Basel ~ Switzerland,* [2]*ETH ~ Zürich ~ Switzerland*

Maternity care, is increasingly challenged by a mismatch between staff supply and care needs, which threatens the health of mothers and newborns through inadequate care[1]. However, there is a lack of a detailed perspective describing patient-level needs and available nursing resources to identify and predict mismatches within organizations. Our aim is therefore to examine the exact nursing time spent for different care activities on maternity care patients with different clinical characteristics in a multidimensional perspective. The key strategy of our project is network methods to process high- dimensional data and characterize complex interdependence between variables in health services research. Nursing activities per mother in a large university hospital in Switzerland are recorded in detail through the 'Leistungserfassung Pflege' (LEP) system in the electronic health record. Longitudinal LEP data of around 2`500 mothers with 300`000 recordings on 150 different care activities is used to determine care activity modules through a network analysis. Partial correlation, accounting for repeated measures, provide a measure of interaction between care activities and a network structure that captures higher-order relationships [2]. Substructures of the determined network are described as care-activity modules, which are characterized by their functional meaning, indicate highly relevant features of the care demand system. Further, cluster analysis applied on data derived by the network can be used to obtain subgroups of individuals with similar care needs for later stratification. Considering nursing activities as a system of previously hidden structures of service provision has been explored, providing multidimensional views and new perspectives on care demands. This detailed picture of care activities and their mutual needs related to maternity care may help to adjust staff resources and ensure more adequate care. Identification and characterization of patient subgroups sharing similar characteristics of care demand on the one side and diagnoses or socio- demographic information on the other side would provide insights into how care needs are related to patient characteristics. Therefore, further research should address how to predict care needs of individuals based on network data on care activities to allocate resources accordingly.

[1] Nove A., ten Hoope-Bender P., Boyce M. et al., Hum Resour Health, 19(146), 2021.
[2] Zhang B., Tian Y., Zhang Z., Circ Cardiovasc Genet. 7(4), 2014, 536-547.

Poster Sessions

MP58 Deep learning based pipeline for automatic classification of breast microcalcifications

Albasini S.*[1], Gerbasi A.[2], Malovini A.[1], Quaglini S.[2], Bellazzi R.[2], Corsi F.[1]
[1]*Istituti Clinici Scientifici Maugeri IRCCS ~ Pavia ~ Italy,* [2]*University of Pavia ~ Pavia ~ Italy*

Breast microcalcifications are currently classified using the BI-RADS radiological scale. In case of suspicious microcalcifications (B3), it is recommended to perform a biopsy assessment for histopathological evaluation. However, about 70-80% of performed biopsies shows benign histology that does not require surgical treatment. Core biopsies are invasive procedures with a biological, psychological (patient discomfort), organizational and economic (for the Health Care System) costs. Therefore, accuracy's improvement in radiological classification of microcalcifications is essential. Convolutional Neural Networks (CNNs) are today the state-of-the-art AI-based models for image classification in computer vision. CNNs' main advantage lies in their ability to automatically detect hidden patterns within images able to guide their classification. We developed a fully automated pipeline based on state-of-the-art Fully Convolutional Networks (FCN) for microcalcifications' detection and classification starting from raw mammograms. Firstly, the original scan is pre-processed to enhance the contrast and remove artifacts, then microcalcifications are automatically detected by a U- net FCN trained to precisely segment the lesions of interest and finally, each cluster is classified as benign or malignant by a ResNet18 - FCN using a deep transfer-learning approach. As the last step of the proposed pipeline, state-of-the-art eXplainable AI (eXAI) methods (Grad-Cam and DeepSHAP) are used to generate maps able to show the areas of the images where the network is mostly focusing its attention to make the final classification. Preliminary results on publicly available datasets (INbreast and CBIS-DDSM) show a classification accuracy and area under the ROC curve (IC 95%) of 0.83 (0.77 – 0.89) and 0.89 (0.84 – 0.94), respectively, on the test set (n = 170). Although it is designed to be fully automated and work with any input scan, the proposed pipeline is easily customizable to meet radiologists needs. Moreover, the visual inspection of eXAI maps overcomes the limitation of dealing with a completely black-box approach by offering the user the possibility of intuitively assessing the degree of reliability of each classification result. The proposed pipeline is a promising interpretable decision support system able to guide the diagnosis and possibly helpful for gaining new insights on the disease mechanisms.

[1] Breast Cancer Statistics, https://www.wcrf.org/cancer-trends/breast-cancer-statistics/.
[2] L. Caplan, Delay in breast cancer: implications for stage at diagnosis and survival, Frontiers inpublic health 2 (2014) 87.
[3] S. O'Grady, M. Morgan, Microcalcifications in breast cancer: From pathophysiology to diagnosisand prognosis, Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1869 (2) (2018) 310–320.doi:https://doi.org/10.1016/j.bbcan.2018.04.006.
[4] S. Azam, M. Eriksson, A. Sjolander, M. Gabrielson, R. Hellgren, K. Czene, P. Hall, Mammographicmicrocalcifications and risk of breast cancer, British journal of cancer 125 (5) (2021) 759–765.
[5] M. Ciecholewski, Microcalcification segmentation from mammograms: A morphological approach,Journal of digital imaging 30 (2) (2017) 172–184.
[6] M. Melloul, L. Joskowicz, Segmentation of microcalcification in x-ray mammograms using entropythresholding, in: CARS 2002 computer assisted radiology and surgery, Springer, 2002, pp. 671–676.
[7] L. Hussain, W. Aziz, S. Saeed, S. Rathore, M. Rafique, Automated breast cancer detection usingmachine learning techniques by extracting different feature extracting strategies, in: 2018 17th IEEEInternational Conference On Trust, Security And Privacy In Computing And Communications/12thIEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), IEEE,2018, pp. 327– 331.

MP59 Prediction of neurodevelopment at two years of age among very preterm infants in the île-de-france region

Anzelin L.*[1], Thiébaut A.[2], Desplanques L.[3], Granier M.[4], Leloup L.[5], Pierrat V.[6], Lapillonne A.[7], Tubert-- Bitter P.[2], Ahmed I.[2], Hanf M.[1]
[1]Data science department, Sesan ~ Paris ~ France, [2]Université Paris-Saclay, UVSQ, INSERM, CESP High Dimensional Biostatistics Team ~ Villejuif ~ France, [3]Perinatal, Child and Women's Health Department - Health Promotion and Inequalities Reduction Department - ARS Île-de-France ~ Saint-Denis ~ France, [4]Neonatology and Neonatal Intensive Care Unit, Centre Hospitalier Sud Francilien ~ Corbeil-Essonnes ~ France, [5]Réseau Pédiatrique Sud et Ouest Francilien - Association pour le Suivi des Nouveau-nés à Risque ~ Clamart ~ France, [6]Réseau Périnatal du Val de Marne ~ Créteil ~ France, [7]Université Paris-Descartes, Hôpital universitaire Necker-Enfants Malades ~ Paris ~ France

Monitoring of very preterm infants is essential for early detection of developmental anomalies. The aim of this study was to build and evaluate a prediction model of suboptimal neurodevelopment at two years of age in very preterm infants.
The study population consisted of infants who were born before 33 weeks of gestational age in the Île- de-France region and enrolled in a large, prospective, population-based open cohort, HYGIE-SEV, between October 2015 and November 2020. Suboptimal neurodevelopment was defined as having at least one motor or cognitive impairment as assessed at the two-year visit. A set of 56 exploratory variables describing perinatal and neonatal characteristics, socio-economic environment and health system efficiency was collected at enrollment and considered for the prediction using Random Forests. Data preprocessing included imputation of missing values, data enrichment, centering and scaling, one-hot encoding and a smoothed bootstrap approach for class imbalance. The dataset was split into train (80%) and test (20%) subsets. In the train subset, hyperparameters were tuned using maximum entropy grid search and ten-fold cross-validation, with the area under the ROC curve (AUC) as the criterium to maximize. The final AUC and importance measures of all variables were computed in the test subset. Among 6,868 very preterm infants who attended the two-year visit, 2,135 (31.1%) were classified as having a suboptimal neurodevelopment. The AUC of the final model on the test subset was 78.0%. The ten variables with highest importance measures were neurologic exploration, network of inclusion, department of family residence, birth maternity, distance between the referring physician and the place of residence, economic resources, birth month, father's professional situation, length of neonatal hospitalization and surgical intervention at birth. Using machine learning was helpful to gain insight into predictors of neurodevelopment in very preterm infants. In addition to clinical variables previously identified in the literature, contextual variables were also highlighted to predict neurodevelopment. These findings may help health professionals in the timely tailoring of follow-up care for very preterm infants who are at risk of developing neurodevelopmental anomalies.

Poster Sessions

MP60 Understanding random forests from a statistician's point of view: a simulation study

Barreñada L.*[2], Boulesteix A.[1], Van Calster B.[2]
[1]Biometry in Molecular Medicine, LMU Munich ~ Munich ~ Germany, [2]Department of development and Regeneration, KU Leuven ~ Leuven ~ Belgium

Random forests have become popular for clinical risk prediction modeling. In a case study on predicting ovarian malignancy, we observed training c-statistics close to 1. Although this suggests overfitting, performance was competitive on test data. We aimed to understand the behavior of random forests by (1) visualizing data space for the case study and (2) a simulation study.
Visualization of data space suggested that the model learned 'spikes of probability' around training set events . A cluster of events created a big peak (signal), isolated events local peaks (noise). The simulation study included 48 logistic data generating mechanisms (DGM) with event fraction of 0.2, varying the predictor distribution (binary vs continuous), the predictors (4 true, 4 true and 12 noise, 16 true predictors), the correlation between predictors (0 vs 0.4), the true c-statistic (0.75 vs 0.90) and the strength of true predictors (equal vs unequal). For each DGM, 1000 training datasets of size 200 or 4000 were simulated and random forest models trained with minimum node size 2 or 20 using the ranger R package. Model performance was evaluated on large test datasets (N=100.000). Median training c-statistics were between 0.97 and 1 unless there were 4 binary predictors or 16 binary predictors and minimum node size 20. In 114/192 scenarios with median training c-statistic ≥0.99, the discrimination loss (difference between true c-statistic and median test c-statistic) was small: median 0.04 (range 0.00-0.13). across all scenarios, the Spearman correlation between median train c-statistic and discrimination loss was 0.62. Median test c-statistics were higher with higher events per variable, higher minimum node size, and binary predictors. Median training calibration slopes ranged between 1.10 and 19.4. Median test calibration slopes ranged between 0.45 and 2.34, and were not related to median training slopes (Spearman correlation -0.11). Median test slopes were higher with higher true c-statistic, higher minimum node size, and higher sample size. Random forests learn local probability peaks, often yielding near perfect training c-statistics. Our results go against the recommendation to use fully grown trees in random forest models. Calibration performance was whimsical.

[1] E. W. Steyerberg, Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-16399-0.
[2] A. J. Wyner, M. Olson, J. Bleich, and D. Mease, 'Explaining the Success of AdaBoost and Random Forests as Interpolating Classifiers'.

MP61 Holographic microscopy data analysis with deep learning for the detection of antimicrobial mechanism of action

Courbon B.*[1], Dixneuf S.[1], Sedaghat Z.[1], Vedrine C.[2]
[1]BIOASTER ~ Lyon ~ France, [2]BIOASTER ~ Paris ~ France

Facing the rapid increase of antimicrobial resistance, there is urge in developing new antimicrobials, in particular with new mechanisms of action (MoA). In the current work, we propose an innovative technology to classify the MoA of an antimicrobial, and possibly detect its novelty. This technology is based on the combination of dynamic Digital Inline Holographic Microscopy (DIHM) and Deep Learning (DL). DIHM provides a label-free, high-throughput time-resolved screening of bacteria morphology to reveal phenotypic responses to antibiotics. DL techniques are powerful tools to extract discriminative features from sequences of images and classify them. We assess the performance of our approach in the challenging context of time-lapse single-cell holographic imaging, characterized by the high dimensionality and biological variability of data. Incubations of Escherichia coli with antibiotics are monitored using time-lapse DIHM for 2 hours with time resolution of 3 minutes. We include 22 antibiotics in our database, corresponding to 5 MoA classes, as well as control samples without antibiotic. We use up to 3 replicates per antibiotic to check for experimental reproducibility. Holograms are reconstructed with a back-propagation algorithm, then phase images are segmented into patches centered around individual bacteria tracked over time. Our final dataset contains around 2000 time-series of bacteria images. We first develop DL models to classify these time-series among 6 MoA classes. Our models are based on Convolutional Neural Networks. The time dimension of the data can be analyzed either using 3- dimensional convolutional kernels (CNN3D) or adding a recurrent neural network on top of time- distributed 2-dimensional convolutional networks (CRNN). Networks hyperparameters are optimized, then algorithms performance is assessed using 10-fold cross-validation. We obtain a 74% classification accuracy at the bacteria-level. Moreover, 89% of the samples are correctly classified by aggregating the predictions for each of their bacteria. Finally we show how our models can be used for the novelty assessment of the MoA of a candidate antibiotic. The combination of Holographic Microscopy and Deep Learning provides a good identification of the MoA of an antibiotic. It represents a promising technology for the screening of new antimicrobials provided that a large and complete database of known molecules is available.

[1] Ouyang X, Hoeksma J, Lubbers RJM, Siersma TK, Hamoen LW, den Hertog J. Classification of antimicrobial mechanism of action using dynamic bacterial morphology imaging. Sci Rep. 12:11162 (2022).
[2] Mahé P, El Azami M, Degout-Charmette E, Sedhagat Z, Josso Q, Rol F. Method for classifying a sequence of input images representing a particle in a sample over time. WO2022084616A1 (2022).

Poster Sessions

MP62 Personalized diagnosis in suspected myocardial infarction: the artemis study

Neumann J.T.[2], Ziegler A.[1], Twerenbold R.[2], Ojeda F.[2], Di Carluccio E.*[1], Aldous S.[3], Allen B.[4], Apple F.[5], Babel H.[1], Christenson R.[6], Cullen L.[7], Doudesis D.[8], Ekelund U.[9], Giannitsis E.[10], Greenslade J.[7], Inoue K.[11], Jernberg T.[12], Kavsak P.[13], Keller T.[14], Lee K.K.[8], Lindahl B.[15], Lorenz T.[2], Mahler S.[16], Mills N.[8], Mokhtari A.[17], Parsonage W.[18], Pickering J.[19], Pemberton C.[20], Reich C.[21], Richards M.[19], Sandoval Y.[22], Than M.[23], Toprak B.[2], Troughton R.[20], Worster A.[24], Zeller T.[2], Blankenberg S.[2]
[1]Cardio-CARE, Medizincampus Davos ~ Davos ~ Switzerland, [2]Department of Cardiology, University Heart and Vascular Center, University Medical Center Hamburg-Eppendorf, Hamburg, Germany ~ Hamburg ~ Germany, [3]Department of Cardiology, Christchurch Hospital ~ Christchurch ~ New Zealand, [4]Department of Emergency Medicine, College of Medicine, University of Florida ~ Gainesville, FL ~ United States of America, [5]Departments of Laboratory Medicine and Pathology, Hennepin Healthcare/HCMC and University of Minnesota ~ Minneapolis, MN ~ United States of America, [6]Department of Pathology, University of Maryland School of Medicine ~ Baltimore, MD ~ United States of America, [7]Department of Emergency Medicine, Royal Brisbane and Women's Hospital ~ Herston, Queensland ~ Australia, [8]BHF Centre for Cardiovascular Science, University of Edinburgh ~ Edinburgh ~ United Kingdom, [9]Lund University, Skåne University Hospital, Department of Internal and Emergency Medicine ~ Lund ~ Sweden, [10]Department of Cardiology, Heidelberg University Hospital ~ Heidelberg ~ Germany, [11]Juntendo University Nerima Hospital ~ Tokyo ~ Japan, [12]Department of Clinical Sciences, Danderyd University Hospital, Karolinska Institutet ~ Stockholm ~ Sweden, [13]Department of Pathology and Molecular Medicine, McMaster University ~ Hamilton, Ontario ~ Canada, [14]Department of Cardiology, Kerckhoff Heart and Thorax Center ~ Bad Nauheim ~ Germany, [15]Department of Medical Sciences and Uppsala Clinical Research Center, Uppsala University ~ Uppsala ~ Sweden, [16]Department of Emergency Medicine, Wake Forest School of Medicine ~ Winston-Salem, NC ~ United States of America, [17]Department of Internal Medicine and Emergency Medicine and Department of Cardiology, Lund University, Skåne University Hospital ~ Lund ~ Sweden, [18]Australian Centre for Health Service Innovation, Queensland University of Technology ~ Kelvin Grove ~ Australia, [19]Department of Medicine, University of Otago Christchurch and Emergency Department, Christchurch Hospital ~ Christchurch ~ New Zealand, [20]: Department of Medicine, Christchurch Heart Institute, University of Otago ~ Otago ~ New Zealand, [21]Department of Cardiology, Heidelberg University Hospital ~ Heidelberg ~ Germany, [22]Minneapolis Heart Institute, Abbott Northwestern Hospital, and Minneapolis Heart Institute Foundation ~ Minneapolis, MN ~ United States of America, [23]Department of Medicine, University of Otago Christchurch and Emergency Department, Christchurch Hospital ~ Christchurch ~ New Zealand, [24]Division of Emergency Medicine, McMaster University ~ Hamilton, ON ~ Canada

In suspected myocardial infarction (MI), guidelines recommend using high-sensitivity cardiac troponin (hs-cTn)-based approaches. These require fixed assay-specific thresholds and timepoints, without directly integrating clinical information. Using machine-learning techniques including hs-cTn and clinical routine variables, we developed and validated a model to estimate the individual probability of MI, while allowing for numerous hs-cTn assays. The aim of this presentation is to describe the approach for developing and validating this diagnostic model In 2,575 patients presenting to the emergency department with suspected MI, two ensembles of machine-learning models using single or serial concentrations of six different hs cTn assays were derived to estimate the individual MI probability. Twelve routinely available variables including age, sex, cardiovascular risk factors, electrocardiography, and hs-cTn were included in the ARTEMIS models. First, multiple imputation was performed. Second, full models were trained using ten-fold cross-validation. Third, variable selection was performed by using the information from all hs-cTn models. Fourth, reduced models were trained using ten-fold cross-validation. Fifth, a superlearner with equal weights was estimated. Model performance was assessed using the logLoss. It was validated in an external cohort with 1,688 patients and tested for global generalizability in 13 international cohorts with 23,411 patients after calibration. Model performance of the reduced models was superior to the full model and superior to the hs-cTn only-model in the training data. It performed best on the validation and generalization data, and it was significantly better than the hs-cTn only-model. Models based on the superlearner generally outperformed the single learners. Using a single hs-cTn measurement, the ARTEMIS model allowed direct rule-out of MI with very high and similar safety but up to tripled efficiency compared to the guideline-recommended strategy. We developed and validated diagnostic models to accurately estimate the individual probability of MI, which allow for variable hs-cTn use and flexible timing of resampling. The development of models using different hs-cTn assays led to substantial greater stability in the model performance due to improved variable selection properties. Furthermore, the use of an equal-weights superlearner further increased the stability of the machine learning models.

1. Nawar EW, Niska RW, Xu J. National Hospital Ambulatory Medical Care Survey. 2005 emergency department summary. Advance data. 2007(386):1-32.
2. Westermann D, Neumann JT, Sorensen NA, Blankenberg S. High-sensitivity assays for troponin in patients with cardiac disease. Nat Rev Cardiol. 2017;14(8):472-83.
3. Thygesen K, Alpert JS, Jaffe AS, Chaitman BR, Bax JJ, Morrow DA, et al. Fourth Universal Definition of Myocardial Infarction (2018). J Am Coll Cardiol. 2018.
4. Collet JP, Thiele H, Barbato E, Barthelemy O, Bauersachs J, Bhatt DL, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST- segment elevation. Eur Heart J. 2021;42(14):1289-367.
5. Writing Committee M, Gulati M, Levy PD, Mukherjee D, Amsterdam E, Bhatt DL, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol. 2021;78(22):e187-e285.
6. Sandoval Y, Apple FS, Mahler SA, Body R, Collinson PO, Jaffe AS, et al. High-Sensitivity Cardiac Troponin and the 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guidelines for the Evaluation and Diagnosis of Acute Chest Pain. Circulation. 2022;146(7):569-81.
7. Writing C, Kontos MC, de Lemos JA, Deitelzweig SB, Diercks DB, Gore MO, et al. 2022 ACC Expert Consensus Decision Pathway on the Evaluation and Disposition of Acute Chest Pain in the Emergency Department: A Report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol. 2022.

MP63 Regularization approaches in clinical biostatistics: a review of methods and their applications

Friedrich S.*[1], Groll A.[2], Ickstadt K.[2], Kneib T.[3], Pauly M.[2], Rahnenführer J.[2], Friede T.[4]
[1]University of Augsburg ~ Augsburg ~ Germany, [2]TU Dortmund University ~ Dortmund ~ Germany, [3]Georg-August-University Göttingen ~ Göttingen ~ Germany, [4]University Medical Center Göttingen ~ Göttingen ~ Germany

A range of regularization approaches have been proposed in the data sciences to overcome overfitting, to exploit sparsity or to improve prediction. Using a broad definition of regularization, namely controlling model complexity by adding information in order to solve ill-posed problems or to prevent overfitting, we review a range of approaches within this framework including penalization, early stopping, ensembling and model averaging. Although there is a growing literature on regularization with a wealth of techniques being available, it is currently largely unknown to what extent these methods are actually used in clinical medicine and what type of problems are addressed by their use. To assess the extent to which these approaches are used in medicine, we systematically reviewed recent volumes of three journals publishing in general medicine, namely the Journal of the American Medical Association (JAMA), the New England Journal of Medicine (NEJM) and the British Medical Journal (BMJ). Moreover, we provide an overview of methods that fall into this category, discuss the context of their application and give examples on statistical software packages. Finally, aspects of the practical implementation of these methods are discussed in a data example on prostate cancer. We demonstrate the implementation of six different regularization approaches, namely a classification tree (CART), a random forest, subset selection, rigde regression, LASSO, and elastic net and compare them to standard logistic regression by means of area under the curve (AUC) and the mean classification error (MCE). In this example, standard logistic regression is outperformed by the regularization methods. Our literature review revealed that regularization approaches are rarely applied in practical clinical applications, with the exception of random effects models. As demonstrated by our application examples, however, statistical software is available and implementation is straightforward. In situations where also other approaches work well, the only downside of the regularization approaches is increased complexity in the conduct of the analyses. Hence, we suggest a more frequent use of regularization approaches in medical research.[1]

[1] Friedrich et al, Statistical Methods in Medical Research, Vol. 32(2), 2023, 425-440

MP64 Effects of irregular sleep behavior on physiological and perceived health in shift workers

Jeon S.*[1], Kim J.K.[2], Song Y.[2], Chung S.[4], Ahn Y.[5], Suh S.[3]
[1]Mokwon University ~ Daejeon ~ Korea, Republic of, [2]Korea Advanced Institute of Science and Technology ~ Daejeon ~ Korea, Republic of, [3]Sungshin Women's University ~ Seoul ~ Korea, Republic of, [4]University of Ulsan College of Medicine ~ Seoul ~ Korea, Republic of, [5]Yonsei University Wonju College of Medicine ~ Wonju ~ Korea, Republic of

Sleep is a fundamental need that plays a vital role in maintaining overall health and well-being. Adequate and quality sleep enables the body to repair and regenerate, bolster the immune system, and promote health. On the other hand, poor sleep quality is known to lead to mood disorders, anxiety, and depression, which can negatively impact mental. Moreover, the impact of poor sleep quality can go beyond mental health to affect physical health, increasing the risk of accidents and injuries, particularly in high-risk professions such as healthcare, transportation, and emergency. Emergency shift workers, such as firefighters, are particularly vulnerable to sleep disorders and associated health risks, given their irregular sleep-wake patterns. Limitations have existed in studies investigating the health risks posed by irregular sleep patterns in shift workers due to their reliance on self-reported data. To better understand and mitigate the health risks of irregular sleep patterns in emergency shift workers, this study aims to acquire direct and indirect information on real and perceived sleep behavior through wearable devices and to implement personalized intervention based on the findings, which can lead to improve sleep quality, cognitive health, mental health, and overall quality of life. The measured human activity data was converted into physiological condition indicators such as homeostatic sleep pressure, circadian rhythm, and alertness by mathematical model of sleep regulation. The findings of predictive machine learning showed that sleep quality is influenced not only by sleep duration, sleep efficiency and sleep-related physiological indicators, but also by work shift characteristics. Moreover, the study revealed that interventions can be effective in improving sleep quality among emergency shift workers. Additionally, the research indicated that quality sleep can improve daytime physical rhythm and reduce perceived shift work disorder, which highlights the importance of prioritizing good sleep habits and implementing personalized sleep behavior interventions. In conclusion, maintaining good health and well-being necessitates adequate and quality sleep, which is beneficial for both physical and cognitive health. Hong, J., Choi, S. J., Park, S. H., Hong, H., Booth, V., Joo, E. Y., & Kim, J. K. (2021). Personalized sleep-wake patterns aligned with circadian rhythm relieve daytime sleepiness. Iscience, 24(10), 103129. Jang, E. H., Hong, Y., Kim, Y., Lee, S., Ahn, Y., Jeong, K. S., ... & Suh, S. (2020). The development of a sleep intervention for firefighters: the FIT-IN (Firefighter's therapy for insomnia and nightmares) Study. International journal of environmental research and public health, 17(23), 8738. Ryu, J. Y., Cho, M. S., Kim, J. S., & Lee, C. W. (2018). A study on the effects of job characteristics by type of shift work system on physical and psychological status of firefighters in Korea. The Journal of Public Policy and Governance, 12(3), 197-231.

MP65 Using deep learning for time series regression models in health: a simulation study

Mahmudimanesh M.*, Bahrampour A.
Department of Biostatistics and Epidemiology, Faculty of Health, Modeling in Health Research Center, Institute for Future Studies in Health ~ Kerman ~ Iran, Islamic Republic of

Finding the best model to forecast time series data is important in all sciences, especially medicine. Artificial intelligence and deep learning models are increasingly used in such data today. The aim of this study is to find the best deep learning model for time series data on heart disease deaths as well as simulated data. There were real data on heart mortality in Tehran city as a response variable and air pollution as an explanatory variable. Also, simulated data from the Autoregressive integrated moving average with explanatory variables (ARIMAX) model to determine the best model based on deep learning methods for fitting time series regression data. To reach the goal, we reviewed all models for fitting such data before using the simulated data from the ARIMAX model to determine the best model. This hybrid model, combining convolution neural networks and long short-term memory (LSTM), performed well on both training and testing datasets with a MSE values of 0.0107 for training and 0.0150 for testing. As a result of this paper, it was concluded that recurrent neural network and LSTM provide appropriate results for univariate time series in comparison to other models. Additionally, the CNN-LSTM model is a good choice if other variables are included in the model and the goal is to fit a time series regression.

1. Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks:: The state of the art. International journal of forecasting. 1998;14(1):35-62.
2. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50:159-75.
3. Vakili M, Taheri M, Sartipzadeh N. Study of risk factors for acute myocardial infarction in patients registered at shahid Sadooghi hospital in Yazd: a case-control study. Quarterly J Sabzevar Univ Medical Sci. 2015;22(1):144-22.
Asadabadi A, Bahrampour A, Haghdoost A. Prediction of breast cancer survival by logistic regres4. sion and artificial neural network models. Iranian Journal of Epidemiology. 2014;10(3):1-8.
5. Illingworth WT, editor Beginner's guide to neural networks. Proceedings of the IEEE National Aerospace and Electronics Conference; 1989: IEEE.
6. Aggarwal CC. Neural networks and deep learning. Springer. 2018;10:978-3.
7. Freeman BS, Taylor G, Gharabaghi B, Thé J. Forecasting air quality time series using deep learning. Journal of the Air & Waste Management Association. 2018;68(8):866-86.

MP66 A systematic review on machine learning techniques for survival analysis in cancer

O"Donnell A.*[1], Wolsztynski E.[1], Cronin M.[1], Moghaddam S.[2]
[1]University College Cork ~ Cork ~ Ireland, [2]University of Limerick ~ Limerick ~ Ireland

Introduction and Objective(s): Machine learning (ML) methodologies for survival analysis have been the topic of much research in recent years, in parallel with the increased availability of high-dimensional data. Many ML classification techniques have been adapted for analysis of censored data in time- to-event studies, including cancer research, with varied levels of performance across cancer types and cohorts [1-3]. A systematic review was undertaken to (i) determine how machine learning methodologies for survival analysis in cancer compare to traditional statistical methods, and (ii) what machine learning methodology tend to demonstrate superior predictive performance for survival analysis in cancer. **Method(s) and Results:** The protocol was register with the international prospective register of systematic reviews (PROSPERO) [4]. Five mainstream databases of research publications were searched for works reporting on the use of ML for continuous survival analysis in cancer cohorts, yielding 2,968 results after the removal of duplicates. Title and abstracts as well as the full-text of the studies obtained will be screened for inclusion in the systematic review by two authors, independently. Studies that only report on classification models will be excluded as well as review papers and animal experimental studies. Data analysis includes a tabular and narrative synthesis of the findings structured around the methodological approach investigated. It also includes a report of the outcome measures and experimental details (e.g. accuracy of survival prediction) as reported in the publications. **Conclusions:** This presentation reports on the complete results of the systematic review in order to provide statisticians and clinicians with an insight into the performance of machine learning for cancer survival analysis.

[1] H. Wang and L. Zhou, "Random survival forest with space extensions for censored data," Artifical Intelligence in Medicine, vol. 79, no. 1, pp. 52-61, 2017.
[2] I. K. Omurlu, M. Ture and F. Tokatli, "The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer," Expert Systems with Applications, vol. 36, no. 4, pp. 8582-8588, 2009.
[3] A. Moncada-Torress, M. C. van Maaren, M. P. Hendriks, S. Siesling and G. Geleijnse, "Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival," Scientific Reports, vol. 11, no. 1, p. 6968, 2021.
[4] A. O'Donnell, E. Wolsztynski and S. Moghaddam, "A systematic review on machine learning techniques for survival analysis in cancer," 23 February 2023. [Online]. Available: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=391624. [Accessed 28 March 2023].

Poster Sessions

MP67 Using machine learning to predict multi-class functional outcomes and death 3 months after stroke in sweden

Otieno J.A.*[2], Häggström J.[2], Darehed D.[1], Eriksson M.[2]
[1]Department of Public Health and Clinical Medicine, Sunderby Research Unit, Umeå University ~ Umeå ~ Sweden,
[2]Department of Statistics, USBE, Umeå University ~ Umeå ~ Sweden

Globally, stroke is the third-leading cause of mortality and disability combined.[1] Accurate prediction of post-stroke disability could provide guidance for the continued care and rehabilitation planning. We aimed to develop and compare the performance of three supervised machine learning algorithms with the traditional logistic regression (LR) model in predicting disability and death 3 months after stroke based on the modified Rankin Scale (mRS) using routinely collected data. We also explored the explainability of these algorithms by revealing the most important variables and how they contribute to the prediction. All adult patients registered with a stroke in the Swedish Stroke Register between 2015 to 2020 were included. Prognostic factors comprised amongst others age, sex, cardiovascular risk factors, medications before stroke, stroke subtype, and stroke severity measured by the National Institutes of Health Stroke Scale (NIHSS). Missing data of NIHSS were handled using the Multivariate Imputation by Chained Equations technique while assigned a new category for missing values in other variables. Feature scaling and label encoding were based on Min-Max and one-hot encoding methods, respectively. The main outcome for prediction was mRS at 3 months after stroke (0-2 independent, 3-5 dependent, and 6 dead). Classifiers included support vector machines, artificial neural networks (ANN), eXtreme Gradient Boosting (XGBoost), and LR. They were trained and tested on 75% and 25% of the dataset, respectively, and their predictive performances were assessed using different performance metrics. The predictions were explained by SHAP values.[2] 85.8% had ischemic stroke and 53.3% were male. The mean age at admission was 75.8 years. The overall accuracy was similar in all models, with a value of more than 68% (95%CI:68-70%). The ANN and XGBoost classifiers performed significantly better than the traditional LR in classifying dependence, with an F1-score of 0.603 (95%CI:0.594-0.611) and 0.577 (95%CI:0.568-0.586), respectively, compared to the LR model (0.554, 95%CI:0.545-0.563). Death was most strongly associated with NIHSS, whereas functional independence was related to male sex, stroke alerts, and lipid lowering drugs. Our ANN and XGBoost classifiers showed a modest improvement in prediction performance compared to LR using routinely collected data. Existing methods can be used for explainability of these algorithms.

[1] Feigin VL, Stark B, Johnson C, Roth G, Bisignano C, Abady G. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021; 20: 795-820.
[2] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

MP68 Post-selection confidence bounds for prediction performance

Rink P.*, Brannath W.
Competence Center for Clinical Trials, University of Bremen ~ Bremen ~ Germany

Machine learning is a promising tool for improving the efficiency and quality of clinical research, and, accordingly, is playing an increasingly important role, for instance, in predicting complex outcomes from a potentially large number of competing models. Model selection and the assessment of the generalization performance are critical tasks that need careful consideration, since the inferred decisions may affect patients' lives and health directly. Typically, to control the type 1 error, model selection and evaluation are strictly separated endeavors, splitting the sample at hand into a training, validation, and evaluation set, and only compute a single confidence interval for the prediction performance of the final selected model. This however comes at the cost of possibly using the data not to its fullest potential, especially when only little data is available to perform both tasks. By allocating a greater fraction of the data towards model selection the goodness assessment gets less reliable, and allocation of a greater fraction towards goodness assessment poses the risk of selecting a sub-par prediction model. We propose an algorithm that resolves this problem reliably. We compute valid lower confidence bounds for multiple models that have been selected based on their prediction performances in the evaluation set by interpreting the selection problem as a simultaneous inference problem. We use bootstrap tilting and a maxT-type multiplicity correction. The approach is universally applicable for any combination of prediction models, any model selection strategy, and many prediction performance measure, for example accuracy or the AUC. Our proposed approach yields lower confidence bounds that are at least comparably good as bounds from standard approaches, and that reliably reach the nominal coverage probability. In addition, especially when sample size is small, our proposed approach yields better performing prediction models than the default selection of only one model for evaluation does. [1]
Deviating from the default approach to strictly separate selection and evaluation, and using the same data for selection and evaluation plus an appropriate multiplicity correction instead, may result in comparatively larger lower confidence bounds and better performing models. [2]

[1] Rink, Pascal and Brannath, Werner (2022). Post-Selection Confidence Bounds for Prediction Performance. arXiv preprint, https://doi.org/10.48550/arxiv.2210.13206. Submitted to the Springer Machine Learning Journal.
[2] Westphal, Max and Brannath, Werner (2020). Evaluation of multiple prediction models: A novel view on model selection and performance assessment. Statistical Methods in Medical Research 29(6): 1728–1745. https://doi.org/10.1177/0962280219854487

MP69 Are lifestyle factors the most important predictors of common mental disorders? A systematic review

Todd E.*[1], Orr R.[1], Loughman A.[1], Khosravi A.[2], Jacka F.[1], Dawson S.[1]
[1]Deakin University, IMPACT – the Institute for Mental and Physical Health and Clinical Translation, Food & Mood Centre, School of Medicine, Barwon Health ~ Geelong ~ Australia, [2]Institute for Intelligent Systems Research and Innovation, Deakin University ~ Waurn Ponds ~ Australia

Depression and anxiety, known as common mental disorders (CMDs), affect millions worldwide and impose a huge cost on individuals and communities [1]. Recently, machine learning (ML) has been used to interrogate large datasets and develop models that predict prevalent or incident CMDs based on lifestyle, demographic, or biological risk factors. The aim of this review is to synthesise the existing literature from such studies and assess whether lifestyle factors are consistently reported to be more predictive of CMDs than less-modifiable factors, such as demographics. This knowledge could provide new avenues for prevention through targeted lifestyle interventions and/or future risk prediction tools. The systematic review will be performed in accordance with the PRISMA statement and registered with PROSPERO. Databases searched will include MEDLINE, EMBASE, PsycInfo, IEEE Xplore, and Engineering Village. The search strategy will identify studies that use ML to predict depression and/or anxiety in adults. Studies will be included if they use a method that produces a measure of variable 'importance' (e.g. Shapley Additive Explanations, Percent Increase in Mean Squared Error) and differentiates people with CMDs from controls. Search terms include 'common mental disorders' (e.g. depression, anxiety) in the title, 'diagnosis or differentiation' (e.g. diagnose, screen) and ML methods that perform 'variable importance analyses' (e.g. random forest, XGBoost) in the abstract. The search will be limited to English language articles with no date restrictions. Anticipated Results: Our study will generate a ranked list of predictors that will determine whether modifiable lifestyle factors are more important for predicting CMDs than less-modifiable factors. We will present our protocol, available preliminary results, and background information at the meeting.

[1] Institute of Health Metrics and Evaluation, Global Health Data Exchange (2019), https://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019- permalink/87120109249ceec600942153d36ee021

Poster Sessions

MP70 Classification of metabolic syndrome using arterial pulse wave via machine learning approaches

Yim M.H.*
Korea Institute of Oriental Medicine ~ Daejeon ~ Korea, Republic of

Metabolic syndrome (MS) is a commonly occurring chronic metabolic disorder and is strongly associated with cardiovascular mortality [1]. Several studies based on noninvasive measurements such as anthropometric measures and sociodemographic characteristics have been conducted to predict or classify MS. However, studies based on variables extracted from pulse wave to predict or classify MS have not been conducted yet. This study aimed to classify MS using pulse wave variables through several machine learning approaches. A total of 215 women who were recruited from November 2021 to July 2022 at the Cheonan Oriental Hospital of Daejeon University, Republic of Korea were included in this study. The MS classification models using pulse wave variables were built through six machine learning approaches: elastic net, k- nearest neighbor, random forest, support vector machine, extreme gradient boosting, and neural network. Relative variable importance was calculated to identify the contribution of individual pulse wave variables selected from the final six models. The performance of each model was evaluated using nested cross-validation with 5 outer and 5 inner splits [2], resulting in accuracy, Kappa, precision, F1 score, sensitivity, specificity, and areas under the receiver operating characteristic curve (AUC) value along with individual 95% confidence intervals (CIs). The optimal threshold was determined by the Youden index, and CI was calculated using 2000 bootstrap replicates. The model using neural network approach reported the highest AUC value, Kappa, and F1 score (AUC = 0.849 [95% CI, 0.773-0.907]; Kappa = 0.420 [95% CI, 0.302-0.545]; F1 score = 0.544 [95% CI, 0.429- 0.652]). The models using K-nearest neighbor, random forest, and support vector machine showed relatively low AUC values of 0.783 (0.702-0.854), 0.777 (0.701-0.851), and 0.772 (0.685-0.846), respectively. The influential pulse wave variables were height of tidal wave, standard deviation of pulse rate, body surface area, 3-dimensional pulse volume, etc. These results showed the potential of pulse wave variables for classification of MS using noninvasive measurement. In addition, classification using machine learning algorithms can help clinicians gain insights and make clinical decisions.

[1] J.-P. Després, I. Lemieux, Abdominal obesity and metabolic syndrome, Nature, vol. 444, no. 7121, 2006, p. 881–887.
[2] Bates, Stephen, Trevor Hastie, Robert Tibshirani, Cross-validation: what does it estimate and how well does it do it?, arXiv preprint arXiv:2104.00673, 2021.

Poster Sessions

MP71 Advanced statistical methods of handling ordinal missing data

Aluko O.*[1], Mwambi H.[2]
[1]University of the Free State ~ Bloemfontein ~ South Africa, [2]University of KwaZulu-Natal ~ Pietermaritzburg ~ South Africa

As of 2011, approximately 33 million persons have been diagnosed with human immunodeficiency virus (HIV) as reported by the national heart, lung, and blood institute (NHLBI). The rate of survival has improved and turned HIV into a chronic disease based on the development of antiretroviral therapy and other clinical procedures. However, the continuous loss of immunity from HIV-infected individuals has resulted in the growing evolution of comorbidity diseases, remarkably increasing the death rate among the infected individuals. . The interest is in the description of the axial emphysema distribution outcomes which were classified as ordinal. For clarity purposes, ordinal data property needs to be applied to the ordinal response categories. Using the methods of analyzing categorical data to ordered categorical data may lead to loss of information. Using models specifically developed for ordinal categorical data has the advantage of considering the ordering pattern of the response categories. The use of methods from a simple binomial distribution to more flexible distributions that allow for over-dispersion. However, longitudinal data is one of the methods used for analyzing ordered categorical responses with missing data either in covariate or outcome/ both. The models are direct likelihood (DL), mixed-effects proportional odds, multiple imputations generalized estimating equations (MI-GEE), and ordinal negative binomial (ONB). In addition, we explore the simulation study of different sizes of monotone missing data from four visits to mimic the original dataset. The simulation study was conducted to ascertain the effectiveness and evaluate the properties of the selected methods to handle missing data. The DL method was introduced by [3] as a better method to handle missing data for ignorable missing mechanisms as missing completely at random (MCAR) and missing at random (MAR). This is called likelihood-based ignorable analysis. Mixed-effects proportional model is referred to as the cumulative logit model [2]. Generalized estimating equations (GEEs) is a likelihood-based model which has been explored because it is popular in analyzing non-Gaussian correlated data [1]. The ordinal negative binomial model without any form of imputation performs greatly in simulation studies and real applications than multiple imputation-based generalized estimating equations (MI- GEE) and other models used.

1. Donneau A. F., Mauer M, Lambert P, Molenberghs G, Albert A, "Simulation-based study comparing multiple imputation methods for non-monotone missing ordinal data in longitudinal settings," Journal of Biopharmaucetical Statistical, vol. 25, no. 3, pp. 560-601, 2015
2. Hedeker D, "A mixed-effects multinomial logistic regression model," Statistics in Medicine, vol. 22, pp. 1433-1446, 2003
3. Mallinckrodt CH, Clark SWS, Carroll RJ, and Molenberghs G "Assessing response profiles from incomplete longitudinal clinical trial data undeer regulatory considerations." Journal of Biopharmaceutical Statistics,13, 179-190, 2003

MP72 Using early or baseline data in a trial with missingness in a continuous primary endpoint

Basu J.*, Stallard N.
University of Warwick ~ Coventry ~ United Kingdom

In interrupted trials, we often encounter missingness of observations in a monotone way, such that we lose data for some patients in the follow-up stage. We consider a special case of this problem in which we have primary outcomes missing for some patients for whom early outcome or baseline covariate data are available. Galbraith and Marschner[1] show how to improve the precision of an estimator for the mean of the primary endpoint by constructing the MLE utilising all available observations from each follow-up stage under multivariate normality assumption of the observation vector for each patient. Van Lancker et al.[2] discuss how the precision of treatment effect can be improved by incorporating baseline covariates and short-term endpoints. In this work, we propose an imputation procedure by fitting a regression line assuming the conditional distribution of primary endpoint given baseline as normal. This allows us to achieve flexibility over the constraint of multivariate normal assumption of the observation vector and it gives the same result in spite of relaxed assumptions. We perform Monte-Carlo simulations to compare our approach with that of Galbraith and Marschner[1]. Limitations and further possible extensions of the method are also discussed.

[1] Galbraith, S., Marschner, I.C. Interim analysis of continuous long-term endpoints in clinical trials with longitudinal outcomes, Statistics in Medicine, 22, 2003, 1787-1805.
[2] Van Lancker, K., Vandebosch, A., Vansteelandt, S. Improving interim decisions in randomized trials by exploiting information on short-term endpoints and prognostic baseline covariates. Pharmaceutical Statistics, 19, 2020, 583-601.

MP73 Evaluating bias when estimating causal mediation estimands with non- adherence and missing data

Chis Ster A.*, Landau S., Emsley R.
King's College London ~ London ~ United Kingdom

Many clinical trials report participant non-adherence. An intention-to-treat (ITT) analysis will estimate the causal effect of treatment offer without bias, though ignores the impact of non-adherence. To account for non-adherence, one can estimate a complier-average causal effect (CACE), the average causal effect of treatment receipt in the subgroup of participants who would comply with their randomisation. Evaluating how interventions lead to changes in the outcome (the mechanism) is key for the development of more effective interventions. A mediation analysis aims to decompose a total treatment effect into an indirect effect and a direct effect. However, most current methods for mediation analysis in clinical trials focus on decomposing the ITT effect, and the corresponding effects ignore the impact of participant non-adherence. Previous work has combined these issues and decomposed the CACE into a direct effect, the Complier Average Natural Direct Effect (CANDE), and an indirect effect, the Complier Average Causal Mediated Effect (CACME). These estimands can be estimated using Structural Equation Models (SEMs). However, the reliability and interpretability of the estimates are affected by missing data. The aim of this work is to evaluate the bias of linear SEMs for estimating these estimands when there are missing data under various missingness assumptions. A Monte Carlo simulation study following the ADEMP framework is conducted to evaluate the bias in CACE, CACME, and CANDE when there are missing data in the pairwise combinations of missing mediator and outcome. We construct three scenarios where the missing data are MCAR, six MAR scenarios, and ten MNAR scenarios, to cover a range of realistic trial scenarios. We vary 8 parameters, including the trial size, the proportion of non-adherence, and the proportion of missing data. Our findings show that linear SEMs provide unbiased estimates of CACE, CACME, and CANDE for all MCAR, some MAR scenarios, but not for any MNAR scenarios. Trialists should consider the missing mechanisms in their study, and, if appropriate, use linear SEMs to estimate the CACE, CANDE, and CACME. An accompanying Stata command (compmed) has been developed for practical implementation of this method and will be illustrated. Park Soojin & Kürüm Esra, 2020. "A Two-Stage Joint Modeling Method for Causal Mediation Analysis in the Presence of Treatment Noncompliance," Journal of Causal Inference, De Gruyter, vol. 8(1), pages 131-149, January.

MP74 Simulation study of four methods for sensitivity analysis of the mar assumption with an application to rcts

Gaunt D.M.*[1], Chris M.[1], Hughes R.A.[2]
[1]Population Health Science, Bristol Medical School, University of Bristol ~ Bristol ~ United Kingdom, [2]Intergrative Epidemiology Unit, University of Bristol ~ Bristol ~ United Kingdom

Analyses of randomised controlled trials (RCTs) are often carried out under the Missing At Random (MAR) assumption, although sensitivity analyses using statistical methods that allow missing data to be Missing Not At Random (MNAR) are advised. In this simulation study, we compare four MNAR analysis methods in the novel situation where the simulated MNAR mechanism differs between the control and intervention groups. A key comparison investigated the effect of assuming, incorrectly, that the MNAR mechanism was the same in both groups. This simulation study was based on an RCT testing the effectiveness of a physical activity programme on the moderate to vigorous physical activity of adolescent girls. The data generation scenarios included two proportions of missingness (30%, 50%), two treatment effects (null effect, 10 points), three strengths of bias arising from the association between missingness and outcome (MAR, or MNAR, bias of -3 points, or bias of -5 points), and two trial sample sizes (500, 2000). Four MNAR analysis methods were compared; mean-score method, selection model with inverse-probability weighting (SM-IPW), delta-adjusted multiple imputation (delta-MI) and stacked-MI, and their implementation in software packages. The treatment effect estimated from each method were compared using the bias, empirical standard error, average model-based standard error, and confidence interval coverage. There were limited differences with regards to the bias and standard errors of all methods. Under the correct assumption the stacked-MI method seemed to perform worst, followed by SM-IPW, delta-MI, and mean-score. This pattern was seen in the run times, with the stacked-MI method, using bootstrapping for standard errors, taking 247 times longer than the mean- score method. Under the incorrect assumption, all methods seemed to perform similarly. We will apply these methods to the RCT data and implement a tipping point analysis. It is important to account for data MNAR to avoid misleading conclusions, however even making the incorrect assumption of the same MNAR mechanism in both arms is preferable to assuming MAR. White I. A mean score method for sensitivity analysis to departures from the missing at random assumption in randomised trials. Statistica Sinica. 2018;28(4):1985–2003-1985–2003.
Beesley LJ, Taylor JMG. Accounting for not-at-random missingness through imputation stacking. Statistics in Medicine. 2021;40(27):6118–32-–32.

Poster Sessions

MP75 Expert elicitation methods for the evaluation of missing data assumptions: a scoping review

Greenwood S.*[1], Morris T.P.[2], Aucott L.[1], O"Malley L.[3], Goulao B.[1]
[1]*University of Aberdeen ~ Aberdeen ~ United Kingdom,* [2]*University College London ~ London ~ United Kingdom,*
[3]*University of Manchester ~ Manchester ~ United Kingdom*

Most trials have missing data. To handle its occurrence, trials should have strategies and assumptions built into trial design and analyses, else they risk bias, loss of study power and increased variability (1). The evidence available to help trialists form their missing data assumptions can be very sparse, as the information needed is the information that is missing. How can we gather more evidence to help form these assumptions? A potential method is expert elicitation, where opinions are elicited from experts who are thought to have informed knowledge or beliefs on an unknown quantity of interest. The objective of this scoping review is to map out expert elicitation methods that can be used to assess missing data assumptions in clinical trials. As the review results will inform later stages of research, all the identified methods will be evaluated. We selected a scoping review to address our research objective as the evidence will be from multidisciplinary sources, and it is subsequently not clearly mapped with consistent terms. The inclusion criteria limits the results to sources describing expert elicitation methods in sufficient detail to allow comprehension and reproduction of the method. Identified methods will be evaluated against a critical appraisal criterion based on previous review examples (2). The team expects to find evidence after initial searches revealed: protocols containing guidance on elicitation methods; examples applying methods for the assessment of missing data assumptions; literature reviews identifying expert elicitation resources; increasing trends regarding application of expert elicitation methods for trials; and extensive literature on the topic. Given the initial search findings, this review is expected to find expert elicitation methods that could be adopted into clinical trial analyses for the assessment of missing data assumptions.

[1] European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. 2010; Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data- confirmatory-clinical-trials_en.pdf. Accessed 07 March, 2023.
[2] Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. Journal of clinical epidemiology 2010;63(4):355-369.

MP76 Correcting for bias due to missing data in longitudinal studies on quality of cancer patients

Haug N.*[1], Jänicke M.[1], Kasenda B.[2], Marschner N.[1], Frank M.[1]
[1]*iOMEDICO AG ~ Freiburg im Breisgau ~ Germany,* [2]*University Hospital Basel ~ Basel ~ Switzerland*

Many studies on cancer patients investigate the impact of treatment on health-related quality of life (QoL). Typically, QoL is measured longitudinally, meaning that study subjects receive questionnaires at predefined time points throughout the observation period. At each time point, a patient may not respond to individual questionnaire items or may not return it as a whole. This leads analysts to the problem of how to deal with these missing data when evaluating outcomes. Common approaches include available case (AC) and last-observation-carried-forward (LOCF) analyses. Counterintuitively, results obtained via these methods often do not detect a decrease of mean QoL of advanced-stage (terminal) cancer patients over time. A partial explanation for this phenomenon is that patients with lower baseline QoL have shorter survival times - we call this the survivor effect. Additionally, results may be biased if missingness of questionnaires does not occur completely at random (MCAR) - for instance, alive patients with lower QoL may be less prone to sending back questionnaires. To quantify and correct for this bias we here apply an approach which does not require data to be MCAR. Additionally, we also report results from an exploratory analysis highlighting the extent of the survivor effect.
We perform our study on two large cancer registries including 1,927 patients with advanced pancreatic and 797 patients with advanced breast cancer. Using augmented inverse probability weighting (AIPW) [1] we estimate the population average of the QoL of all registry patients at each time point including those who did not return their questionnaire. This method only requires data to be missing at random (MAR). The AIPW estimates are then compared to those obtained from using AC and LOCF approaches. Using a pattern mixture model (PMM), we also analyze sensitivity of our results against violation of the MAR assumption. We do not observe a significant difference between the considered estimators, indicating that the counterintuitive result of stable mean QoL in cohorts of advanced-stage cancer patients is predominantly driven by the survivor effect.

[1] S. R. Seaman, S. Vansteelandt, Statistical Science, Vol. 33, No. 2, 2018, 184-197.

MP77 Accounting for data missing not at random: comparison of bayesian and monte carlo probabilistic bias analyses

Clayton G.[1], Kawabata E.[1], Major--Smith D.[1], Shapland C.Y.[1], Carter A.[1], Fernández--Sanlés A.[2], Borges M.C.[1], Morris T.[3], Tilling K.[1], Griffiths G.[1], Millard L.[1], Davey Smith G.[1], Lawlor D.[1], Hughes R.*[1]
[1]*University of Bristol ~ Bristol ~ United Kingdom,* [2]*University College London ~ London ~ United Kingdom,* [3]*MRC Clinical Trials Unit at UCL ~ London ~ United Kingdom*

Bias from data missing not at random (MNAR) is a persistent concern in health-related research[1-3]. A bias analysis quantitatively assesses whether conclusions change under different assumptions about missingness[4,5]. The dependency between missingness and the outcome is typically modelled using either a selection or pattern-mixture model[4]. Both include non-identifiable bias parameters which govern the magnitude and direction of the bias. Probabilistic bias analysis (PBA) specifies a prior distribution for the bias parameters, explicitly incorporating available information and uncertainty about their true values[5]. A Bayesian PBA combines the prior distribution with the data's likelihood function whilst a Monte Carlo PBA samples the bias parameters directly from its prior distribution[6]. We compared a Bayesian PBA with a Monte Carlo PBA when a large proportion of outcome data were MNAR. Motivated by a real example, we investigated fitting a logistic regression when the outcome was MNAR for most participants and missingness depended on the outcome, exposure and auxiliary variables. Via simulations, we compare complete case analysis (CCA), MAR implementations of multiple imputation (MI_MAR) and inverse probability weighting (IPW_MAR) and two PBAs: fully Bayesian selection model (BSM) and Monte Carlo pattern-mixture (MCPM) imputation approach. We repeated the simulation study for (1) causal and null exposure effects, (2) informative and weakly informative prior for the bias parameters, and (3) data generated according to a selection and pattern- mixture model. Estimates of the CCA, MI_MAR and IPW_MAR were substantially biased with 95% confidence interval (CI) coverages of 7-64%. Including auxiliary variables in MI_MAR's imputation model amplified the bias due to assuming data were MAR. In the causal setting, applying MCPM with an informative or weakly informative prior resulted in negligible bias and close to nominal CI coverage. In comparison, levels of bias were 3-9 times higher for BSM resulting in slight CI under-coverage (88-92%). Also, BSM failed to fit in over 9% of the simulated datasets. The same patterns were found for the null setting, although CI over-coverage was observed for both PBAs. Relative performance of the PBAs was unaffected by the data generation model. Monte Carlo PBA is a viable alternative to a fully Bayesian PBA.

[1] G. Carreras, G. Miccinesi, A. Wilcock, N. Preston, D. Nieboer, L. Deliens, M. Groenvold, U. Lunder, A. van der Heide, M. Baccini and ACTION consortium, "Missing not at random in end of life care studies: multiple imputation and sensitivity analysis on data from the ACTION study," BMC Medical Research Methodology, vol. 21, no. 13, 2021.
[2] A.-D. Hazewinkel, J. Bowden, K. H. Wade, T. Palmer, N. J. Wiles and K. Tilling, "Sensitivity to missing not at random dropout in clinical trials: Use and interpretation of the trimmed means estimator," Statistics in Medicine, vol. 41, no. 8, pp. 1462-1481, 2022.
[3] I. Petersen, C. A. Welch, I. Nazareth, K. Walters, L. Marston, R. W. Morris, J. R. Carpenter, T. P. Morris and T. M. Pham, "Health indicator recording in UK primary care electronic health records: key implications for handling missing data," Clinical Epidemiology, vol. 11, pp. 157-167, 2019.
[4] J. R. Carpenter and M. Smuk, "Missing data: A statistical framework for practice," Biometrical Journal, vol. 63, pp. 915-947, 2021.
[5] J. N. Hunnicutt, C. M. Ulbricht, S. A. Chrysanthopoulou and K. L. Lapane, "Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review," pharmacoepidemiology and drug safety, vol. 25, pp. 1343-1353, 2016.
[6] L. C. McCandless and P. Gustafson, "A comparison of Bayesian and Monte Carlo sensitivity analysis for unmeasured confounding," Stat Med, vol. 36, no. 18, pp. 2887-2901, 2017.

MP78 Methods of dealing with attrition bias due to non-random dropout in models with binary outcome

Janošová M.*[1], Katina S.[2]
[1]RECETOX, Faculty of Science, Masaryk University, and Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, Czech Republic ~ Brno ~ Czech Republic, [2]Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno, Czech Republic and Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic ~ Brno ~ Czech Republic

Missing values of some variables and missing subjects at or after some time point due to dropout in longitudinal studies are known sources of bias. While several methods were developed to combat missing values, their use often assumes that data are missing at random (MAR) or missing completely at random (MCAR). When the dropout is dependent on the outcome of interest, i.e. data are missing not at random (MNAR), the MAR/MCAR assumptions are invalid and the performance of these methods is in question. In this presentation we specifically look at models with binary outcome, such as occurence of a disease, when this outcome is unknown in some subjects. Our aim is to compare several methods, namely complete cases analysis, imputation of missing data, inverse probability of censoring weighting and some of its variations. To compare these methods we designed a simulation study in which we look at biases of parameter estimates and their statistical significance for selected combinations of outcome probabilities and dropout probabilities. Based on the results we form a recommendation for practical data analysis. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme project R-Exposome Chair (Grant Agreement 857487) and Masaryk University's specific research support for student projects MUNI/A/1132/2022.

[1] Little R, Rubin D, Statistical Analysis with Missing Data, 3rd Edition, Wiley, 2020.
[2] Su L, Seaman SR, Yiu S, Sensitivity analysis for calibrated inverse probability-of-censoring weighted estimators under non-ignorable dropout. Statistical Methods in Medical Research. 2022;31(7):1374-1391.

MP79 Imputation of cross-classified multilevel models

Jolani S.*
Maastricht University ~ Maastricht ~ Netherlands

In the last few years, multiple imputation of missing data in multilevel models has received many attentions in medical as well as social and behavioral research. When units have a purely hierarchical or nested structure, imputation task is rather straightforward, and typically the imputation model includes dummy variables for clusters to account for the nested structure of the data. However, there are situations in which the hierarchical structure is not complete. In health studies, for instance, patients are nested within hospitals they attend and also within general practitioners they have registered. In such cases, a unit (e.g., a patient) is classified along more than one dimension, the so- called cross-classified structure. It is unclear how correctly the missing data should be imputed in cross-classified models. We focus on two-way cross-classified models where units at level 1 are classified by two variables at level 2. We show how correctly missing data should be imputed to preserve the cross-classification. The results of simulation studies as well as an application in health domain showed that the cross- classified structure of the data should be considered in the imputation step to avoid bias in the parameter estimates. When there is a cross-classified structure in the data, it should be reflected in the imputation process to prevent incorrect conclusions.

MP80 Combining multiple imputation and inverse probability weighting to handle missing data in longitudinal studies

Middleton M.*[1], Nguyen C.[2], Carlin J.[2], Moreno--Betancur M.[2], Lee K.[2]
[1]University of Melbourne ~ Melbourne ~ Australia, [2]Murdoch Children's Research Institute ~ Melbourne ~ Australia

Missing data are ubiquitous in longitudinal studies, with loss to follow-up often resulting in a large proportion of participants with missing outcome data, typically with some sporadic missingness across other analysis variables. Omitting incomplete records from analyses may result in biased parameter estimates as well as a loss of precision. Multiple imputation (MI) is often used to handle missing data, but imputation of a large proportion of missing information may result in biased inferences if the imputation model is misspecified. An alternative approach is to use MI in combination with inverse probability weighting (IPW) where MI is used to handle sporadic missing covariate data, while IPW handles missing outcome data, in a potentially large proportion of the sample. The aim of this research was to compare the performance of a combined MI/IPW approach to either MI or IPW alone when handling missing data in longitudinal studies. We conducted a simulation study to assess the performance of MI/IPW, MI-only, IPW-only, and a complete case analysis, in a longitudinal study with sporadic missingness in covariates and loss to follow-up where the primary interest is the causal effect of an exposure on either a continuous, normally distributed or binary outcome, estimated via regression. We considered a range of realistic scenarios based on a case study, varying the amount of missing data, outcome type, missing data mechanism, and sample size, and illustrated these approaches in the case study. MI/IPW showed slight bias in the effect estimate when considering the continuous outcome and high levels of missing data, but not with a binary outcome. IPW-only produced biased effect estimates and standard errors in small samples, while MI-only was approximately unbiased for the effect estimate, with correct standard errors, across all scenarios. We found no substantial benefit of using a combined MI/IPW approach over MI-only, with the latter providing unbiased inference and a relatively simpler implementation. Overall, these results suggest that MI-only may be the preferred analytical approach when there are large amounts of missing outcome data and sporadic missingness in covariates in longitudinal studies.

Hughes, R.A., Heron, J., Sterne, J.A.C., and Tilling, K., Accounting for missing data in statistical analyses: multiple imputation is not always the answer. International Journal of Epidemiology, 2019. 48(4): 1294-1304 https://doi.org/10.1093/ije/dyz032.
Little, R.J., Carpenter, J.R., and Lee, K.J., A comparison of three popular methods for handling missing data: complete-case analysis, inverse probability weighting, and multiple imputation. Sociological Methods & Research, 2022. 0(0) https://doi.org/10.1177/00491241221113873.
Lee, K.J. and Simpson, J.A., Introduction to multiple imputation for dealing with missing data. Respirology, 2014. 19(2): 162-167 https://doi.org/10.1111/resp.12226.
Seaman, S.R., White, I.R., Copas, A.J., and Li, L., Combining multiple imputation and inverse- probability weighting. Biometrics, 2012. 68(1): 129-137 https://doi.org/10.1111/j.1541-0420.2011.01666.x.

Poster Sessions

MP81 Instability of clinical prediction models caused by dichotomisation of a continuous outcome

Archer L.*[1], Snell K.I.[1], Collins G.S.[2], Riley R.D.[1]
[1]University of Birmingham ~ Birmingham ~ United Kingdom, [2]University of Oxford ~ Oxford ~ United Kingdom

Dichotomisation of continuous outcomes is not recommended for prediction modelling; it reduces power to detect predictor effects and may result in misleading conclusions regarding predictor- outcome associations. Individual-level instability in predictions can occur where sample sizes are small [1] and is an issue for modelling of outcomes on both the continuous and binary scales, possibly leading to poor external validity. We compared instability in individual-level predictions from example models for a continuous outcome, forced expiratory volume [2], when dichotomised either before or after model development. Instability in individual-level predicted risks from the two modelling approaches was assessed for various sample sizes and dichotomisation cut points. Models were compared using prediction instability plots (predicted risks across bootstrap models plotted against predictions from the original model), instability indices (mean absolute difference between predicted risks for an individual, as calculated from the original and bootstrap models), and classification indices (proportion of bootstrap models giving different outcome classifications to the original model, given a pre-defined, clinically relevant threshold probability). Consistency in smoothed calibration curves and decision curves across bootstrap models was also assessed. Predicted risks from the linear regression approach were more stable than those from logistic regression in all examples, with a narrower spread across 500 bootstrap models. Mean absolute differences between predicted risks from original and bootstrap models were lower for linear regression, for example with median instability index across individuals of 0.001 (LQ to UQ: 0.000 to 0.007) for a 0.1 outcome prevalence, compared to 0.002 (0.000 to 0.014) for the corresponding logistic model. Classification was also more consistent for the linear model, where those close to the threshold had lower probabilities of differing classification across bootstrap models. While individual- level predictions were more stable, this was not reflected in increased stability of model calibration or net benefit. Modelling continuous outcomes on their continuous scale and dichotomising after the modelling stage appears to give more stable predictions than logistic models developed using the pre-dichotomised outcome. This is likely due to more efficient use of data when modelling on the continuous scale, with a larger effective sample size for analysis.

[1] Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. arXiv preprint arXiv:2211.01061. 2022 Nov 2.
[2] Rosner, B. (1999), Fundamentals of Biostatistics, 5th ed., Pacific Grove, CA: Duxbury

MP82 Real-time detection of hiv outbreaks among people who inject drugs and modelling of subsequent epidemic states

Baralou V.*[1], Thomadakis C.[1], Demiris N.[2], Gountas I.[3], Touloumi G.[1]
[1]Department of Hygiene, Epidemiology & Medical Statistics, Medical School, National & Kapodistrian University of Athens ~ Athens ~ Greece, [2]Department of Statistics, Athens University of Economics and Business ~ Athens ~ Greece, [3]Medical School, University of Cyprus ~ Nicosia ~ Cyprus

HIV outbreaks among people who inject drugs (PWID) are not uncommon, but differ in their size, duration and post-epidemic level. We aimed to compare through simulations the performance of various methods for detecting in real time the growth, non-increasing (plateau and/or decline) and post-epidemic state of such outbreaks.
Simulations were based on real data from a previous HIV outbreak among PWID in Greece. Data on weekly HIV diagnoses were provided by The European Surveillance System (TESSy) of the European Centre for Disease Prevention and Control (ECDC). Eight scenarios were generated assuming different epidemic states (with/without plateau), shapes (constant/exponential), time progress (abrupt/gradual) and pre-outbreak levels (low/high). To identify each state, we developed a two-state hidden Markov model (HMM), a method based on prediction interval (PI) and a novel one that combines the two methods by assigning weights to HMM's transition probabilities at each time point based on the PI limits (HMM-PI). We also applied classic methods for outbreak detection including regression-based methods and control charts [1]. Methods were applied prospectively; performance was assessed by sensitivity, specificity and timeliness (i.e., the difference between the start of each state and the first alarm after its onset). As for detecting the outbreak onset, HMM-PI performed at least similarly with well-established methods in all scenarios. When applied for detecting the growth and non-increasing state, HMM-PI improved the sensitivity of both HMM and PI but resulted in similar timeliness with PI. However, PI gave the best balance between sensitivity, specificity and timeliness across all settings concerning the non-increasing state. When applied for detecting the post-epidemic state, no method reached a satisfying balance between these metrics; PI though performed better in most scenarios giving high specificity and moderate timeliness. Overall, methods' performance deteriorated when assuming gradual rather than abrupt epidemics. Real data application results were similar to those of simulation scenarios for abrupt outbreaks .
No method is a panacea for identifying all states of an outbreak. HMM-PI seems promising for detecting the outbreak onset while PI is preferable for the remaining states. HMM performed poorly in real time but can provide useful insights if applied retrospectively.

[1] S. Unkel, C. P. Farrington, P. H. Garthwaite, C. Robertson, N. Andrews. Statistical methods for the prospective detection of infectious disease outbreaks: a review. Journal of the Royal Statistical Society, Series A (Statistics in Society), 175, 2012, 49-82.

MP83 Studies developing prediction models are not considering sample size requirements: a systematic review

Dhiman P.*[1], Ma J.[1], Qi C.[2], Bullock G.[3], Sergeant J.[4], Riley R.[5], Collins G.[1]
[1]*University of Oxford ~ Oxford ~ United Kingdom,* [2]*Swansea University ~ Swansea ~ United Kingdom,* [3]*Wake Forest School of Medicine ~ Winston-Salem ~ United States of America,* [4]*University of Manchester ~ Manchester ~ United Kingdom,* [5]*University of Birmingham ~ Birmingham ~ United Kingdom*

Having an appropriate sample size is important when developing a clinical prediction model. Using an insufficient sample size when developing a prediction model leads to imprecise parameter estimates and increases the risk of overfitting, which can yield inaccurate and unstable predictions leading to reduced model performance when evaluated in 'new' individuals from the same population, and ultimately limits generalisability of the model[1]. The aim of this study is to review how sample size is considered in studies developing a prediction model for a binary outcome. We searched PubMed for prediction model studies developing a prediction model for a binary outcome, published between 01/07/2020 and 30/07/2020. We reviewed the sample size calculations and investigated the sample sizes used to develop the models. We calculated the minimum sample size needed to estimate the overall risk and minimise overfitting in each study (Riley et al sample size criteria[2]) and summarised the difference between the calculated and used sample size. Sample size justification was provided in only nine out of 119 included studies (8%), of which five studies cited using an events per variable (EPV) rule of thumb. The recommended minimum sample size could be calculated for 94 studies: 73% did not meet the minimum required sample size to estimate the overall risk and minimise overfitting, with a median deficit of 75 events [IQR: 234 lower to 7 higher]; and 26% did not meet the minimum required sample size to estimate the overall risk. Sample size calculation and justification is rarely reported in studies developing a prediction model for a binary outcome using logistic regression and studies do not use enough data to meet minimum sample size requirements for their prediction model scenario. Models developed using insufficient data may lead to model instability and unreliable predictions, that if used to guide clinical decision making have the potential to cause harm. With formal sample size and reporting guidance now available, we strongly encourage researchers to fully and transparently perform and report their sample size calculations, to meet minimum sample size and reporting requirements for their studies.

[1] Riley RD, Collins GS, arXiv, 2022.
[2] Riley RD, Ensor J, Snell KIE, BMJ,368, 2020, m441.

MP84 A comparison of hyperparameter tuning procedures for clinical prediction models: a simulation study

Dunias Z.*[1], Van Calster B.[2], Timmerman D.[3], Boulesteix A.[4], Van Smeden M.[1]
[1]*University Medical Center Utrecht ~ Utrecht ~ Netherlands,* [2]*KU Leuven ~ Leuven ~ Belgium,* [3]*University Hospitals Leuven ~ Leuven ~ Belgium,* [4]*University of Munich ~ Munich ~ Germany*

Many methods for developing risk prediction models involve one or more hyperparameters. Tuning hyperparameters, such as the regularization parameter in a Ridge or Lasso regression, is often aimed at improving the (out-of-sample) predictive performance of the model. In this study, various hyperparameter tuning procedures for clinical prediction models were systematically compared and evaluated. The focus was on out-of-sample predictive performance (discrimination, calibration and overall prediction error) of risk prediction models developed using Ridge, Lasso, Elastic Net or Random Forest. The influence of sample size, number of predictors and events fraction on the performance of the hyperparameter tuning procedures was studied using extensive simulations. The results indicate important differences between tuning procedures in calibration performance (both in terms of average performance and variability), while generally showing similar discriminative performance. The one-standard-error rule for tuning applied to cross-validation (1SE CV) generally resulted in surprisingly severe miscalibration. Standard non-repeated and repeated cross-validation (both 5-fold and 10-fold) performed similarly well and appeared to outperform the other tuning procedures. Bootstrap showed a slight tendency to more severe miscalibration than the standard cross-validation based tuning procedures. These differences between tuning procedures were larger for smaller sample sizes, lower events fractions and fewer predictors. These results imply that the choice of tuning procedure can have a profound influence on the predictive performance of prediction models. We warn of the potentially detrimental effects on model calibration of the popular 1SE CV rule for tuning prediction models in low-dimensional settings.

MP85 Clinical utility curve: a new proposal to analyze the utility of predictive models

Escorihuela--Sahún M.*[1], Esteban L.M.[1], Borque--Fernando Á.[2], Morote J.[3], Savirón--Cornudella R.[4], Lou-- Mercadé A.C.[6], Sanz G.[5]
[1]*Escuela Universitaria Politécnica de La Almunia, Universidad de Zaragoza ~ La Almunia de Doña Godina ~ Spain,* [2]*Department of Urology, Miguel Servet University Hospital, IIS-Aragon ~ Zaragoza ~ Spain,* [3]*Department of Urology and Surgery, Vall d'Hebron Hospital and Autònoma of Barcelona University ~ Barcelona ~ Spain,* [4]*Department of Obstetrics and Gynecology, San Carlos Hospital and San Carlos Health Research Institute (IdISSC) Complutense University Madrid ~ Madrid ~ Spain,* [5]*Department of Statistical Methods and Institute for Biocomputation and Physics of Complex Systems-BIFI ~ Zaragoza ~ Spain,* [6]*Department of Obstetrics and Gynecology, Hospital Universitario Lozano Blesa ~ Zaragoza ~ Spain*

Predictive models are characterized by three main properties, the calibration, discrimination and clinical utility. The Calibration and discrimination are extensive analyzed by means of receiver characteristic curve and calibration curves. The clinical utility is related with the benefit of the use of the model with a cutoff point, but with less extensive proposal for its analysis. The most used parameter to study the clinical utility is the net benefit, that is a weighted function of true positive and false negative cases [1]. The decision curve compares the net benefit with treat all or none of patients and it provides a good catalogue of threshold points where the model has clinical utility, but its interpretability is not easy, because the net benefit is not a clinical parameter.
Here, we proposed the clinical utility curve [2] that plot two functions for different cutoff points, the first of them is the false negative rate, the patients that are going to be misclassified below the threshold point, and the second one the number of patients that has a value below the cutoff point, that is the patients that are not going to be treated because are classified as negative. The best model corresponds to the lower values of the first curve and the greater ones to the second, that is why we proposal the clinical utility as the parameter defined as the difference of both values. We illustrated the clinical utility curve with two real cases. The first one is the prediction of acidosis (pH < 7.10) in arterial cord blood at birth by using Electronic fetal monitoring showing that for 33% cutoff point, there is a missing of 5% acidotic cases and 46% of unnecessary cesarean sections could be prevented. The second one is a magnetic-resonance-imaging-based predictive model for the prediction of clinically significant prostate cancer in prostate biopsies. The clinical utility curve shows that selecting a 15% threshold avoided 40.1% of prostate biopsies and missed 5.4% of the 36.9% csPCa detected.
We derived a new procedure to estimate the clinical utility of predictive models based on saved treatments and missing diagnosis.

[1] Vickers, A. J., van Calster, B., & Steyerberg, E. W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic and prognostic research, 3(1), 1-8.
[2] Esteban LM, Escorihuela-Sahún ME, Sanz G, Muñoz-Rivero MV, Borque-Fernando Á. Statistical characterization of a good biomarker in oncology. Arch Esp Urol. 2022 Mar;75(2):95-102.

Poster Sessions

MP86 Predicting central line-associated bloodstream infections in hospitalized patients: a systematic review

Gao S.*[1], Albu E.[1], Tuand K.[1], Cossey V.[2], Rademakers F.[2], Van Calster B.[1], Wynants L.[3]
[1]KU Leuven ~ Leuven ~ Belgium, [2]UZ Leuven ~ Leuven ~ Belgium, [3]Maastricht University ~ Maastricht ~ Netherlands

Central line-associated bloodstream infections (CLA-BSIs) are the most common source of hospital- acquired infections (HAIs) associated with higher morbidity and costs, and considered a priority target for prevention. Tools to predict individuals' CLA-BSI risk may help improve infection control in hospitals. In this systematic review, we aimed to evaluate the risk of bias and applicability of published CLA-BSI prediction models, and discuss practical problems for implementing them. Searches were conducted on June 10, 2022 using PubMed, Embase, Web of Science Core Collection and Scopus, including studies describing the development or validation of predictive models for CLA- BSI that have at least two predictors. Articles that did not report the original research (i.e., reviews and conference abstracts), without full text, or qualitative studies were excluded. Two authors independently appraised risk models using CHARMS 1 and assessed their risk of bias and applicability using PROBAST 2. Fifteen risk prediction studies of CLA-BSI were included, describing 37 models. When different algorithms were compared, we focused on the model that was selected as the best by the authors. Eventually we appraised 18 models, including 13 regression models and 5 machine learning models. The C-indexes ranged from 0.67 to 0.82 for internal validation and 0.53 to 0.77 for external validation. No internally validated calibration assessments was provided and only 1 externally validated model plotted the calibration curve. All models were at a high risk of bias. Common reasons were using an inappropriate proxy outcome, measuring predictors that were unavailable at the moment the prediction is needed in practice, inadequate number of events per variables (median = 3), negligence of missing data, absence of model validation particularly calibration assessment. 17 out of 18 models had high applicability concerns, 1 model had unclear concern for applicability due to incomplete reporting. We critically evaluated 18 risk prediction models for CLA-BSI and did not identify any model suitable for practically clinical use. Therefore, a well-developed and applicable model using either regression or machine learning techniques is needed. Additionally, there is an urgent need to improve the methodological conduct of risk prediction studies for any clinical outcome of interest.

[1] Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Annals of internal medicine, 170(1), W1–W33. https://doi.org/10.7326/M18-1377
[2] Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., Reitsma, J. B., & Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS medicine, 11(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744

MP87 Extended joint models for longitudinal viral load and virologic failure in hiv at western cape (south africa)

Honwana F.*[1], Mukonda E.[1], Gumedze F.[3], Myer L.[1], Hsiao M.[2]
[1]Division of Epidemiology and Biostatistics, School of Public Health, University of Cape Town ~ Cape Town ~ South Africa, [2]Division of Medical Virology, University of Cape Town/National Health Laboratory Service Virology, Groote Schuur Hospital ~ Cape Town ~ South Africa, [3]Department of Statistical Sciences, Faculty of Science, University of Cape Town ~ Cape Town ~ South Africa

Routine collection of plasma HIV RNA 'viral load' provides a characterization of HIV burden in resource-constrained health care systems such as those from low-middle-income countries. However, viral load can only be reliably quantified to a certain limit of detection. Resulting in viral load data often left truncated below the lower limit of quantification and right-skewed. These characteristics may lead to methodological issues. For individuals above the detection limit, elevated viral loads are often observed, and these elevated viral loads are more likely to be associated with negative outcomes such as the risk of virologic failure. Joint models assume a normal distribution for the longitudinal outcome. This assumption may not be reasonable for viral load data. It is reasonable to jointly model semicontinuous longitudinal biomarker measures and time to virologic failure data to enable improved predictions of virologic failure. We used longitudinal viral load measures taken from 91,818 individuals on the Western Cape routine data between 2008 and 2018. Of these individuals, the majority (71.1%) of them had viral load values that were below the detection limit of 50 copies/mL at the first observed study time. Under a Bayesian framework, we defined a two-part joint model for time-to-virologic failure and longitudinal viral load biomarkers characterized by a right-skewed distribution and exhibiting heterogeneity. We used an adaptive Gauss-Hermite method to estimate the parameters. The results from the application of the two-part joint model on a subset (10%) of the individuals in the routine data were unsatisfactory due to the difficult nature of routine data. Nonetheless, posterior predictive checks revealed evidence of robustness in the parametrization of the model. The current underlying trajectory of a semicontinuous viral load marker was predictive of individuals' virologic failure. It is important to assume a two-part joint model for time-to-virologic failure and longitudinal viral load biomarker data when biomarker viral load data has a large portion of values that are below the detection limit. To achieve a reliable result, high-performance computing is required to enable more iterations and applications on all the individuals in the biomarker viral load data.

[1] Yu T, Wu L, Gilbert PB. A joint model for mixed and truncated longitudinal data and survival data, with application to HIV vaccine studies. Biostatistics. 2018 Jul 1;19(3):374-90.
[2] Chen Q, May RC, Ibrahim JG, Chu H, Cole SR. Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. Statistics in medicine. 2014 Nov 20;33(26):4560-76.
[3] Dagne GA. Joint two-part Tobit models for longitudinal and time-to-event data. Statistics in Medicine. 2017 Nov 20;36(26):4214-29.
[4] Brilleman SL, Crowther MJ, May MT, Gompels M, Abrams KR. Joint longitudinal hurdle and time-to-event models: an application related to viral load and duration of the first treatment regimen in ith HIV initiating therapy. Statistics in Medicine. 2016 Sep 10;35(20):3583-94.

MP88 Clinical prediction models for transition to psychosis in individuals meeting at risk mental state criteria

Hunt A.*, Bonnett L.
University of Liverpool ~ Liverpool ~ United Kingdom

Psychotic disorders affect 1% of the UK and the At Risk Mental State (ARMS) criteria identifies individuals at high risk for psychotic disorders.[1] Although multivariable models exist to predict an individual's risk of transition to psychosis, there is a need for improvement, as currently only about 18% of individuals meeting ARMS criteria transition within 12 months. [2] The project aimed to synthesise evidence about existing and validating prediction models for the transition of psychosis within the ARMS criteria and appraise the risk of bias of identified models. This informed the development of a new improved prognostic tool to predict ARMS individuals at risk of transitioning. The systematic review and individual participant data meta-analysis was conducted to assess the risk of bias in other, similar, prediction models. The following bibliographic databases were searched: PsycINFO, Medline, EMBASE and CINAHL from 1994 to 2022. The results from the review also assisted in identifying the prognostic factors used for our own developed prediction model. The model was developed using logistic regression with backwards selection and an individual participant dataset, and internally validated using bootstrap resampling. Model performance was evaluated via discrimination and calibration. Results were reported inline with the TRIPOD guidelines. The systematic review identified 69 unique prediction models related to a risk of transition to psychosis. However, the quality assessment highlighted only 4 unique prediction models presenting an overall low risk of bias. The meta-analysis received data from 26 studies contributing 3,739 participants, 2909 of whom are available for model building. Participants were from 20 studies, 359 developed a psychotic disorder. Our final model included the following prognostic factors; disorder of thought content, disorganised speech, and functioning. Discrimination was 0.68 and the calibration slope was 0.91. The systematic review provided critical insight into the choice of prognostic factors within our novel prediction model. Our developed model performed well in the dataset comprising data from 20 studies. The external validity should also be tested to ensure clinical utility of the model. The project would contribute to routine practice at mental health services and ultimately, improve the lives of people meeting ARMS criteria.

[1] Onwumere, J., D. Shiers, and C. Chew-Graham, Understanding the needs of carers of people with psychosis in primary care. 2016, British Journal of General Practice. p. 400-401.
[2] Fusar-Poli, P., et al., Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. Archives of general psychiatry, 2012. 69(3): p. 220-229.

Poster Sessions

MP89 Risk prediction models for individual diagnosis: a prostate cancer case study

Jalali A.*[1], Newell J.[2], Foley R.[3], Maweni R.[3], Murphy K.[4], Lundon D.[3], Lynch T.[5], Power R.[6], O'brien F.[7], O'malley K.[8], Galvin D.[8], Durkan G.[9], Murphy B.[3], Watson W.[3]
[1]University of Limerick ~ Limerick ~ Ireland, [2]University of Galway ~ Galway ~ Ireland, [3]University College Dublin ~ Dublin ~ Ireland, [4]Maynooth University ~ Kildare ~ Ireland, [5]St. James University Hospital ~ Dublin ~ Ireland, [6]Beaumont Hospital ~ Dublin ~ Ireland, [7]University Hospital Waterford ~ Waterford ~ Ireland, [8]Mater Misericordiae University Hospital ~ Dublin ~ Ireland, [9]University Hospital Galway ~ Galway ~ Ireland

Risk prediction models are commonly used for risk stratification in various areas including cancer research. Prostate cancer (PCa) represents a significant healthcare problem, where the first critical clinical question is the need for a biopsy. The study aims to determine whether accurate risk stratification of patients would help to reduce the need for a biopsy and their exposure to its side effects. Statistical and machine-learning approaches were applied to the clinical data of 4801 prostate cancer patients from an Irish cohort in order to develop a clinical risk calculator to accurately predict the risk of prostate cancer and high-grade (Gleason ≥7) PCa [1]. The discrimination ability of the model is internally validated using cross-validation to reduce overfitting, and the model performance is compared with PSA and the American risk calculator (PCPT), Prostate Biopsy Collaborative Group (PBCG) and European risk calculator (ERSPC) through various performance outcome summaries. The risk calculator demonstrated significant improvements in the stratification of PCa patients in an Irish setting, and a higher net benefit from decision curve analysis. Risk prediction models were developed that have the potential to improve the predictive ability for prostate cancer diagnosis. A decision-making tool was also created to facilitate the use of risk prediction models to inform better clinical decision-making and reduce overdiagnosis. This could contribute to reducing the number of men requiring a biopsy and their exposure to its side effects. Effective communication of prediction models is as important as developing, evaluating, and validating them in order to integrate them into clinical practice and facilitate clinical decision-making. The 'DynNom' R package [2] will be introduced which facilitates the development of such decision-making tools for a variety of prediction models.

[1] Jalali, A., Foley, R. W., Maweni, R. M., Murphy, K., Lundon, D. J., Lynch, T., ... & Watson, R. W. (2020). A risk calculator to inform the need for a prostate biopsy: a rapid access clinic cohort. BMC medical informatics and decision making, 20(1), 1-11.
[2] Jalali, A., Alvarez-Iglesias, A., Roshan, D., & Newell, J. (2019). Visualising statistical models using dynamic nomograms. PloS one, 14(11), e0225253.

Poster Sessions

MP90 Real time prediction of infectious disease outbreaks based on google trend data in africa

Muchene E.*
University of Nairobi ~ Nairobi ~ Kenya

New infections with infectious diseases occur quite often in a given susceptible community. However, they do not always lead to an outbreak which would warrant & trigger massive government intervention at the right time. With the advancement in information technology, real-time data collection and dissemination has grown significantly. One of the greatest success stories in real-time disease analytics has been in influenza research using Google flu trends which can predict regional outbreaks of influenza 7-10 days before the center for disease control and prevention surveillance systems. Other real-time tools for data collection, based on mobile network coverage for instance, have been adopted by aid agencies such as the Red cross during the Ebola 2015 outbreak in West Africa We propose to obtain data as reported in the World health Organization website and where possible, data from ministries and other research institutions in addition to Google trends data. Where possible, other population demographics will be incorporated in the data. For the modeling strategy, we propose a joint linear/non-linear model as may be deemed appropriated. Joint modeling especially using random effects to capture the association between two or more outcomes has commonly been applied to jointly model longitudinal and survival data. The proposed model will be enriched with additional covariates which potentially adjust for internet coverage, population composition and literacy levels amongst others. As the research evolves, if deemed plausible, Bayesian hierarchical joint models will be evaluated since they allow for inclusion of prior (historical) information about prevalence of particular infectious diseases in the respective regions under analysis. Finally, the model will be validated through a cross-validation procedure and potentially, a simulation study performed to evaluate the robustness metrics of the model. If validated, the proposed models will not only provide a cheap tool for infectious diseases data collection, but also promote timely intervention through real-time prediction of the incidence of infectious disease. This will in turn enhance the preparedness of disaster response teams in handling disease outbreaks and more so, in channeling the resources to timely interventions that may prevent a more costly outbreak. Amarasinghe, Ananda, Joel N Kuritsk, G William Letson, and Harold S Margolis. 2011. "Dengue Virus Infection in Africa." Emerging Infectious Diseases 17 (8): 1349–54. doi:10.3201/eid1708.101515.

Anders, Katherine L, and Simon I Hay. 2012. "Lessons from Malaria Control to Help Meet the Rising Challenge of Dengue." The Lancet. Infectious Diseases 12 (12): 977–84. doi:10.1016/S1473- 3099(12)70246-3.
Camacho, Anton, Adam Kucharski, Yvonne Aki-Sawyerr, Mark A White, Stefan Flasche, Marc Baguelin, Timothy Pollington, et al. 2015. "Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: A Real-Time Modelling Study." PLoS Currents 7 (January). doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2.
Carneiro, Herman Anthony, and Eleftherios Mylonakis. 2009. "Google Trends: A Web-Based Tool for
Real-Time Surveillance of Disease Outbreaks." Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America 49 (10): 1557–64. doi:10.1086/630200.
Dugas, Andrea Freyer, Mehdi Jalalpour, Yulia Gel, Scott Levin, Fred Torcaso, Takeru Igusa, and Richard E Rothman. 2013. "Influenza Forecasting with Google Flu Trends." PloS One 8 (2). Public Library of Science: e56176. doi:10.1371/journal.pone.0056176.
Gluskin, Rebecca Tave, Michael A Johansson, Mauricio Santillana, and John S Brownstein. 2014. "Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends." PLoS Neglected Tropical Diseases 8 (2). Public Library of Science: e2713. doi:10.1371/journal.pntd.0002713.
"Google Trends." 2016. Accessed March 27. https://www.google.com/trends/.
IFRC. 2015. "Using Real-Time Data to Improve Emergency Response - IFRC." The International Federation of Red Cross and Red Crescent Societies. http://www.ifrc.org/en/news-and-media/news- stories/africa/liberia/using-real-time-data-to-improve-emergency-response-68958/.
Meystre, Stephane. 2005. "The Current State of Telemonitoring: A Comment on the Literature." Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association 11
(1). Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA: 63–69. doi:10.1089/tmj.2005.11.63.
Moorthy, Vasee S, Michael F Good, and Adrian V S Hill. 2004. "Malaria Vaccine Developments." Lancet 363 (9403). Elsevier: 150–56. doi:10.1016/S0140-6736(03)15267-1.
Njagi, E. N., D. Rizopoulos, G. Molenberghs, P. Dendale, and K. Willekens. 2013. "A Joint Survival- Longitudinal Modelling Approach for the Dynamic Prediction of Rehospitalization in Telemonitored Chronic Heart Failure Patients." Statistical Modelling 13 (3): 179–98. doi:10.1177/1471082X13478880.
Okiro, Emelda A, Simon I Hay, Priscilla W Gikandi, Shahnaaz K Sharif, Abdisalan M Noor, Norbert Peshu, Kevin Marsh, and Robert W Snow. 2007. "The Decline in Paediatric Malaria Admissions on the Coast of Kenya." Malaria Journal 6 (1): 151. doi:10.1186/1475-2875-6-151.
Plachouras, D, B Sudre, M Testa, E Robesyn, and D Coulombier. 2014. "Early Transmission Dynamics of Ebola Virus Disease (EVD), West Africa, March to August 2014 - Euro Surveillance 17 September 2014." Euro Surveillance : Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin 19 (37). http://europepmc.org/abstract/med/25259536.
Rizopoulos, Dimitris. 2012. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R.
https://books.google.be/books/about/Joint_Models_for_Longitudinal_and_Time_t.html?id=xotIpb2dua MC&pgis=1.
Wesolowski, Amy, Caroline O Buckee, Linus Bengtsson, Erik Wetter, Xin Lu, and Andrew J Tatem. 2014. "Commentary: Containing the Ebola Outbreak - the Potential and Challenge of Mobile Network Data." PLoS Currents 6 (January). doi:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e

MP91 Developing a prediction model for survival – a case study

Kilian S.*[3], Burgmaier K.[1], Liebau M.[2], Kieser M.[3]
[1]Department of Pediatrics, Faculty of Medicine, University Hospital Cologne and University of Cologne ~ Cologne ~ Germany, [2]Department of Pediatrics, Center for Family Health, Center for Rare Diseases, and Center for Molecular Medicine, University Hospital Cologne and Faculty of Medicine, University of Cologne ~ Cologne ~ Germany, [3]Institute of Medical Biometry, Heidelberg University ~ Heidelberg ~ Germany

Developing a prediction model requires a careful choice of methods. In particular, this applies to survival endpoints since the usual practice in regression analysis to predict the mean of a distribution may not be appropriate. The TRIPOD statement provides a framework for specifying and reporting the process [1]. This includes aspects like outcome, predictors, missing data, model specification, and validation. We discuss advantages and disadvantages of different ways to handle each aspect and we present the choices we made for developing and validating a prediction model for kidney survival of patients with the Autosomal Recessive Polycystic Kidney Disease. For example, the type of prediction made for a patient could be a relative risk score, a complete survival distribution, or something in between. Also, the type of model has to be chosen. While the often used Cox model is easily applicable, machine learning methods like random survival forests may give better predictions due to their flexibility. If a small set of predictors is desirable, some kind of variable selection has to be performed. The metric to assess model performance has to be chosen carefully considering the prediction objective. Technical details like the pooling of Kaplan-Meier curves have to be considered when missing values are handled by multiple imputation. Model validation should be prespecified and can be done within the development process by cross validation or on a separate validation. When a dataset is split into development and validation set, representativeness of both sets should be ensured. When developing and validating a prediction model for survival, multiple aspects have to be considered. We illustrate these aspects by a real clinical example.

[1] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery, 148-158.

MP92 Development of blood stasis questionnaire on gynecological diseases

Ko M.M.*
Korea Institute of Oriental Medicine ~ Daejeon ~ Korea, Republic of

Blood stasis is a pathophysiological concept refers to blood flow when it is not smooth or becomes stagnant in traditional East Asia medicine. Korean medical treatments, including those for infertility and dysmenorrhea, have received increasing attention as therapies for women's health. Among women's diseases, menstrual pain, infertility, and uterine fibroids are associated with blood stasis. Accordingly, herbal medicines that disperse blood stasis can be more effective for these diseases than other interventions. However, a precise diagnosis should be made before choosing the treatments because there are kinds induced by other pattern identifications. This study aimed to develop the Blood Stasis Questionnaire for gynecological disease (BSQ-GD) by extracting clinical indicators related to gynecological diseases. In total, 103 patients who met gynecological disease criteria were enrolled in this cross-sectional, observational . This study was approved by the Institutional Review Board of the Korea Institute of Oriental Medicine (IRB No. KIOM I-1310/001-001-03). The reliability and validity of the BSQ-GD were assessed using Cronbach's α, and the prediction accuracy was determined using logistic regression. The BSQ-GD showed satisfactory internal consistency (Cronbach's α coefficient=0.71) and validity, with significant differences in mean scores between blood stasis (22.30±3.34) and non-blood stasis (14.93±3.49) groups. The cut-off value of the BSQ-GD score was 19 points when the Youden Index (73.45), and the concordance probability (0.75) were at their maximum. The area under the receiver operating characteristic curve was approximately 96%, and the sensitivity and specificity of the diagnostic accuracy according to the cut-off values were 80.95% and 92.50%, respectively. This study developed and validated a 7-item BSQ-GD. It satisfied the reliability (α=0.70) and constructed validity requirements; the BSS score of the BSS group with gynecological disorders was higher than that of the non-BSS group. The BSQ-GD had a high discriminative ability for BSS-GD. Further studies are required to overcome limitations related to menstruation and post-menopausal women.

1. Chen KJ. Blood stasis syndrome and its treatment with activating blood circulation to remove blood stasis therapy. Chin J Integr Med 2012;18:891-6.
2.Jung J, Ko MM, Lee MS, Lee SM, Lee JA. Diagnostic indicators for blood stasis syndrome patients with gynaecological diseases. Chin J Integr Med 2018;24:752-7.
3. Kang BK, Park TY, Lee JA, Jung J, Lee MS. Development of a blood stasis syndrome questionnaire and its reliability and validity. Eur J Integr Med 2016;8:942-6.

MP93 Performance evaluation and improvement of the framingham diabetes risk model using community-based koges data

Lee H.A.*[3], Park H.[1], Hong Y.S.[2]
[1]Department of Preventive Medicine, College of Medicine, Ewha Womans University ~ Seoul ~ Korea, Republic of, 2Department of Internal Medicine, College of Medicine, Ewha Womans University ~ Seoul ~ Korea, Republic of, 3Clinical Trial Center, Ewha Womans University Mokdong Hospital ~ Seoul ~ Korea, Republic of

We evaluated the predictive performance of the Framingham Diabetes Risk Model (FDRM)1 using data from an independent cohort in Korea. To improve the FDRM, we developed a modified FDRM by redefining the predictors based on current knowledge, and evaluated the internal and external validity. We used data from a community-based cohort of the Korean Genome Epidemiology Study (age range at baseline 40–69 years), and split the data 7:3 for validation (n = 5409: n = 2318). We calculated the probability of diabetes based on the FDRM. We developed a modified FDRM based on modified definitions of hypertension and diabetes. We also added clinical features related to diabetes to the predictive model. Model performance was evaluated and compared using the area under the curve (AUC). During an 8-year follow-up, we observed 460 and 180 (cumulative incidence 8.5% and 7.8%) incident diabetes cases among diabetes-free subjects in the training and validation data sets, respectively. The modified FDRM consisted of age, body mass index (cut-off points 25.0 and 30.0 kg/m2), hypertension (≥ 130/80 mmHg or taking antihypertensive drugs), elevated triglyceride (≥ 150 mg/dL), hypo high-density lipoprotein cholesterol (< 40 mg/dL for males and < 50 mg/dL for females), elevated fasting blood glucose (≥ 100 mg/dL), and elevated hemoglobin A1c (≥ 5.7%). The expanded clinical model added white blood cells and γ-glutamyl transpeptidase to the modified FDRM. The estimated AUC was 0.71 for the FDRM, which was lower than the 0.85 reported in the Framingham offspring study. Model performance improved when the redefined predictor was applied to the predictive model (AUC 0.82). However, adding clinical features to the modified FDRM produced no further improvement in discrimination (AUC 0.83). External validation was evaluated on cross- sectional survey data, and both models performed well with an AUC of 0.90. The performance of the FDRM in the Korean population was acceptable for predicting incident diabetes, but it was improved when corrected with redefined predictors. The validity of the modified model needs to be evaluated further.
Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1003176). Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Archives of internal medicine. 2007 May 28;167(10):1068-74. PMID:17533210.

Poster Sessions

MP94 Development and validation of a prognostic model for institutionalisation in parksinsonism: ipd meta-analysis

Li Y.*[1], Macleod A.[1], Lawson R.[2], Yarnall A.[2], Bäckström D.[3], Forsgren L.[3], Camacho M.[4], Williams--Gray C.[4], Maple--Grødem J.[5], Alves G.[5], Tysnes O.[6], Counsell C.[1], Mclernon D.[1]
[1]University of Aberdeen ~ Aberdeen ~ United Kingdom, [2]Newcastle University ~ Newcastle ~ United Kingdom, [3]Umeå University ~ Umeå ~ Sweden, [4]University of Cambridge ~ Cambridge ~ United Kingdom, [5]University of Stavanger ~ Stavanger ~ Norway, [6]University of Bergen ~ Bergen ~ Norway

When people with Parkinson's disease (PD) can no longer look after themselves, they may move to a nursing home (institution). Identifying those at high risk of institutionalisation may lead to improved management. We aimed to develop and validate a flexible parametric proportional-odds model and recalibrate if necessary. We analysed 4 prospective PD incidence cohorts from the UK (CamPaIGN, PICNICS, PINE) and Sweden (NYPUM) using a one-stage individual-participant-data (IPD) meta-analysis. Missing data were imputed with multilevel multiple imputation. A flexible parametric proportional-odds model was developed to predict 10-year risk of institutionalisation and performance was tested using internal- external cross-validation. We adjusted the model for age, sex, MDS-UPDRS III (disease impairment), disease stage, and MMSE. Discrimination was assessed using Harrell's Concordance (C). Moderate calibration was assessed by fitting a second flexible parametric proportional-odds model to the external data adjusted for the complementary log-log transform of the predicted 10-year risk (estimated using the model based on the 3 development studies). We plotted the predicted risk from this second model against the predicted risk from the original model.[1] Recalibration was conducted by updating spline knot positions, intercept and slope. From 723 PD patients (individual study sizes ranged from 124–270), there were 215 PD patients institutionalised within 10 years (individual study events ranged from 27–63). From the internal-external cross-validation, Harrell's C ranged from 0.73–0.81, mean calibration from 0.59–1.27 and calibration slope from 0.84–1.29. Moderate calibration was good in PICNICS and PINE (calibration curves were closed to the perfect calibration line), but the model overpredicted in CamPaIGN and underpredicted in NYPUM (calibration curves deviated from the perfect calibration line). After recalibration, mean calibration improved as expected (CaMPAIGN from 0.59 to 0.88, NYPUM from 1.27 to 0.94) and calibration curves were close to the perfect calibration line. We adapted an approach for smooth calibration curves for the Cox model to the flexible parametric proportional-odds model which allows a more accurate visual assessment of calibration. Cross- validation showed good discrimination performance in all studies, good calibration in PICNICS and PINE and inadequate calibration in CamPaIGN and NYPUM, which was improved by recalibration.

1. Austin PC, Harrell Jr FE, van Klaveren D. Graphical calibration curves and the integrated calibration index (ICI) for survival models. Statistics in Medicine. 2020;39(21):2714-42.

MP95 A method for predicting clinical trial enrollment under restrictive constraints

Lunn D.*, Johnson V., Perevozskaya I., Perperoglou A.
GSK ~ London ~ United Kingdom

The Poisson-Gamma model popularised by Anisimov and others (e.g. Anisimov & Fedorov, 2007, Stat Med) can be used to predict clinical trial enrollment, with uncertainty, across a global footprint of clinical sites. The model can be used at the trial planning stage or for ongoing studies, with Bayesian updates taking account of current enrollment data. The method is efficient but primarily geared for "competitive" enrollment scenarios, whereby all sites are assumed to enroll independently. In reality, however, a number of material constraints apply, which render such independence assumptions invalid. For example, most countries will have upper limits set on the number of patients allowed, and some countries may need to recruit a minimum number of patients (to satisfy regulatory requirements, say). We present a method for projecting (and reprojecting) enrollment, with full uncertainty, for such "restricted" enrollment scenarios. The method combines the Poisson-Gamma model with Monte Carlo simulation and ensures that the required constraints are always met. We illustrate the method, its efficient implementation, and ways of presenting/interpreting the output, all of which form part of GSK's new platform for planning and monitoring studies. Anisimov, V, and Fedorov, V (2007) Modelling, prediction and adaptive adjustment of recruitment in multicentre trials, Statistics in Medicine, 26, pp 4958-4975

MP96 An external validation of the kfre in ckd patients of south asian ethnicity

Maher F.*[1], Teece L.[1], Major R.[2], Medcalf J.[2], Brunskill N.J.[2], Sarah B.[1], Gray L.J.[1]
[1]University of Leicester ~ Leicester ~ United Kingdom, [2]University Hospitals Trust ~ Leicester ~ United Kingdom

The Kidney Failure Risk Equation (KFRE) predicts the 2- and 5-year risk of needing kidney replacement therapy (KRT) using four risk factors – age, sex, urine albumin-to-creatinine ratio (ACR), and creatinine-based estimated glomerular filtration rate (eGFR). Although the KFRE has been recalibrated in a UK cohort, this did not consider minority ethnic groups. Further validation of the KFRE in different ethnicities is a research priority. The KFRE also does not consider the competing risk of death, which may lead to overestimation of KRT risk. This study externally validates the KFRE for patients of South Asian ethnicity and compares methods for accounting for ethnicity and the competing event of death. Data were gathered from an established UK cohort containing 35,539 individuals diagnosed with chronic kidney disease. The KFRE was externally validated and updated in several ways taking into account ethnicity, using recognised methods for time-to-event data, including the competing risk of death. A clinical impact assessment compared the updated models through consideration of referrals made to secondary care. The external validation showed the risk of KRT differed by ethnicity. Model validation performance improved when incorporating ethnicity and its interactions with ACR and eGFR as additional risk factors. Further, accounting for the competing risk of death improved prediction. Using a criteria of 5 year ≥5% predicted KRT risk, the competing risks model resulted in an extra 3 unnecessary referrals (0.59% increase) but identified an extra 1 KRT case (1.92% decrease) compared to the previous best model. A hybrid criteria of predicted risk using the competing risks model and ACR ≥70mg/mmol should be used in referrals to secondary care. The accuracy of KFRE prediction improves when updated to consider South Asian ethnicity and to account for the competing risk of death. This may reduce unnecessary referrals whilst identifying risks of KRT and could further individualise the KFRE and improve its clinical utility. Further research should consider other ethnicities. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al. A predictive model for progression of chronic kidney disease to kidney failure. Jama. 2011;305(15):1553-9.
Major RW, Shepherd D, Medcalf JF, Xu G, Gray LJ, Brunskill NJ. The Kidney Failure Risk Equation for prediction of end stage renal disease in UK primary care: An external validation and clinical impact projection cohort study. Public Library of Science (PLoS); 2019.

MP97 Population modeling for circannual rhythms of hba1c in type 2 diabetic patients using large registry data

Nishikawa M.*[1], Yuki M.[2], Sakamoto M.[3]
[1]The Jikei University School of Medicine ~ Tokyo ~ Japan, [2]International University of Health and Welfare Graduate School ~ Tokyo ~ Japan, [3]International University of Health and Welfare Mita Hospital ~ Tokyo ~ Japan

It is said that greater visit to visit HbA1c variability is associated with increased cardiovascular events of type 2 diabetic patients (Hirakawa et al., 2014), and standard deviation (SD) and/or coefficient of variation (CV) in a patient was used as the measure of variability. However, these reports did not reach to a consistent result, since conventional way of estimating variability is only to calculate individual CV/SD in a patient whose number of measurements is more than a certain number, neglecting the timing of measurements and/or other relevant clinical variables. Recently, it has been revealed by Sakamoto et al. (2019) that HbA1c, blood pressure (BP), and LDL-C may have certain circannual rhythms. Evaluating variability or absolute value in HbA1c, taking circannual rhythms into consideration, will improve accuracy and precision for risk prediction. However, it seems difficult to take HbA1c measurement every month in general practice due to medical cost. In our research, we build population PD model in type 2 diabetic patients in steady state on drug treatment using a large registry dataset of JDDM Study Group, to express circannual rhythm of HbA1c. The statistical challenges are how to express the circannual rhythms and to borrow alternative information from correlated variables. For patients data on treatment more than 6 months, we assumed non-linear mixed effect models, where certain combination of a periodic non-linear function and linear functions were applied for population mean PD (HbA1c) model. Month (season), age, sex, and duration of diabetics were modelled as fixed effects, and subject as a random effect. BP and BMI may be included as time-dependent covariates. Another approach was to treat BP as dependent variable and building similar non-linear mixed effect models, then, HbA1c profiles were estimated by some function of BP. Among about 30000 patients, 70% were used for model building with cross-validation and 30% for model validation after reducing the number of measurements. Linear trends over time with different slopes depending on covariates were observed. Circannual rhythms expressed by cosine function would be maintained until covid19. Our model could predict HbA1c profile through several years. The concept of HbA1c variation might become more important in near future. Hirakawa et al. Diabetes Care. 2014;37(8):2359-56. Sakamoto et al. Diabetes Care. 2019;42(5):816-23.

MP98 Current methodological challenges in prognostic modeling - use case multiple sclerosis

Reeve K.[1], On B.I.*[2], Havla J.[3], Burns J.[2], Gosteli M.[4], Mansmann U.[2], Held U.[1]
[1]Epidemiology, Biostatistics and Prevention Institute, University of Zürich ~ Zurich ~ Switzerland, [2]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München ~ Munich ~ Germany, [3]Institute of Clinical Neuroimmunology, LMU Hospital, Ludwig-Maximilians-Universität München ~ Munich ~ Germany, [4]University Library, University of Zurich ~ Zurich ~ Switzerland

Systematic reviews of prognostic models have revealed that reporting quality is often poor, that the studies suffer from high risk of bias, and that they lack external validation studies. In a Cochrane systematic review, we aimed to identify and summarize multivariable prognostic models for quantifying the risk of clinical disease progression, worsening, and activity in adults with multiple sclerosis (MS).
Relevant databases were searched up to July 2021. Validation studies evaluating model performance were also included. More than 13,000 records were identified, and of these 57 studies reporting on 75 model developments were included. Of these, 35 models were developed using traditional statistical methods, whereas the rest using machine learning (ML) methods. Only two of the included models were evaluated externally multiple times. None of the validations were performed by researchers independent from those that developed the model. Over half (52%) of the models were not accompanied by model coefficients, tools, or instructions, hindering their implementation or independent validation. Most of the models (61%) contained predictors requiring specialist equipment likely to be absent from standard clinical or hospital settings. All but one of the model developments or validations was rated as having high overall risk of bias, as assessed by the PROBAST tool [1]. The primary reason for this was the use of inappropriate statistical methods during prognostic model development or evaluation. Over time, we observed an increase in the use of ML methods, starting 2009. Reporting was assessed according to the TRIPOD statement [2], and major deficiencies were identified, especially in the studies using ML. The current evidence is not sufficient for recommending the use of any of the published prognostic models for people with MS. The MS prognostic research community should adhere to the current reporting and methodological guidelines and conduct many more state-of-the-art external validation studies for the existing or newly developed models. Also identified were gaps in methodological guidance regarding the assessment of models developed using complex ML methods.

[1] R.F. Wolff, K.G.M. Moons, R.D. Riley, et al, PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies, Annals of Internal Medicine,170(1), 2019, 51
[2] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement, Journal of Clinical Epidemiology, 68(2), 2015, 112-121

Poster Sessions

MP99 Longitudinal and survival joint prediction: time reparameterization in amyotrophic lateral sclerosis context

Ortholand J.*, Durrleman S., Tezenas Du Montcel S.
Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013 ~ Paris ~ France

Joint modelling has been widely used to improve the estimation accuracy of both longitudinal and survival outcomes. Several have already been developed and they differ by the latent variables used to link the two processes. Shared random effects models are often using Generalized Linear Models (GLM). A non-linear relation for random effects, named time reparameterization, has been shown to offer good performance and interpretability in longitudinal modelling[1]. This work aimed to extend time reparameterization to univariate joint modelling in the context of Amyotrophic Lateral Sclerosis (ALS), for which death or tracheotomy is often associated with the study of the clinical score ALSFRSr. The proposed model (PropM) was benchmarked against three models: the longitudinal model alone (LongMA), a Cox model (CoxM) and a GLM, using the JMbayes2 package[2]. We used 5-fold cross- validation on simulated and real ALS data. Prediction performances were measured using signed error, absolute error, C-index (at 1, 1.5, 2 years), and Integrated Brier score (IBS). The interpretability of the intercept estimated by the model was assessed by the intra-class correlation between this intercept and, on simulated data, the one used for simulation, or, on real data, the age at first symptoms. On simulation data, the PM reduces the bias on signed error compared to LongMA. No performance was significantly different between PropM and CoxM. Compared to GLM, PropM got significantly better results for absolute error and C-index at each time. The agreement, between the simulated and the PropM estimated random effects, was at least 0.703 (CI95%=[0.67,0.74]) for the intercept and at least 0.228 (CI95%=[0.16,0.29]) for the log speed. On real data, PropM reduces the bias for longitudinal metrics compared to both GLM and LongMA. For survival modelling, PropM was significantly better for IBS compared to GLM. CoxM was significantly better, compared to PropM, for C-index whatever the time. Finally, PropM estimated intercept correlates well with age at first symptoms: 0.892 (CI95%=[0.88,0.9]). The proposed joint model using time reparameterization, both on simulated and real data, reduces the bias compared to the existing longitudinal model and gets better results than the state-of-the-art model with good random-effect interpretability.

[1] JB. Schiratti, S. Allassonniere, O. Colliot, S. Durrleman, Journal of Machine Learning Research, 18, 2017, pp.1-33
[2] Rizopoulos D. Journal of Statistical Software, 2022

MP100 Introducing the 'tetris' plot for visualising the instability in variable selection

Thompson D.*, Perperoglou A.
GlaxoSmithKline ~ London ~ United Kingdom

The challenges in data-driven variable selection are widely recognised within the prediction modelling literature. Methods such as stepwise selection for example, adopt numerous conditional steps which (when ignored) complicate the statistical inference that one can draw and additionally invite a 'one-true model' interpretation of the final variable selection. In cases without independent data to replicate findings, or indeed without a suitable body of published scientific literature, we run the considerable risk of playing forward an incorrect selection into a future hypothesis or development investment. The primary issue is that, had the data presented slightly differently the analysis may well have landed on a possibly different selection. Bootstrap resampling is invaluable for exploring this sensitivity and offers an empirical way to interrogate the non-tractable interplay between algorithm and sample data. Through the repeat application of the selection algorithm across 100s of resampled bootstrap draws it is possible to quantify (amongst other things) joint-selection probabilities and marginal probabilities of the variables being chosen. In this presentation we discuss bootstrap sampling as a method of validation for variable selection and introduce the 'tetris' plot as an appealing and simple way to communicate model instability with non-quantitative colleagues. We outline the core building blocks for creating the plot and illustrate its interpretation across a selection of worked examples. We propose the tetris plot, with its broad applicability, as a useful and novel aid in helping highlight the risks in the reification of data-driven variable selection.

MP101 Transforming a published nomogram back to formulas: how to do this manually or with ai

Wang J.*[1], Lu Z.[2]
[1]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University ~ Utrecht ~ Netherlands, [2]Faculty of Science, Utrecht University ~ Utrecht ~ Netherlands

Nomograms are widely used to present clinical prediction models (CPMs), and claimed as a useful tool for users to make individual level prediction. There were still more than 2000 nomograms published in 2020. However, nomograms are from the pencil-and-paper age, they are neither convenient nor efficient, because the users can only predict the probability for one patient at one time, which makes the external validation (usually with >1000 patients) of the CPMs not feasible. In this tutorial, we will show two approaches to transform a published nomogram (as a figure) to mathematical formulas, without getting access to the original data or re-fitting a model. The first approach is based on WebPlotDigitizer or similar tools which can extract the underlying numerical data from figures as images. This approach needs some manual data extraction work. The second approach is based on an AI algorithm to detect information in a nomogram, to automate the data extraction in the first approach. The data extracted from a nomogram will be further process in R. The most frequently used statistical models visualized with nomograms include Logistic regression, Cox model, and linear regression. We will show how to transform nomograms back the these models as formulas step by step. The nomogram extractor tool can revive an existing CPM, get the risk equation behind the nomogram and (re)implement the CPM in a more user friend way, i.e. an online calculator. We also aim to promote stopping researchers from using nomograms in future development of CPMs.

MP102 Dynamic prediction based on conditional restricted mean survival time for right-censored data

Yang Z.*[1], Zhang C.[2], Hou Y.[3], Chen Z.[2]
[1]Stomatological Hospital, School of Stomatology, Southern Medical University ~ Guangzhou ~ China, [2]Department of Biostatistics, School of Public Health (Guangdong Provincial Key Laboratory of Tropical Disease Research), Southern Medical University ~ Guangzhou ~ China, [3]Department of Statistics, School of Economics, Jinan University ~ Guangzhou ~ China

In clinical follow-up studies with a time-to-event end point, the difference in the restricted mean survival time (RMST) between groups is a suitable substitute for the hazard ratio (HR). However, the RMST only measures the survival of patients over a period of time from the baseline and cannot reflect changes in life expectancy over time. Based on the RMST, we study the conditional restricted mean survival time (cRMST) by estimating life expectancy in the future according to the time that patients have survived, reflecting the dynamic survival status of patients during follow-up. We introduce the estimation method of the cRMST based on pseudo-observations, the statistical inference concerning the difference between two cRMSTs (cRMSTd), and the establishment of the robust dynamic prediction model using the landmark method (the dynamic RMST model). Extensive simulation studies are conducted to evaluate the statistical properties of these methods, which are also applied to a real example of patients with chronic kidney disease who received renal transplantations. Simulation results indicate that, when combining different sample sizes, censoring rates, and values of prediction time s and prediction window w, the estimation of the cRMSTd is accurate, and the hypothesis test based on the cRMSTd can effectively control type I error rates. In addition, the regression coefficients of the dynamic RMST model can also be well estimated with very small bias and good coverage. From the results of the C-index and prediction error, it can be seen that the prediction performance of the dynamic RMST model is better than that of the "static" RMST model, which only uses information from the start of follow-up. The cRMST proposed in this paper has a wide range of applicability, which can analyze patients' life expectancy from any prediction time. Considering the time-dependent covariates and time-varying effects of covariates, the dynamic RMST model based on the cRMST can effectively offer more scientific prognostic information for patients who have survived initial s years.

[1] Z. Yang, H. Wu, Y. Hou, H. Yuan, Z. Chen, Dynamic prediction and analysis based on restricted mean survival time in survival analysis with nonproportional hazards, Computer Methods and Programs in Biomedicine, 207, 2021, 106155.
[2] Z. Yang, Y. Hou, J. Lyu, D. Liu, Z. Chen, Dynamic prediction and prognostic analysis of patients with cervical cancer: a landmarking analysis approach, Annals of Epidemiology, 44, 2020, 45-51.

Poster Sessions

MP103 Segmented regression in the context of kidney function after transplantation

Cleenders E.*, Coemans M., Callemeyn J., Kuypers D., Van Loon E., Wellekens K., Verbeke G., Naesens M.
KU Leuven ~ Leuven ~ Belgium

After kidney transplantation, kidney function is characterized by an initial rapid improvement, followed by a change point that introduces a stabilization period. The "stabilized" kidney function level is often used in clinical decision-making and clinical trials, but remains poorly defined. We aimed to characterize the evolution of kidney function in the first year after transplantation by objectively defining the change point which indicates stabilization. We analyzed data from a retrospective cohort study of 921 kidney transplant recipients, who were alive with a functioning graft at 1 year post-transplant. Only observations of kidney function (measured by estimated glomerular filtration rate, eGFR) within the first year were included (N = 49 233 in total; median = 49 observations per transplant; IQR = 41–61 observations). For each patient individually, a segmented regression model with one change point was used to estimate change point timing, eGFR value at change point, rate of change before and rate of change after change point.[1] Associations of those estimated quantities with recipient/donor characteristics and graft failure rate were assessed with linear regression and Cox regression respectively. The change point occurred at a median time of 6.5 days since transplantation (IQR = 3.5–15.8 days). The estimates for change point timing and initial slope were log-transformed to meet the assumptions of linear regression. The initial increase in kidney function was steeper (P<0.001) and the change point occurred earlier (P<0.001) in case of a living donor. Among deceased donor transplantations, the change point occurred later in case of donation after cardiac death, compared to donation after brain death (P=0.002). All aspects of the early evolution were associated with graft failure rate beyond one year post-transplant, although these associations were insignificant after correcting for eGFR value at one year post-transplant. Segmented regression successfully modeled kidney function evolution in the first year after transplantation. For each patient, we objectively defined the change point between the initial increase of eGFR and the subsequent stabilization phase. Long-term kidney failure was shown to be affected by the eGFR level at one year post-transplant, rather than by the shape of the eGFR evolution during the first year.

[1] V. Muggeo, Estimating regression models with unknown break-points. Statistics in Medicine. 2003;22(19):3055-3071.

MP104 Benchmarking an emulated trial against a real target trial: challenges and illustration in cystic fibrosis

Granger E.*[1], Davies G.[2], Frost F.[3], Keogh R.[1]
[1]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [2]UCL Great Ormond Street Institute of Child Health ~ London ~ United Kingdom, [3]University of Liverpool ~ Liverpool ~ United Kingdom

Randomised controlled trials (RCTs) are the gold standard for evaluating the effect of treatments. However, many questions relating to treatments are challenging to address in RCTs. An alternative approach is to use observational data. Target trial emulation using observational data aims to help avoid common biases that can occur, by applying the study design principles of RCTs, combined with an analysis accounting for confounding. Recent years have seen an uptake of trial emulation in several disease areas; however there remains works to be done to establish whether nonrandomised data can reliably assess treatment effects. The aim of this study is to investigate whether we can replicate the results of a published RCT in cystic fibrosis (CF) using patient registry data. We designed a protocol to emulate an RCT that found evidence for a beneficial effect of azithromycin on lung function in people with CF [1]. The protocol describes key components of the RCT, how we plan to emulate these using UK CF Registry data and how we plan to benchmark the results of the emulated trial against the RCT. We also identify key challenges which may lead to discrepancies in the results. For example, it is difficult to emulate the RCTs inclusion and exclusion criteria exactly using the available data and similarly, it is difficult to emulate the RCT treatment regime precisely with no data on treatment doses in the registry. We use the e-values approach to assess the potential impact of unmeasured confounding. This work is the focus of a new international collaboration network of CF researchers (CF Trial Emulation Network). We plan to emulate several RCTs in CF using UK and US CF Registry data. Emulating existing RCTs with registry data will contribute to the evidence base for this approach. This will inform extensions in which we will use target trial emulation to investigate questions that have not been (and may not be) addressed using RCTs, which is particularly relevant in CF due to the high treatment burden and introduction of new precision medicines.

A. Clement, A. Tamalet, E. Leroux, et al.,Thorax, 61, 2006, 895-902.

MP105 Extending interventional disparity (in)direct effects to investigate inequalities in adverse stroke outcome

Lindmark A.*[1], Eriksson M.[1], Darehed D.[2]
[1]Umeå School of Business, Economics and Statistics, Umeå University ~ Umeå ~ Sweden, [2]Department of Public Health and Clinical Medicine, Sunderby Research Unit, Umeå University ~ Umeå ~ Sweden

Low socioeconomic status (SES) is associated with increased risk of death and disability after stroke. Minimizing disparities is warranted, but interventional targets are unclear. We aim to evaluate to what extent SES-based disparities in death and dependency at three months after stroke could be eliminated by offsetting differences in comorbidity, stroke severity, and acute care. Due to the observational nature of our data, we aim to avoid making strong assumptions on the directions of association between the mediators.
We used a causal mediation analysis approach, focusing on interventional disparity effects that target the reduction in observed SES-disparities accomplished by intervening to equalize the distributions of intermediate variables. In addition to being agnostic regarding the causal ordering of the mediators, these have the advantage of shifting the focus from infeasible interventions on SES itself to intervention targets that are more informative from a policy standpoint. We extended the two-mediator group setting in Micali et al. [1] to a setting with four mediator groups, with effect estimation based on Monte Carlo simulation. Results were based on nationwide Swedish register data on patients with acute ischemic stroke in 2015-2016 (n=25 846). SES was defined by both education and income, and categorized into low, mid, and high. Overall, 26.3% of all patients were dead or ADL-dependent three months after stroke. After adjustments for confounding, low SES was associated with an increased absolute risk of 5.4% (95% CI: 3.9%–6.9%) compared to mid SES, and 10.1% (95% CI: 8.1%–12.2%) compared to high SES. Intervening to shift the distribution of all mediators among patients with low SES to those of the more privileged groups would result in an absolute reduction of these associations by 2.2% (95% CI: 1.2%– 3.2%) and 4.0% (95% CI: 2.6%–5.5%), respectively. Intervening on each mediator individually the largest reductions, 1.5% (95% CI: 0.6%–2.3%) and 2.6% (95% CI: 1.5%–3.8%), would be accomplished by equalizing stroke severity. Targeted interventions to equalize SES-related differences in comorbidity, acute care, and particularly stroke severity, could mitigate a considerable part of the SES-gap in adverse outcome after stroke.

[1] Micali N, Daniel RM, Ploubidis GB, De Stavola BL. Maternal Prepregnancy Weight Status and Adolescent Eating Disorder Behaviors: A Longitudinal Study of Risk Pathways. Epidemiology. Jul 2018;29(4):579-589. doi:10.1097/ede.0000000000000850

MP106 Age and time-trends of the association between body mass index and mortality: evidence from the odds study

Mboya I.*[1], Fritz J.[1], Da Silva M.[1], Häggström C.[2], Stocks T.[1]
[1]Department of Translational Medicine, Lund University ~ Malmo ~ Sweden, [2]Department of Public Health and Clinical Medicine, Umeå University ~ Umeå ~ Sweden

The association between body mass index (BMI) and all-cause mortality is U-shaped; however, the BMI associated with the lowest mortality risk (nadir) has been suggested to increase over time, whilst the elevated risk for obesity has decreased. In a large pooled cohort, we investigated these findings further by studying time-trends of the association between BMI and mortality in groups according to sex, baseline age, and cause-specific mortality. We analyzed nationwide data of 3,715,443 men and women from the Obesity and Disease Development Sweden (ODDS) study, with baseline examinations between 1963-2016, and a median baseline age (interquartile range) of 21 (18-30) years. Over a median follow-up of 27 years (interquartile range 17-37 years), 315,259 deaths were observed. We used Cox models to examine the association between BMI (using restricted cubic splines and WHO categories) and mortality. Overall, we observed J- or U-shaped associations and nadirs within the normal weight BMI category. Whereas the nadir of all-cause mortality increased slightly with higher baseline age and calendar year, the increased risk for obesity compared to normal weight decreased with higher age and calendar year. Further time-trend analyses focused on individuals <40 years of age (92% of the population), stratified by death cause and sex. In these younger individuals, the association between BMI and cardiovascular disease (CVD) mortality in obese individuals increased over time in men (hazard ratio (HR) 2.81, 95%CI, 2.64-2.99 before 1980, and 4.05, 3.28-5.00 in 1990 onwards) and women (HR 2.48, 1.83-3.36 before 1980, and 3.20, 2.69-3.80 in 1990 onwards), but decreased over time for other death causes (cancer and other death causes combined). We are currently investigating whether, e.g. residual confounding and time-trends of different CVD death causes over time, which are differentially associated with BMI, explain these findings. Overall, BMI-mortality associations were either J- or U-shaped. Despite the increase across all age groups and calendar years, the nadir remained within the normal weight BMI category. The HRs for CVD mortality in obese individuals increased over the calendar year at baseline in the younger population but decreased for other death causes. Potential causes for these findings are under investigation.

[1] Afzal S, Tybjærg-Hansen A, Jensen GB, Nordestgaard BG. Change in body mass index associated with lowest mortality in Denmark, 1976-2013. Jama. 2016 May 10;315(18):1989-96.
[2] Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3· 6 million adults in the UK. The Lancet Diabetes & endocrinology. 2018 Dec 1;6(12):944-53.

Poster Sessions

MP107 Interobserver variability of recall decisions between mammogram readers in breast cancer screening

Quinn L.*[1], Jenkinson D.[2], Taylor--Philips S.[1], Takwoingi Y.[1], Sitch A.[1]
[1]University of Birmingham ~ Birmingham ~ United Kingdom, [2]University of Warwick ~ Coventry ~ United Kingdom

Breast cancer is one of the most common types of cancer, usually diagnosed after routine breast cancer screening or referral to a specialist breast cancer clinic after a GP consultation. In England, women between 50 and 70 years old are invited to receive a mammogram every three years as part of the breast screening programme. Two independent readers interpret mammograms to decide whether or not to recall a woman on suspicion of breast cancer, the objective of this study was to measure the interobserver variability between mammogram readers in the National Health Service (NHS) breast screening programme.
Across 22 English NHS centres, 401,682 women underwent breast cancer screening mammograms. Descriptive summaries of the women, readers, recall rates, and cancers detected were reported. Interobserver variability was calculated using percentage agreement and Prevalence-Adjusted Bias- Adjusted Kappa (PABAK), reported with 95% confidence intervals. Secondary analysis was completed by age group, cancer diagnosis and reader recall rates. All analysis was completed separately for the first (prevalent) and subsequent (incident) screening. Two mammography readers independently interpreted screening mammograms for 401,682 women. Final recall rates were 7.52% and 3.01% for prevalent and incident screening, respectively. Interobserver variability for recall rates for prevalent screening was 93.6% (95% CI: 93.4% to 93.7%), PABAK was 87.2 (95% CI: 86.9 to 87.4) and for incident screening was 97.2% (95% CI: 97.2 to 97.3), PABAK was 94.4 (95% CI: 94.3 to 94.5). Interobserver agreement between readers was lower when either of the readers had high recall rates. Agreement between readers was higher for women without a cancer diagnosis compared to those with a diagnosis. For women with a diagnosis, the agreement is similar for prevalent and incident screening. Interobserver agreement was similar across age groups when split by prevalent and incident screening. Interobserver agreement between readers for all outcomes was high, and usually higher for incident compared to prevalent screening. Readers with high recall rates lead to more disagreements between readers and there are no differences in agreement across age groups when prevalent and incident screening are separated.

MP108 Assessing the external validity of the validate-swedeheart trial

Rylance R.*[1], Wagner P.[2], Omerovic E.[3], Held C.[4], James S.[4], Koul S.[1], Erlinge D.[1]
[1]Lund University ~ Lund ~ Sweden, [2]Uppsala University ~ Västerås ~ Sweden, [3]Sahlgrenska University hospital ~ Gothenburg ~ Sweden, [4]Uppsala University ~ Uppsala ~ Sweden

The VALIDATE-SWEDEHEART trial was a registry-based randomized trial comparing bivalirudin and heparin in patients with acute myocardial infarction undergoing percutaneous coronary intervention. It showed no differences in mortality at 30 or 180 days. This study examines how well the trial population results may generalize to the population of all screened patients with fulfilled inclusion criteria in regard to mortality at 30 and 180 days. The standardized difference in the mean propensity score for trial inclusion between trial population and the screened not-enrolled with fulfilled inclusion criteria was calculated as a metric of similarity. Propensity scores were then used in an inverse-probability weighted Cox regression analysis using the trial population only to estimate the difference in mortality as it would have been had the trial included all screened patients with fulfilled inclusion criteria. Patients who were very likely to be included were weighted down and those who had a very low probability of being in the trial were weighted up. The propensity score difference was 0.61. There were no significant differences in mortality between bivalirudin and heparin in the inverse-probability weighted analysis (hazard ratio 1.11, 95% confidence interval (0.73, 1.68)) at 30 days or 180 days (hazard ratio 0.98, 95% confidence interval (0.70, 1.36)). The propensity score difference demonstrated that the screened not-enrolled with fulfilled inclusion criteria and trial population were not similar. The inverse-probability weighted analysis showed no significant differences in mortality. From this, we conclude that the VALIDATE results may be generalized to the screened not-enrolled with fulfilled inclusion criteria.

1. Marchand E, Stice E, Rohde P, et al. Moving from efficacy to effectiveness trials in prevention research. Behav Res Ther 2011; 49(1): 32–41. [PMC free article] [PubMed] [Google Scholar]
2. Akobeng AK. Assessing the validity of clinical trials. J Pediatr Gastroenterol Nutr 2008; 47(3): 277–
282. [PubMed] [Google Scholar]
3. Dekkers OM, von Elm E, Algra A, et al. How to assess the external validity of therapeutic trials: a conceptual approach. Int J Epidemiol 2010; 39: 89–94. [PubMed] [Google Scholar]
Erlinge D, Omerovic E, Frobert O, et al. Bivalirudin versus heparin monotherapy in myocardial infarction4. . N Eng J Med 2017; 377: 1132–1142. [PubMed] [Google Scholar]
5. Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. Lancet 2005; 365: 82–93. [PubMed] [Google Scholar]
6. Paul M, Bronstein E, Yahav D, et al. External validity of a randomised controlled trial on the treatment of severe infections caused by MRSA. BMJ Open 2015; 5: e008838. [PMC free article] [PubMed] [Google Scholar]
7. Erlinge D, Koul S, Omerovic E, et al. Bivalirudin versus heparin monotherapy in non-ST-segment elevation myocardial infarction. Eur Heart J Acute Cardiovasc Care 2019; 8(6): 492–501. [PubMed] [Google Scholar]

Poster Sessions

MP109 Exploring the sample quality: the comparison of data from the mapme2 intervention with ncmp, england, 2021/22

Shojaei Shahrokhabadi M.*, Adamson A.J., Teare M.D., Jones A.R., Basterfield L., Matthews J.N., Hiu S.
Population Health Sciences Institute, Newcastle University ~ Newcastle ~ United Kingdom

The MapMe2 study is a cluster randomised trial designed to assess the effectiveness and feasibility of implementing an intervention within the National Child Measurement Programme (NCMP). The NCMP routinely measures the height and weight of children in Reception (ages 4/5 years) and Year 6 (ages 10/11 years) in England and provides feedback to parents about their child's weight status. The primary objective of the MapMe2 study is to evaluate whether a web-based intervention delivered within the NCMP feedback can enhance parents' ability to identify their child's weight status accurately and, ultimately, improve their child's weight status at 12mth. The trial has recruited approximately 56,000 children from 10 participating Local Authorities who are part of the NCMP; primary schools were randomised to one of three intervention arms: MapMe-web access detail provided in NCMP feedback letter only; the former with an additional reminder letter at 6mths; standard NCMP feedback letter (control). The baseline measurement in the MapMe2 study was taken from the NCMP measures. One of the most challenging aspects of this intervention was drawing a sample from the target population, to which the study results would be generalised. This abstract describes an ongoing study that takes a closer look at two existing datasets, namely the trial sampled data and national data, to evaluate the degree to which the characteristics of the sampled data match those of the wider population. To assess sampling bias, we summarised the statistical properties of data obtained from the MapMe2 intervention and the NCMP in England for the 2021/22 school year. Various conventional methods, including two-sample rank tests, were employed to identify similarities and differences between datasets. Based on the study results, the sample data's characteristics at baseline were consistent with those of the larger population. Using a range of statistical methods, we have gained valuable insight into the extent to which the sample represents the larger population. By leveraging this information, we can argue how far the findings from the MapMe2 intervention can be generalised to the entire population and assess the feasibility of introducing this intervention into the NCMP.

1. Adamson AJ, et al. Can embedding the MapMe2 intervention in the National Child Measurement Programme lead to improved child weight outcomes at one year? 2021. Trial registration: [ISRCTN12378125]. Available from: https://www.isrctn.com/ISRCTN12378125.
2. National Child Measurement Programme, England, 2021/22 school year. 3 Nov 2022 [cited 24 Mar 2023]. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/national-child- measurement-programme/2021-22-school-year.

MP110 High-dimensional selection in a matched case-control study: cardiotoxicity in childhood cancer survivors

Aba N.*[1], Belhechemi S.[2], Fresneau B.[1], Allodji S R.[1], Koscielny S.[2], Vu--Bezin G.[1], De Vathaire F.[1], Le Teuff G.[2], Haddy N.[1]
[1]Université Paris-Saclay, UVSQ, Inserm, CESP ~ Villejuif ~ France, [2]Gustave Roussy Cancer Center, Université Paris-Saclay, Biostatistics and Epidemiology Unit ~ Villejuif ~ France

While matched case-control designs are widely used in epidemiological studies, and genomic studies, statistical analyses that allow for high-dimensional feature selection generally do not take adequately account of matching. Consequently, some methods based on conditional logistic Lasso penalization [1] were developed. We illustrated one of them [2] for the research of genetic markers related to cardiac disease (CD) occurrences in childhood cancer survivors (CCS). We conducted a matched case-control study of 330 patients nested within the 7670 patients of the French Childhood Cancer Survivor Study (FCCSS) cohort. Each validated CD-case (165 cases) was matched to one CD-free control based on age at cancer diagnosis, gender, type of primary cancer, and follow-up duration (controls were followed at least until their matched caseq developed the event). The expression of 33000 genes was obtained through a transcriptome microarray. In order to to identify the best features while taking into account the matched design of the data, we performed a conditional logistic Lasso regression. To address the instability of the Lasso approach on the shrinkage parameter, robust analyses were performed using Bolasso, Sublasso and Percentile Lasso. The first two are based on resampling methods, and the latter selects the Lasso model for a tuning parameter corresponding to an appropriate percentile of the distribution of the optimal tuning parameter. The final selection corresponds to the intersection of the three sets of selected genes. Model classification performance was evaluated using Area under the Receiver Operating Characteristic curve (AUC) through an unconditional logistic regression adjusted on matching variables, and cancer treatment doses. Five promising genes were identified, their addition to the model, including clinical and treatment data, improved the discriminant ability, as compared to the model based solely on clinical and treatment data (bias-corrected AUC = 0.84 vs 0.58; P < 0.001). In summary, we illustrated one statistical method for high-dimensional genomic data selection in matched case-control study. Further works comparing other methods such as boosting for classification of matched-pairs, and random forest penalized conditional logistic regression will be of great interest.

[1] Avalos M, Grandvalet Y, Duran Adroher N, et al., Statistics in Medicine, vol. 31, 2012, p 2290-2302.
[2] Reid S, Tibshirani R., Journal of Statistical Software, vol. 58, p 12

MP111 Estimating unobserved prevalence of opiate and crack use in england

Djennad A.*[1], Harris R.[1], Jahr S.[2], Charlett A.[1], Presanis A.[3], De Angelis D.[3]
[1]UK Health Security Agency ~ London ~ United Kingdom, [2]Department of Health and Social Care ~ London ~ United Kingdom, [3]MRC Biostatistics Unit University of Cambridge ~ Cambridge ~ United Kingdom

Opiate and crack cocaine use (OCU) and injecting are associated with hepatitis C, HIV infection, drug overdose mortality, and crime. Prevalence estimates support government and treatment providers' resource and strategy planning, but many individuals with treatment needs are not in contact with the relevant services. Capture-recapture methods have been used to estimate this unobserved population, although data are sparse for small areas. We propose the use of random effects models to estimate prevalence at local and national levels. OCUs in 2018/19 from three sources - community treatment, combined Criminal Justice System (CJS; including arrests, prison treatment, probation) and drug-related deaths - were linked using Fellegi– Sunter probabilistic linkage. The resulting dataset consisted of contingency tables for individuals observed/not observed in the three above sources, stratified by age, sex, drug type, injecting status and 151 local authority areas. A log-linear area level mixed effects model was fitted, including fixed effects of the three sources, group covariates, and two-way interactions, and random effects for sources and covariates. In 2018/19 there were 136,996 individuals observed in treatment, 56,874 observed in the CJS and 2,255 drug related deaths in England for those aged 15-64 years old. The estimate of total OCU was over 300,000. Of estimated OCUs, 47.7% had a history of injecting drugs, while 52.3% have never injected any substance. Rates of OCU ranged from <5 to 29 per 1000 population across areas. Younger individuals were less likely to appear in community treatment, but more likely to be in contact with the CJS. Those with a history of injecting were more likely to be observed in community treatment but had more drug-related deaths. Large number of individuals are using opiate and/or crack cocaine in England, many who are not in treatment, or having a history of injecting and risk of exposure to blood-borne viruses. We developed a plausible, coherent model to estimate unobserved prevalence, with random effects providing efficient estimates for small areas. Probabilistic matching avoids relying on exact matching, which may underestimate source overlap and lead to biased estimates.

[1] I.P. Fellegi, A.B. Sunter, A theory for record linkage. J. Am. Stat. Ass., 64, (1969), pp. 1183–1210.
[2] G. Hay, A. Rael dos Santos, H. Reed, V. Hope, Estimates of the Prevalence of Opiate Use and/or Crack Cocaine Use, 2016/17. Sweep 13 report. Public Health Institute, Faculty of Education, Health and Community, Liverpool John Moores University.

Poster Sessions

MP112 Impact of combined exposure to multiple air pollutants on breast cancer risk using bayesian profile regression

Giampiccolo C.*[6], Amadou A.[4], Coudon T.[4], Praud D.[4], Grassot L.[4], Faure E.[5], Coudivat F.[1], Severi G.[2], Mancini F.R.[5], Fervers B.[4], Roy P.[3]
[1]*National Institute for industrial Environment and Risks (INERIS) ~ Verneuil-en-Halatte ~ France,* [2]*Centre de Recherche en Epidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine, Université Paris-Saclay, UPS UVSQ, Gustave Roussy, ~ Villejuif ~ France,* [3]*Pole Sante Publique, Hospices Civils de Lyon ~ Lyon ~ France,* [4]*Department of Prevention Cancer Environment, Centre Léon Bérard ~ Lyon ~ France,* [5]*Centre de Recherche en Epidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Médecine, Université Paris-Saclay, UPS UVSQ, Gustave Roussy ~ Villejuif ~ France,* [6]*Laboratoire de Biometrie Et Biologie Evolutive, CNRS UMR 5558 ~ Villeurbanne ~ France* T

The general population is continuously exposed to multiple and correlated air pollutants. To date, very few studies has assessed the impact of multiple exposures to air pollutants on breast cancer risk. To address this issue, specific statistical approaches are needed. Our objective is to estimate the association between the joint exposure to 8 air pollutants (benzo(a)oyrene, cadmium, dioxins, nitrogen dioxide, ozone, polychlorinated biphenyls, particulate matter and fine particles) and breast cancer risk. We use a case-control study nested within the French E3N cohort, involving 5222 incident breast cancer cases and 5222 matched controls. For each woman, an average annual exposure to the pollutants was estimated from 1990 to 2011. Two different statistical approaches were compared. The first approach consists of grouping individuals according to their exposure to pollutants, by a Dirichlet process [1], and then applying conditional logistic regression. The second method consists of using the Bayesian Profile Regression (BPR) [2] model, which groups individuals according to their exposure and risk levels, and assigns a risk to each of these groups. In both methods, odds ratios(ORs) and their 95% confidence(CI) / credible(CrI) intervals were estimated. In both approaches, 9 clusters were identified, many similar clusters in terms of exposure levels were observed, the cluster characterised by low exposures to all pollutants, except ozone was taken as reference. A consistent increase in breast cancer risk compared to the reference cluster was observed for 3 clusters in the first approach, and for 2 clusters in the second approach. The highest estimated effect is observed for the cluster represented by high exposure to all pollutants except ozone in both approaches (first approach: OR=2.25, CI=(1.59,3.19); second approach: OR=1.39, CrI=(1.06,1.83)). This is the first study assessing the effect of exposure to a mixture of 8 air pollutants on breast cancer risk, using two approaches. The first approach is aimed at describing clusters. The second approach is more interesting because individuals are grouped according to their exposure and risk level. The results show evidence of a positive joint effect of exposure to high levels of all pollutants (except ozone) on the risk of breast cancer.

[1] Molitor J, Papathomas M, Jerrett M, Richardson S. Bayesian profile regression with an application to the National survey of children's health. Biostatistics. 2010;11(3):484-498.
[2] Liverani S, Hastie DI, Azizi L, Papathomas M, Richardson S. PReMiuM: An R Package for Profile Regression Mixture Models Using Dirichlet Processes. J Stat Softw. 2015;64(7):1-30.

MP113 Effect of bias, composition, and decision-making in expert panels on diagnostic accuracy estimates

Kellerhuis B.*, Jenniskens K., Schuit E., Moons C., Hooft L., Reitsma H.
UMC Utrecht ~ Utrecht ~ Netherlands

Expert panels are commonly used as a reference standard in diagnostic test accuracy studies. We assessed how characteristics of an expert panel affect estimates of diagnostic accuracy of an index test and what choices researchers can make during the study design phase to minimize bias. We simulated various scenarios in which an expert panel was used as the reference standard for target condition diagnosis. Individual experts provided a probability estimate that the target condition was present based on component tests' results for each participant. Probability estimates from experts were combined by taking the mean, minimum, or maximum values, yielding the expert panel probability estimate of the target condition being present. Diagnostic accuracy estimates, e.g. sensitivity (Se) and specificity (Sp), of the index test were then calculated by forcing dichotomous target condition classification. We varied the following factors: (1) number of experts in the panel; (2) number of study participants; (3) target condition classification (probability) threshold; (4) prevalence of the target condition. Each scenario was repeated 1000 times. The outcome of interest was mean standard error (MSE) in the estimates of accuracy of the index test. Across varying conditions for differences between experts, the number of experts did not substantially affect the MSE of estimates of sensitivity (95% CI -0.7% to 0.2%) and specificity (-0.1% to 0.1%), nor did the number of study participants (Se [-0.1%, 0.2%]; Sp [0.0%, 0.0%]). A classification threshold of 0.5 reduced bias in sensitivity estimates compared to thresholds of 0.2 (Se [-2.1%, 0.0%]; Sp [0.0%, 1.1%]) and 0.3 (Se [-1.9%, 0.1%]; Sp [0.0%, 0.6%]). A prevalence of 0.5 reduced bias in sensitivity estimates and increased bias in specificity estimates compared to prevalences of 0.2 (Se [-7.1%, 0.2%]; Sp [0.0%, 3.8%]) and 0.4 (Se [-2.0%, 0.4%]; Sp [-0.2%, 1.7%]). Researchers can reduce bias in sensitivity and specificity estimates by selecting a target condition classification threshold of >= 0.5. Simulated expert panels with more than 2 experts or more than 360 participants did not produce less biased diagnostic performance estimates than simulated expert panels with 2 experts or 360 participants.

[1] Jenniskens K, Naaktgeboren CA, Reitsma JB, Hooft L, Moons KGM, van Smeden M. Forcing dichotomous disease classification from reference standards leads to bias in diagnostic accuracy estimates: A simulation study. J Clin Epidemiol. 2019;111:1-10.

MP114 Identification of chronic patients with acute chikungunya: an analytical method based on agreement coefficient

Nizzardo A.[1], Casamassima C.*[1], Calusi G.[1], Federico D.[1], Watson H.[2]
[1]*Aptuit (Verona) Srl, an Evotec company ~ Verona ~ Italy,* [2]*Evotec ID (Lyon) SAS ~ Lyon ~ France*

Chikungunya (CHIK) is a viral disease transmitted to humans by the Aedes mosquito. A chronic disease state with persistent arthralgias/arthritis and a profound impact on quality of life is observed in approximately 43% [1] of infected individuals with a high variability by study possibly due to no agreement in identifying chronic subjects. However, the clinical joint evaluation, patient's pain perception, and overall health status are each considered suitable predictors. To minimise variability due to subjective clinical assessments, we implemented an analytical approach based on agreement coefficients for identifying chronic CHIK subjects.
A prospective study was conducted in Peru on 59 adults with acute CHIK. The number of tender joints out of 40 examined, the pain (VAS scale), and the RAPID3 questionnaire were assessed and considered optimal predictors of chronic condition after three months from infection. Thresholds representing the presence of persisting symptoms were considered for each variable, and the Fleiss kappa and Gwet's AC1 agreement coefficients were calculated. Since there is no standard agreement about chronic thresholds, all the possible combinations have been evaluated. The analysis produced concordant results between the indicators and identified threshold combinations with a high agreement (> 0.75) between the three predictors. Five out of 2100 combinations resulted in the top five for both coefficients. The combination with the greatest agreement averaging the two coefficients' results is tender joints >1, pain >3, and RAPID3 >6; 19 subjects out of 59 (about 30%) were considered chronic. Results from the musculoskeletal stiffness questionnaire (MSQ) and the Synovitis assessment performed by power doppler ultrasound score (PDUS) demonstrated a clear difference between chronic/not-chronic subjects, validating the chronic classification. The MSQ and PDUS scores in the overall population were 3.4 (9.05) and 3.3 (4.24), respectively, versus 8.5 (13.67) and 6.1 (5.47) in the chronic subpopulation. This analytical approach could be beneficial to define the subgroup of chronic subjects in a population affected by CHIK, reducing possible bias deriving from subjective clinical evaluations. This method and the thresholds identified must be verified in other CHIK trials and could be extended for different pathologies using different predictors.

[1] Paixão ES, Rodrigues LC, Costa MDCN, et al. Chikungunya chronic disease: a systematic review and meta-analysis. Trans R Soc Trop Med Hyg. 2018;112(7):301-316.

MP115 Metastatic prostate cancer treatment: umbrella review of systematic Reviews and meta-analyses

Sirisreetreerux P., Poprom N., Numthavaj P.*, Rattanasiri S., Thakkinstian A.
Faculty of Medicine Ramathibodi Hospital, Mahidol University ~ Bangkok ~ Thailand

Prostate cancer can be described as the second most common cancer worldwide in men. The mortality rate of prostate cancer indicates around 6.7% with geographical variation. Systemic treatment initially with androgen deprivation therapy is the standard care for hormone sensitive prostate cancer aiming to reduce androgen receptors resulting in tumor shrinkage. In addition, early combination treatment with ADT and novel anti-androgen agents or ADT with chemotherapy are recommended in selected patients. We conducted an umbrella review aiming to summarize the available SRMA to evaluate the medical treatment for hormone sensitive prostate cancer.
Materials and methods: A literature search was performed for previous systematic review and meta- analysis that included only randomized controlled trials until September 2020. We systematically appraise the results of the previous SRMA, overlapping, excessive significant test and the quality of the studies. Results: A total of 4,191 studies were identified but only 27 SRMAs were included, see Figure 1. Among 27 SRMAs, 12 were network meta-analysis and 15 were direct meta-analysis. Most studies showed no statistical significance difference in overall mortality among GnRH agonists, antagonists and bilateral orchiectomy. Combination treatment are more beneficial than ADT alone in both OS and PFS outcomes with more adverse events. On the other hands, there are no OS advantage of any combination regimen over the others. There were many treatment options for metastatic HSPS patients, either ADT alone and combination treatments. Regarding the previous SRMAs, combination treatments demonstrated the clear benefit in OS and PFS over ADT alone with more AEs. Further studies are needed to compare among combination treatments.

1. Wong MC, Goggins WB, Wang HH, Fung FD, Leung C, Wong SY, et al. Global Incidence and Mortality for Prostate Cancer: Analysis of Temporal Patterns and Trends in 36 Countries. Eur Urol. 2016;70(5):862-74.
2. Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2012;61(6):1079-92.
3. Seidenfeld J, Samson DJ, Hasselblad V, Aronson N, Albertsen PC, Bennett CL, et al. Single-therapy androgen suppression in men with advanced prostate cancer: a systematic review and meta-analysis. Ann Intern Med. 2000;132(7):566-77.
4. Weckermann D, Harzmann R. Hormone therapy in prostate cancer. LHRH antagonists versus LHRH analogues. Eur Urol. 2004;46(3):279-83; discussion 83-4. 5. Cornford P, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al.
EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. Part II-2020 Update: Treatment of Relapsing and Metastatic Prostate Cancer. Eur Urol. 2021;79(2):263-82. 6. Rydzewska LHM, Burdett S, Vale CL, Clarke NW, Fizazi K, Kheoh T, et al. Adding abiraterone to androgen deprivation therapy in men with metastatic hormone-sensitive prostate cancer: A systematic review and meta-analysis. Eur J Cancer. 2017;84:88-101.
7. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225-34.

Poster Sessions

MP116 Challenges in planning and conducting evidence based studies in a pandemic: are rcts really the gold standard?

Ponzi E.*[3], Rueegg C.[4], Leblanc M.[1], Olsen I.C.[3], Valberg M.[2], Pesonen M.[4]
[1]Norwegian Institute of Public Health ~ Oslo ~ Norway, [2]Institute of Health and Society, University of Oslo ~ Oslo ~ Norway, [3]Clinical Trial Unit, Oslo University Hospital and Oslo Center for Biostatistics and Epidemiology, University of Oslo ~ Oslo ~ Norway, [4]Oslo Center for Biostatistics and Epidemiology, University of Oslo and Clinical Trial Unit, Oslo University Hospital ~ Oslo ~ Norway

The Covid-19 pandemic has highlighted the need for high quality, evidence based studies, to base the implementation of decisions and public health policies. In evidence-based medicine RCTs are often considered the gold standard, but while there have been several solid randomized trials for drug interventions and for testing vaccine effects, randomized trials for public health interventions bear many more challenges and difficulties. In this work, we want to highlight such difficulties and show how most standard ways of performing and interpreting the results from RCTs do not apply to a pandemic setting. By following the CONSORT statement, we discuss each of the components of RCTs and explain how standard RCTs can fail in estimating the intervention effect of interest when there is strong interference and contamination between participants. We start by discussing which research questions, and consequently endpoints, are appropriate and of interest for public health interventions, and whether RCTs are equipped to answer them. We then discuss internal and external validity, and which methodological challenges arise in RCTs, from the choice of design and interventions to the measurements and analyses of the outcomes, up to the interpretation and generalizability of the results. For example, although some designs, as cluster-randomized designs, appear to be suitable, they may be difficult to implement correctly and the results may not be causally interpretable on an individual level. Similarly, it is not trivial to define the population of interest, nor appropriate interventions to implement. Most importantly, randomization, which has been the very reason RCTs are considered the gold standard, does not guarantee unbiased estimates in the presence of interference among randomized units. Finally, we discuss how generalizability, and the correct interpretation of the study findings, are undermined by such challenges. We conclude that RCTs might not always be the optimal choice, and that resulting misleading conclusions on the effect of interest are particularly harmful given the status RCTs have in the evidence hierarchy. Therefore, in instances when suitable RCT methods are not available, other types of studies might prove to be a more appropriate and ethical choice.

Schulz K. F. , Altman D. G., Moher D., for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Int Med. 2010;152(11):726-32. PMID: 20335313
Valberg et al., Letter to the editor regarding "Covid-19 transmission in fitness centers in Norway - a randomized trial". BMC Public Health. 2022;22(1):2433

Poster Sessions

MP117 Using nonlinear models to identify breakpoints in disease prevalence among patients with multimorbidity

Puronaite R.*[1], Ramanauskaite D.[2], Švaikeviciene K.[5], Tarutyte G.[3], Trinkunas J.[4], Burneikaite G.[2], Kazenaite E.[6], Glaveckaite S.[2], Jakaitiene A.[7]
[1]Center of Informatics and Development, Vilnius University Hospital Santaros Klinikos; Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University; Institute of Data Science and Digital Technologies, Fac, [2]Center of Cardiology and Angiology,Vilnius University Hospital Santaros Klinikos; Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [3]Center of Informatics and Development, Vilnius University Hospital Santaros Klinikos; Department of Research and Innovation, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [4]Center of Informatics and Development, Vilnius University Hospital Santaros Klinikos; Department of Information Systems, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University ~ Vilnius ~ Lithuania, [5]Center of Endocrinology, Vilnius University Hospital Santaros Klinikos; Clinic of Internal Diseases, Family Medicine and Oncology, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [6]Department of Biomedical Reasearch, Vilnius University Hospital Santaros Klinikos; Department of Pathology, Forensic Medicine and Pharmacology, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [7]Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University; Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Vilnius University ~ Vilnius ~ Lithuania

The prevalence of diabetes and heart failure (HF) increases with age or the number of chronic diseases. However, it is known that there are potential breakpoints where the risk of disease in patients begins to increase. Since we observe that the relationship between the disease prevalence, patients' age and the number of patients' diseases is nonlinear, it is appropriate to use nonlinear methods to model this relationship. The data from the National Health Insurance Fund database, which covered the period from January 2012 to June 2014 and included patients with at least two chronic diseases, were used to estimate the prevalence of HF and diabetes. Groups of men (n = 167,759) and women (n = 260,493) with multimorbidity were divided into training (80% of the data) and testing (20% of data) groups. The methods chosen for the analysis were logistic regression, classification trees, generalised additive models and segmented (logistic) regression. R software was used to perform the analysis. Initial results showed that all methods used distinguish HF patients from healthy controls in male and female groups with relatively similar accuracy (the highest achieved AUC was 0.829 for males and 0.825 for females). The highest AUCs were achieved using generalised additive models and segmented (logistic) regression, with breakpoints at 58 and 75 years and 3 and 5 chronic diseases for men; 59, 63 and 76 years and 4 and 7 chronic diseases for women. Meanwhile, the AUCs for diabetes models were relatively low, 0.627 for men and 0.616 for women, but the models with the highest performance remain the same with segmented (logistic) regression breakpoints at 42 and 66 years and 5 chronic diseases for men and 64 years and 6 chronic diseases for women. The results suggest that nonlinear models can reasonably estimate the prevalence of HF with only the patient's age and number of chronic diseases. In contrast, in the case of diabetes, it is reasonable to look for additional risk factors that may refine the models.

MP118 Comparing approaches to rank interventions in a network meta-analysis: patients with opioid dependence

Rodrigues M.*[1], Dennis B.B.[1], Malik M.[2], Naji L.N.[1], Malhotra D.[1], Hillmer A.[1], Sanger N.[1], Parpia S.[1], Thabane L.[1], Samaan Z.[1], Sadeghirad B.[1]
1 McMaster University ~ Hamilton ~ Canada, [2]University of Toronto ~ Toronto ~ Canada

Despite the widespread use of network meta-analysis (NMA) to synthesize the effectiveness of multiple interventions, there is no consensus on how to rank treatments. Proposed approaches use different criteria for categorizing the hierarchical effectiveness of interventions, which may result in different conclusions. There exists a need to identify the optimal method for ranking treatments for patients with opioid dependence to inform decision-makers about the credibility and effectiveness of presented evidence.
We conducted a literature search from inception to September 26, 2021 to identify randomized controlled trials evaluating opioid substitution and antagonist therapies for patients with opioid dependence, and assessed intervention effectiveness for treatment retention through a random- effects, frequentist NMA. Treatment effectiveness was ranked using three approaches: the rank-heat plot (RHP) method which uses surface under the cumulative ranking curve values, and the minimally- and partially-contextualized methods, which use assessments regarding the quality of evidence, in conjunction with effect estimates or the magnitude of clinically-beneficial effects, respectively. We included 53 trials assessing 12 interventions. All approaches were similar in ranking the interventions which were the most effective (dihydrocodeine and combination medication-assisted therapy with heroin and methadone) and least effective (clonidine) at retaining patients in treatment. However, methadone ranked only moderately using the RHP and partially-contextualized approaches, but was considered 'among the most effective' interventions by the minimally-contextualized method. Different approaches to rank treatments in NMAs may impact clinical decision-making. We suggest use of the minimally contextualized approach for this clinical population, given the arbitrary thresholds and lack of considering the quality of evidence by other methods. Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. bmj. 2020;371. Brignardello-Petersen R, Izcovich A, Rochwerg B, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. bmj. 2020;371.
Ellis SG. Do We Know the Best Treatment for In-Stent Restenosis Via Network Meta-Analysis (NMA)? Simple Methods Any Interventionalist Can Use to Assess NMA Quality and a Call for Better NMA Presentation. 2015. Veroniki AA, Straus SE, Fyraridis A, Tricco AC. The rank-heat plot is a novel way to present the results from a network meta-analysis including multiple outcomes. J Clin Epidemiol. 2016;76:193-199.

MP119 Group penalized exponential tilt model for differentially methylated genes in epigenetic association study

Park H., Huang D., Sun H.*
Pusan National University ~ Busan ~ Korea, Republic of

DNA methylation is a representative epigenetic change that occurs in our body and plays an essential role in regulating gene expression as well as in cancer progressing. Identification of differentially methylated genes between two different biological conditions has been popularly studied in epigenetic association studies. However, most of statistical methods aim to detect differences in mean methylation levels between two conditions. So, they are limited to identify differences in methylation variances which have been recently observed in cancer research. We propose new statistical method based on a group-penalized exponential tilt model that essentially combines exponential tilt model and group lasso. The proposed method can identify differentially methylated genes when two biological conditions are different in methylation mean only, methylation variance only, or both. In our extensive simulation study, we demonstrated that the proposed method has superior selection performance, compared with the existing statistical methods developed for detection of differentially methylated genes. We also applied it to 450K DNA methylation data of The Cancer Genome Atlas Breast Invasive Carcinoma Collection (TCGA-BRCA). We were able to identify potentially cancer-related genes.
The proposed method is able to prioritize differentially methylated genes based on their selection probability. It was designed to detect differentially methylated genes with both methylation mean and variance signals, considering overlapping CpG sties among tens of thousands of genes.

Poster Sessions

MP120 Can we estimate a risk without observing the relevant number of cases ?

Tango T.*
Center for Medical Statistics ~ Tokyo ~ Japan

The relevant important data cannot be always collected due to the difficulty in doing so. In this presentation, we shall consider the following situation as an example: We would like to investigate the association of some adverse reproductive outcomes such as infant deaths with mothers living within 10km from municipal solid waste (MSW) incinerators with high dioxin emission levels in Japan. The study area was defined as circles of radius 10 km from MSW incinerators. To estimate the "risk profile" around the MSW incinerators, each study area was divided into ten sub-areas (called "zone") delimited by ten circles of a radii of 1,2, ...10 km. For each zone, population size or live births is available, but the relevant number of infant deaths is difficult to obtain. But we can obtain the total number of infant deaths in each study area. In this situation, we would like to estimate the profile of crude or standardized infant death rate. Taking advantage of the total observed number of infant deaths in each study area and the population size in each zone, a non-parametric estimator is proposed to get the mean profile of risk assuming the homogeneity among different MSW incinerators. Regarding the estimation of standard error of estimates, the Bootstrap method can be used. The basic idea is similar to that described in a different context elsewhere [1]. The proposed non-parametric estimator might potentially have a wide applicability to some estimation problem where we cannot observe the relevant data.

[1] Tango T. Linear equations with random variables. Statistics in Medicine 24, 2005, 3213 - 3222.

MP121 Bayesian bent-cable model for longitudinal and survival time with heterogenous random-effects

Ariyo O.*
University of Essex ~ Colchester ~ United Kingdom

In most joint modelling, the growth curve is often used to model the response process over time, which comes with the following assumptions: (i) homogeneity of random effects, (ii) normality of the error term and (iii) same monophasic functional form.These assumptions may be questionable, especially if the longitudinal response combines two or more latent classes (subpopulations) and mult-phase trajectories over time. These multiphasic changes occur gradually rather than abruptly due to biological processes developing resistance to treatment over time. This study adopt a joint modelling of Bent-cable model for longitudinal response and survival time with heterogenous random-effects distributions. We evaluate the performance of this model with simulation studies and a longitudinal study on HIV- infected patients.
Dagne GA. Heterogeneous growth bent-cable models for time-to-event and longitudinal data: application to AIDS studies. J Biopharm Stat. 2018;28(6):1216-1230. doi: 10.1080/10543406.2018.1489407. Epub 2018 Jun 28. PMID: 29953318.

*WINNERS OF ISCB CONFERENCE AWARD FOR SCIENTISTS

MP122 Detection of multiple change points in survival analysis with narrowest significant pursuit technique

Ashrafi Samia (WINNER OF ISCB44 CONFERENCE AWARD FOR SCIENTISTS), Khan M.H.R.
University of Dhaka ~ Dhaka ~ Bangladesh

Change point analysis in survival data is one of the most important analytical fields in many areas along side time series analysis. Due to some natural and even man-made unusual activity results in some biological or chemical threats to our health and arises different kinds of critical diseases which causes sudden upward trend in mortality rates and this consequences to the distributional changes in hazard rate. As this distributional shifts are not considered by the general survival time models, we extend Fryzlewicz's (Fryzlewicz, 2020) the Narrowest Significant Pursuit (NSP) technique for detecting multiple change points to the survival analysis through accelerated failure time (AFT) models. This technique identifies certain zones within data sequences automatically at a predetermined global significance level, where each zones must contain at least a single change-point. In this process, we consider change-points as sudden changes to a AFT model's underlying parameters. A multiresolution sup-norm loss is used in NSP to fit the supposed AFT model over several data regions, and the shortest interval in which the linearity is violated significantly is then identified and later the process repeats to the left and to the right till no more significant intervals can be detected. The number of change points is not necessary to assume before applying NSP. Instead, it is deduced from the data. To demonstrate method's accuracy in identifying change points, we implement the proposed method to a several number of simulation scenarios with a variety of settings and to a real SEER Prostate Cancer Data. Both the simulated study and real data application show that the suggested strategy is capable of identifying shifts in the distribution of survival data with a great satisfactory level. Though Sequential testing approach and NSP both methods give same estimation of change point locations, sequential testing approach deals with p parameters assuming p -1 of them will not be changed. But in real cases, existence of change points, all the parameters may have changed values. Our study has shown that this may not happen in applying NSP. So this is suggested that NSP is more preferable compared to other method.

[1] Fryzlewicz, P. (2020). Narrowest significance pursuit: inference for multiple change-points in linear models. arXiv preprint arXiv:2009.05431.

MP123 Comparison of methods to analyze time-to-event endpoints in trials with delayed treatment effects

Behnisch R.*, Kirchner M., Kieser M.
Institute of Medical Biometry, University of Heidelberg ~ Heidelberg ~ Germany

The advance of immuno-oncology therapies comes with the challenge to deal with the unique mechanisms of action of these drugs especially when the primary efficacy endpoint is a time-to-event endpoint. It has been shown that the most powerful methods to compare time-to-event endpoints are weighted log-rank tests with weights proportional to the hazard ratio [1]. This implies that the standard log-rank test, often required by regulatory authorities, is most powerful under proportional-hazards alternatives. This assumption is often violated by the mechanism of action leading to delayed treatment effects or crossing of survival curves resulting in a substantial loss in power. Hence, a rather long follow-up period is required to detect a significant effect in immune-oncology trials when the log- rank test is used. Another way to compensate this loss in power would be to prespecify weights proportional to the hazard ratio but is often not feasible since the exact mechanism is not known in advance. Recently, different alternatives have been advocated, e.g. the modestly weighted log-rank test, the MaxCombo test or tests based on the restricted mean survival time, which shall be investigated to assess their performance when the treatment effect is delayed.
For a better overview over the multitude of methods that have been proposed so far, we have conducted a systematic literature search. The resulting set of methods was then compared systematically with regard to type I error and power in an extensive simulation study. To incorporate different mechanisms of action, we simulate data based on a generalized linear lag model for varying times of study duration, accrual and delay as well as different treatment effects. For methods where parameters need to be prespecified, the influence of misspecification of these parameters on the power was assessed. Most of the methods control type I error and achieve reasonable power in case of proportional hazards. A delayed treatment effect results in a power reduction for all methods, but the extent of this reduction varies between methods. There is no single method that performs best in all scenarios so the choice of the optimal analysis strategy depends on the assumed delay pattern.

[1] R. Peto, J. Peto, Journal of the Royal Statistical Society. Series A, 135(2), 1972, 185-207

Poster Sessions

MP124 Adaptive group sequential designs for clinical trials with multiple time-to- event outcomes in markov models

Danzer M.F.*, Faldum A., Schmidt R.
University of Münster ~ Münster ~ Germany

Adaptive designs for the assessment of a single time-to-event outcome are well established. Problems arise when multiple endpoints are considered simultaneously. In particular, the information used to inform a redesign needs to be chosen carefully. This relates to a general issue in adaptive designs for time-to-event endpoints outlined in [1]. Although group sequential methods for multiple time-to-event outcomes are presented in [2], it is not possible to extend them to an adaptive design without limiting the information available in interim analyses. To overcome this problem, adaptive designs for multiple endpoints shall be developed that allow the use of interim information for all involved endpoints. The dependence structure between the different endpoints must be taken into account. For this purpose, they are embedded in a multi-state model. If this model is Markovian, it is possible to apply a multivariate version of the central limit theorem, as used in [2]. We use this to construct adaptive group sequential designs for testing hypotheses about the joint distribution of multiple time-to-event endpoints. Small sample properties are investigated by simulation. The testing procedure allows for data-dependent design changes based on information from all involved endpoints. It includes the standard adaptive log-rank test as a special case and differs from the methods introduced in [2] by distinguishing between different transitions in the underlying multistate model. Its practical applicability is demonstrated in a case study.

[1] P. Bauer, M. Posch. Letter to the editor. Modification of the sample size and the schedule of interim analyses in survival trials based on data inspections, Statistics in Medicine, 2004, 23, 1333-1334
[2] D. Y. Lin. Nonparametric sequential testing in clinical trials with incomplete multivariate observations, Biometrika, 1991, 78, 123-131

MP125 The effect of immune correlates on the precision of vaccine efficacy evaluation: demographic subgroups

Dudasova J.*[1], Valenta Z.[2], Sachs J.R.[3]
[1]Quantitative Pharmacology and Pharmacometrics, MSD Czech Republic; and First Faculty of Medicine, Charles University ~ Prague ~ Czech Republic, [2]Institute of Computer Science of the Czech Academy of Sciences ~ Prague ~ Czech Republic, [3]Quantitative Pharmacology and Pharmacometrics, Merck & Co., Inc. ~ Rahway, NJ ~ United States of America

Clinical trials assessing vaccine efficacy typically provide data on binary and time-to-event clinical endpoints along with subject-specific characteristics such as immune response to vaccination (immunogenicity) and demographics. Immunogenicity measurements often represent an established or a putative correlate of protection. Simulations have previously revealed that the understanding of efficacy in demographic subgroups can be substantially improved by quantifying the relationship between immunogenicity and probability of disease (or of another binary clinical endpoint) and integrating that relationship with observed immunogenicity data to obtain relatively precise estimates of efficacy in subgroups of interest. Objectives of this work are: (i) to extend the methodology of Bayesian analysis for estimating vaccine efficacy [1] to include time-to-event (TTE) analyses using immunogenicity and TTE endpoints, and (ii) to illustrate the proposed approach using a dataset from immunogenicity sub-study to herpes zoster (shingles) vaccine phase 3 clinical trial ("Shingles Prevention Study", SPS) [2]. Cox proportional hazards model is used to evaluate immunogenicity as correlate of risk (CoR) and correlate of protection (CoP). Efficacy is estimated as the proportional reduction in model-predicted risk of disease of vaccinated subjects relative to control. The 95% confidence interval associated with estimated efficacy includes uncertainty via parametric resampling of the posterior distribution for the model parameters and bootstrapping the observed immunogenicity data. The shingles vaccine example shows that the immunogenicity-based estimation of efficacy using TTE models is more precise than the standard (case-count) estimation when immunogenicity meets criteria for CoR and CoP. Immunogenicity-based zoster vaccine efficacy for 1326 participants of SPS immunogenicity sub-study in the younger age group is 58% (95% CI, 21 to 68%), and, for the older group, is 53% (95% CI, 18 to 63%). Corresponding case-count-based efficacy estimates are 55% (95% CI, -27 to 84%) and 67% (95% CI, -2 to 90%) for the younger and older groups, respectively. Use of immune correlate data in time-to-event models can increase precision of efficacy estimation (i.e., yield narrower confidence intervals) in demographic subgroups, and thus can help support even better-informed decisions by vaccine developers and public health authorities.

[1] Dudasova, J. et al. A method to estimate probability of disease and vaccine efficacy from clinical trial immunogenicity data. npj Vaccines 6, 133 (2021).
[2] Levin, M. J. et al. Varicella-zoster virus–specific immune responses in elderly recipients of a herpes zoster vaccine. The Journal of Infectious Diseases 197, 825-835 (2008). Survival analysis

MP126 Impact of censoring on estimation for restricted mean survival time in small sample size

Hashimoto H.*[1], Kada A.[2]
[1]Nagoya City University ~ Nagoya ~ Japan, [2]National Hospital Organization Nagoya Medical Center ~ Nagoya ~ Japan

The restricted mean survival time (RMST) is one of the indicators used to summarize time to event data. In randomized controlled trials with small sample sizes, it has been pointed out that the distribution of differences in RMST deviates markedly from a normal distribution, and several countermeasures have been proposed [1,2]. In the estimation of one-sample RMST, even when the sample size is small, a good confidence interval can be estimated by applying a logit transformation [3]. However, this has not been examined in situations where there is censoring. We investigated by simulation the situation of censoring for a one-sample Weibull-distributed function. Specifically, we assumed several event and censoring distributions, and combined them into several situations. Confidence intervals for the RMST were estimated using no transformation and variable transformations (arcsine square root, logit, and complementary log-log transformations). Performance was evaluated by the coverage probability and the above and below error probabilities for the true value. In situations where the event occurrence rate was low, the proportion that could not be estimated increased as the censoring probability increased. In terms of the effect on estimation, censoring increased the coverage probability and decreased the above error probability in situations where the event occurrence rate was low. On the other hand, there was no change in the below error probability. Variable transformations reduced these effects of censoring on the estimation compared to no transformation.
In situations where the event occurrence rate is low, the censoring has strong impact and we need to pay attention. In addition, variable transformations may reduce its impact.

[1] Lawrence J, Qiu J, Bai S, Hung HJ. Difference in restricted mean survival time: small sample distribution and asymptotic relative efficiency. Stat Biopharm Res. 2019;11:61-66.
[2] Horiguchi M, Uno H. On permutation tests for comparing restricted mean survival time with small sample from randomized trials. Stat Med. 2020;39(20):2655-2670.
[3] Hashimoto H, Kada A. A Note on Confidence Intervals for the Restricted Mean Survival Time Based on Transformations in Small Sample Size. Pharm Stat. 2022;21(2):309-316.

MP127 Risk factors and transitional probability of clinical events in korean ckd patients using the multistate model

Kim J.*[1], Kim J.H.[2]
[1]University of Suwon ~ Hwaseong ~ Korea, Republic of, [2]Seoul National University Hospital ~ Seoul ~ Korea, Republic of

Compared to western countries, Korean CKD patients show distinctive differences in clinical event outcomes including lower cardiovascular disease (CVD) and higher end-stage kidney disease (ESKD) events. This study analyzed the risk factors, transition probability and cumulative hazards associated with clinical events using the multi-state model. Among 1423 patients (age 54 [44-63] years), the overall prevalence of clinical events were the following: ESKD (22.6%), CVD (7.5%), death (3.3%), death after ESKD (3.6%) and death after CVD (1.2%). Different risk factors were associated with different clinical outcomes and in particular the risk factors associated with higher ESKD event risks were underlying CVD, diabetes, polycystic kidney disease, fibroblast growth factor-23 while hypertension, increased age and estimated glomerular filtration rates were associated with lower risks. The 10-year progression probability for each event status include the following: 0.23 for ESKD, 0.08 for CVD, 0.04 for death, 0.09 for death after ESKD and 0.01 for death after CVD. The 10-year cumulative hazard estimates for each event status were the following: ESKD [0.43, 95% CI (0.37-0.49)], CVD [0.12, (0.10-0.15)], death [0.05, (0.03-0.06)], death after ESKD [0.52, (0.20-0.84)] and death after CVD [0.27, (0.15-0.40)].
Different risk factors were associated with varying clinical outcomes in Korean CKD patients. The 10- year progression probability was the highest in ESKD followed by death after ESKD events. Also, the 10-year cumulative hazard estimate was the highest for death after ESKD followed by ESKD events. These findings correlate with the distinctive clinical outcome features of Korean CKD patients.

Vejakama, P., Ingsathit, A., McEvoy, M., Attia, J., & Thakkinstian, A. (2017). Progression of chronic kidney disease: an illness-death model approach. BMC Nephrology, 18(1), 205.
Ross-Driscoll, K., & Patzer, R. E. (2022). Competing Risks and Multistate Models in Clinical Nephrology Research. Kidney Int Rep, 7(11), 2325-2326.

Poster Sessions

MP128 Bias reduction for semi-competing risks model with rare events: application to a chronic kidney disease

Kim J.*[1], Jeong B.[2], Ha I.D.[3], Oh K.[4], Lee D.[2]
[1]Medical Research Collaborating Center, Seoul National University Hospital ~ Seoul ~ Korea, Republic of, [2]Department of Statistics, Ewha Womans University, ~ Seoul ~ Korea, Republic of, [3]Department of Statistics, Pukyong National University ~ Busan ~ Korea, Republic of, [4]Department of Internal Medicine, Seoul National University Hospital ~ Seoul ~ Korea, Republic of

In a semi-competing risks model in which a terminal event censors a non-terminal event but not vice versa, the conventional method can be used to predict clinical outcomes by maximizing likelihood estimation. However, this method can produce unreliable or biased estimators when the number of events in the datasets is small. Specifically, parameter estimates may converge to infinity, or their standard errors can be very large. Moreover, terminal and non-terminal event times may be correlated, which can account for the frailty term. Here we adapt the penalized likelihood with Firth's correction method for gamma frailty models with semi-competing risks data to reduce the bias caused by rare events. The proposed method is evaluated in terms of relative bias, mean squared error, standard error, and standard deviation in comparison with conventional methods through simulation studies. The results of the proposed method are stable and robust even when data contain only a few events with the misspecification of the baseline hazard function. We also illustrate a real example with a multi-center, patient-based cohort study to identify risk factors related to chronic kidney disease progression or adverse clinical outcomes.
This study will provide a better understanding of semi-competing risk data in which the number of specific diseases or events of interest is rare.

Ha, I. D., Xiang, L., Peng, M., Jeong, J.-H., and Lee, Y. (2020). Frailty modelling approaches for semi- competing risks data. Lifetime data analysis, 26(1):109–133.
Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. Biometrics, 66(3):716–725.
Oh, K.-H., Park, S. K., Park, H. C., Chin, H. J., Chae, D. W., Choi, K. H., Han, S. H., Yoo, T. H., Lee, K., Kim, Y.-S., et al. (2014). Know-ckd (Korean cohort study for outcome in patients with chronic kidney disease): design and methods. BMC nephrology, 15(1):1–9.

MP129 Prevalence-incidence mixture model for the risk of high-grade cervical lesion based on individual risk factors

Kroon K.*, Bogaards J., Berkhof J.
Amsterdam UMC ~ Amsterdam ~ Netherlands

Cervical cancer screening is moving from a 'one-size-fits-all' approach towards a more efficient and effective personalized risk-based approach. This requires accurate risk assessment of high-grade cervical lesions or cancer (CIN2+) in high-risk human papillomavirus (HPV) positive women. Some women have underlying CIN2+ at baseline, which is detected during follow-up, so we need to distinguish between prevalent and incident CIN2+ in a model. Furthermore, incident CIN2+ can be caused by either a temporarily elevated risk due to the HPV infection at baseline or by a new infection acquired during follow-up (background risk of CIN2+). Other models exist for estimating the cumulative risk of CIN2+, however their parameters do not have a meaningful biological interpretation, such as expected time to disease which is important for determining screening intervals, and are difficult to compare between different settings. We present a biologically-driven variant of a prevalence-incidence model for interval-censored data to estimate the cumulative risk of CIN2+ in HPV+ women. Our model includes parameters for the probability of prevalent CIN2+, progression to CIN2+, and viral clearance, which can all depend on covariates. For incident disease our model assumes exponential distributions for the competing progression and clearance event combined with background risk. The parameters are estimated with the expectation-maximisation (EM) gradient algorithm, which we have implemented in a newly developed R package. We validated the model with simulation studies and applied it to the POBASCAM and VUSA-Screen studies –population-based cervical cancer screening studies from two regions in the Netherlands with different HPV positivity rates. We found that the parameter estimates from both studies are equivalent conditional on risk factors. Cumulative risk of CIN2+ in HPV+ women can be accurately predicted by our model. In addition, our model provides parameters as expected time to CIN2+ that are relatively easy to interpret and important for decision-making. Our model can be used to predict the impact of multiple rounds of screening and inform changes in the screening interval when moving towards a more personalised approach based on individual risk-factors. It is also applicable to other situations where a population has a temporary strongly elevated disease risk at baseline.

MP130 Comparison of total event analysis in three cardiovascular trials with composite outcomes

Lee S.*, Ramasundarahettige C., Gerstein H., Mcintyre W.F., Bangdiwala S.I., Thabane L.
Population Health Research Institute ~ Hamilton ~ Canada

: Cardiovascular (CV) randomized controlled trials (RCTs) frequently use composite endpoints for analyzing the time-to-first incidence of any event within the composite. However, incorporating recurrent events of each component could enhance the efficiency of these studies. Unfortunately, utilizing the traditional Cox model for total event analysis may be problematic in the presence of between-subject heterogeneity resulting from multiple events within a subject [1]. To address this issue, several extensions of the Cox model have been developed [2], including the Andersen-Gill (AG), Prentice-Williams-Peterson (PWP), or Wei-Lin-Weissfeld (WLW) approaches. This study aims to compare different methods for analyzing all events within a composite concerning between-subject heterogeneity using three large multi-center cardiovascular trials. The study estimated the between-subject heterogeneity for the composite outcomes and its components and evaluated the treatment effect of each method as a hazard ratio (HR) with its corresponding 95% confidence interval (CI). The two composite outcomes were the occurrence of myocardial infarction (MI), stroke, and cardiovascular (CV) death, as well as the aforementioned composite outcome along with heart failure hospitalization. The proportion of events for the composite increased from 16.4% to 22.3% in ORIGIN, 4.8% to 5.5% in COMPASS, and 13.9% to 17.1% in TRANSCEND, corresponding heterogeneity to each study of 2.4, 4.0 and 2.0, respectively. In ORIGIN, the HR for the first composite was 1.03 (95%CI, 0.94-1.12, width=0.18) for Cox with only first events, 1.01 (95%CI, 0.92-1.11, width=0.19) for AG, 1.02 (95%CI, 0.94-1.10, width=0.16) for PWP total, 1.01 (95%CI, 0.94-1.09, width=0.15) for PWP gap, and 1.03 (95%CI, 0.94-1.12, width=0.18) for WLW. Similar results were found in other studies and when examining the second composite. The results show that the proportion of events for the composite increased substantially in all three trials, with a corresponding small heterogeneity. The treatment effects for time-to-first and total events methods were robust and consistent, with only small differences in treatment effect and CI width across different methods. Therefore, this study suggests incorporating recurrent events could benefit cardiovascular trials of composite outcomes associated with small heterogeneity without compromising the accuracy of the treatment effect.

[1] Claggett B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of Time-to- First Event and Recurrent-Event Methods in Randomized Clinical Trials. Circulation. 2018 Aug 7;138(6):570-577.
[2] Ozga AK, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. BMC Med Res Methodol. 2018 Jan 4;18(1):2.

MP131 Non-parametric bayesian imputation of right censored data in survival analysis

Moghaddam S.*[1], Newell J.[2], Hinde J.[2]
[1]University of Limerick ~ Limerick ~ Ireland, [2]University of Galway ~ Galway ~ Ireland

In survival analysis due to censoring, methods of plotting individual survival time, such as density plots, are invalid. The median survival is often used as a summary of the survival experience of a patients' population. However, it is unlikely that the median is a relevant summary at the patient level [1] and a density plot of the data is perhaps more informative for communication than a single summary statistic. A fundamental idea in this research is to consider censored data as a form of missing, incomplete, data and use approaches from the missing data literature to handle this issue. Non-parametric Bayesian methods are considered to impute right censored survival data to achieve this aim. A common motivation in using non-parametric Bayesian methods is to account for model uncertainty about the choice of a parametric distribution. Under the non-parametric Bayesian paradigm, the unknown distribution of the model is treated as a random parameter with stochastic non-parametric priors, such as the Dirichlet process. Posterior predictive distribution is used to impute censored observations. The results of the non-parametric Bayesian imputation compared to our parametric Bayesian approach [2] and it shows that in the situations where the true distribution of the data is unknown, the non-parametric Bayesian approach provide better estimation of censored observations in comparison to the parametric Bayesian approach. However, this needs to be balanced against the fact that it is much more computationally intensive and more time-consuming. The imputation of censored observations not only allows more interpretable graphics to be produced [3] for a wider general audience (physicians and patients), but it opens up the possibility of the use of standard formal methods of analysis for continuous responses.

[1] Gould, S.J., 2010. The median isn't the message. Ceylon Medical Journal, 49(4).
[2] Moghaddam, S., Newell, J. and Hinde, J., 2022. A Bayesian Approach for Imputation of Censored Survival Data. Stats, 5(1), pp.89-107.
[3] Royston, P., Parmar, M.K. and Altman, D.G., 2008. Visualizing length of survival in time-to-event studies: a complement to Kaplan–Meier plots. Journal of the National Cancer Institute, 100(2), pp.92- 97.

Poster Sessions

MP132 Testing for sufficient follow-up to detect the cured proportion

Musta E.*, Yuen T.P.
University of Amsterdam ~ Amsterdam ~ Netherlands

With increased cure chances for many cancer types, it is clinically of interest to estimate the proportion of cured patients. It is known that, in order to detect the cured proportion, the follow-up needs to be sufficiently long, but it is not clear what is the minimum required follow-up time. In practice, a Kaplan- Meier curve with a long plateau, containing many censored observations, is considered as an indication of sufficient follow-up. However, it is often difficult to assess based on this visual inspection. Maller and Zhou [1] were the first to propose a statistical test for the assumption of sufficient follow-up and few other tests have been later designed based mainly on [1]. However, such attempts have not yet been satisfactory for practical purposes because of their conservative behavior or the underlying assumptions. Our goal is to develop a novel method for testing sufficient follow-up, under general non- parametric assumptions, that is reliable for practical use. We follow a different approach compared to the existing attempts in the literature. First, we formulate the hypotheses in a more general way, not requiring that the times of the event of interest have compact support. Instead, our notion of sufficient follow-up is based on the quantiles of the distribution and is more realistic. The underlying assumption for our method is that the density function of the event times is unimodal, which is reasonable in most of the applications of interest and includes several parametric families commonly used in survival analysis. The test is based on a nonparametric shape constrained density estimate, appropriately corrected at the boundary [2]. In this way, we do not only rely on observations close to the end of the study, where estimators are unstable because of censoring. We investigate the behavior of the test through a simulation study and illustrate its practical use on some open-source data of cancer clinal trials. Testing sufficient follow-up remains a challenging problem and medical knowledge is required aside any statistical procedure. The proposed method shows satisfactory behavior in practice and can also be used as a visual diagnostic tool.

[1] Maller, R. A., and S. Zhou. "Testing for sufficient follow-up and outliers in survival data." Journal of the American Statistical Association 89, no. 428 (1994): 1499-1506.
[2] Lopuhaä, H. P., and E. Musta. "Smooth estimation of a monotone hazard and a monotone density under random censoring." Statistica Neerlandica 71, no. 1 (2017): 58-82.

MP133 Approaches to deal with non-proportional hazards in paediatric oncology - the context matters

Pötschger U.*[1], Heinzl H.[2], Mittlböck M.[2]
[1]Children's Cancer Research Institute ~ Vienna ~ Austria, [2]Medical University of Vienna, Center for Data Science ~ Vienna ~ Austria

In oncology, survival is the most important clinical endpoint and log-rank test, or Cox-regression are almost exclusively used for statistical evaluation. These approaches are not valid in the presence of non-proportional hazards, and there is no universally accepted best way, how to proceed then. Non- proportional hazards are a common problem in paediatric oncology where cure of the disease (increase of long-term survival probabilities) is the most interesting outcome. Including a time-by-covariate interaction in a Cox-regression is the predominant approach to detect non-proportional hazards and assess their effect. However, this approach does neither evaluate survival times nor long-term survival probabilities, and hence, the results frequently do not answer the underlying medical research question of long-term effects. The modelling of restricted mean survival (RMS), that is the average survival from time zero to a pre-specified time point, has recently been proposed to deal with non-proportional hazards. RMS has a clear survival interpretation which can be useful when prolongation of the remaining lifetime is of main interest. RMS is not suitable, when the focus is on patients' cure like in paediatric oncology. The third approach is the so-called pseudo-value regression assessing survival at a pre-specified time point and thus particularly suited to investigate long-term survival probabilities. An illustrative example is provided with data from neuroblastoma patients, where the prognostic value of age and MYCN-amplification on long-term survival is to be assessed. Cox regression, log-rank test, and RMS provide misleading results. In contrast, the results of pseudo-value regression provide an unambiguous survival interpretation. Pseudo-value regression and its extensions can be recommended to evaluate survival in children with cancer. They should become part of the standard repertoire of statisticians working in paediatric oncology. Andersen, P. K. and M. P. Perme "Pseudo-observations in survival analysis." Stat Methods Med Res 2010, 19(1): 71-99. Pötschger U, Heinzl H, Valsecchi MG, Mittlböck M. Assessing the effect of a partly unobserved, exogenous, binary time-dependent covariate on survival probabilities using generalised pseudo- values. BMC Med Res Methodol. 2018 Jan 19;18(1):14. Mittlböck M, Pötschger U, Heinzl H. Weighted pseudo-values for partly unobserved group membership in paediatric stem cell transplantation studies. Stat Methods Med Res. 2022 Jan;31(1) Ambrogi F, Iacobelli S, Andersen PK. Analyzing differences between restricted mean survival time curves using pseudo-values. BMC Med Res Methodol. 2022 Mar 18;22(1):71.

MP134 Asking the right question when assessing os in trials that allow for cross over: considerations and case study

Previtali A.*, Colicino S.
Bristol Myers Squibb ~ Boudry ~ Switzerland

Treatment switching (TS) occurs in randomized clinical trials when patients discontinue their randomly assigned treatment and start a new therapy. Although ethically and clinically justified, TS presents a difficult problem for statisticians trying to ascertain the causal effects of interventions, particularly when assessing overall survival (OS). The difficulty arises as TS may occur after randomization but before observing the variable of interest (e.g., a death occurring after TS). Following the spirit of ICH E9(R1) on estimands, the clinical question should drive how TS is handled. In practice, especially when seeking regulatory approval, OS is assessed using an intention-to-treat (ITT) approach comparing treatments as they were initially randomized (i.e., regardless of TS). Additional methodologies (namely, RPSFT, 2-AFT and IPCW) were developed to assess treatment effect (TE) in the hypothetical scenario in which TS would not have occurred. These methods allow the relative effect of experimental over control treatment to be isolated by removing (or reducing) the potential benefit of switching. Focusing on a special case of TS hereby referred to as cross over (CO), in which patients are only allowed to switch from control to experimental, this work aims to i) contextualize these analyses in terms of the clinical question, ii) clarify their interpretation and role in regulatory submissions and iii) provide an example of their application using a case-study in cell therapy In the case-study TRANSFORM (NCT03575351), when TE was estimated ignoring TS, the study did not demonstrate an improvement in OS (HR = 0.724; 95% CI: 0.443, 1.183). However, when the TE was estimated assuming CO did not occur, results from 2-AFT and RPSFT showed a favorable TE (HR = 0.415; 95% CI: 0.251, 0.686 and HR = 0.279; 95% CI: 0.145, 0.537, respectively). IPCW could not be implemented due to data limitations The definition of clear clinical objectives is paramount to decide how TS should be analytically addressed. While regulators may be more inclined to focus on the ITT analysis, other agencies such as payers may be also interested in evaluating the relative effect assuming TS did not occur. Together, these approaches contribute to a comprehensive efficacy evaluation NA for the abstract

MP135 Adverse events in trials with survival outcomes: from clinical questions to methods for statistical analysis

Tassistro E.*, Antolini L., Bernasconi D.P., Valsecchi M.G.
University of Milano-Bicocca ~ Monza ~ Italy

When studying a novel treatment with a survival time outcome, failure can be defined to include an adverse event (AE) among the endpoints typically considered, for instance relapse. These events act as competing risks, where the occurrence of relapse as first event and the subsequent treatment change exclude the possibility of observing AE related to the treatment itself.
In principle, the analysis of AE could be tackled by two different approaches: It requires a competing risk framework for analysis: the clinical question relates to the observed occurrence of AE as first event, in the presence of the event "relapse"; It requires a counterfactual framework for analysis: the clinical question relates to the treatment causing AE occurrence as if relapse could not occur. This work has two aims: the first is to critically review the standard theoretical quantities and estimators with reference to their appropriateness for dealing with approaches 1 or 2 and to the following features: (a) estimators should address for the presence of right censoring; (b) theoretical quantities and estimators should be functions of time. The second aim is to define a strategy to relax the assumption of independence between the potential times to the competing events of the commonly used estimators when counterfactual approach 2 is of interest.
Method(s) and Results: After reviewing the standard methods we clarify the impact of the crucial assumption of independence between potential times to competing events of the standard estimators used in the counterfactual approach. We propose the use of regression models, stratified Kaplan- Meier curves and inverse probability of censoring weighting to relax the assumption of independence by achieving conditional independence given covariates and we develop a simulation protocol to show the performance of the proposed methods.
Conclusions: The proposed methods overcome the problem due to the dependence between the two potential times. In particular, one can handle patients' selection in the risk sets, and thus obtain conditional independence between the two potential times, adjusting for all the observed covariates that induce dependence.

[1] A. Allignol, J. Beyersmann, C. Schmoor (2016). Statistical issues in the analysis of adverse events in time-to-event data. Pharmaceutical Statistics, 15, 297-305
[2] S.J.W. Willems, A. Schat, M.S. van Noorden, M. Fiocco (2018). Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. Biometrical Journal, 62, 836-851

TUESDAY 29 AUGUST 2023

POSTER SESSIONS 2

TP1	Huda Alsulami	*Methods for combining dependent p-values with side information and application to genetics*
TP2	Maria Blanco	*Deriving interpretable thresholds for variable importance in random forests by permutation*
TP3	Dominic Edelmann	*Fast and precise survival testing for genome-wide association studies*
TP4	Katarzyna Górczak	*Hierarchical negative-binomial model for analysis of correlated sequencing data: practical implementations*
TP5	Benjamin Hivert	*Contributions and challenges of data fission for post-clustering differential analysis*
TP6	Stanislav Katina	*Comparison of standard linear and ridge regression models used in the shape index calculation*
TP7	Claudio Junior Salaroli	*A statistical methodology to select and combine covariates for disease classification in high-dimensional data*
TP8	Michalis Katsoulis	*Reduced visits at emergency departments for cardiac conditions and cardiac mortality in the covid-19 pandemic*
TP9	Raimondo Bruno	*Applying mnlfa to examine consistency in dsm-5 opioid use disorder between pain patients and people who inject*
TP10	Diletta Fabrizi	*Dyadic approaches for patient self-care and caregiver contribution to self-care assessment in type 2 diabetes*
TP11	Laura Savaré	*Latent markov model for profiling heart failure patients' adherence to drugs*
TP12	Daniel Bodden	*Challenges in group sequential designs in rare diseases*
TP13	Audrone Jakaitiene	*Multiple regression models for predicting rare outcomes in hypertrophic cardiomyopathy*
TP14	Pallavi Rane	*A review of the design of clinical trials in rare cancers*
TP15	Tomoharu Sato	*Clinical trial design and estimation methods that can yield additional information than single-arm trials*
TP16	Stefanie Schoenen	*The impact of allocation bias on the test decisions in clinical trials with multiple endpoints*
TP17	Mikolaj Swiderski	*Addressing biases in odds ratios for rare chronic skin conditions using real-world prescription drug exposures*
TP18	SATOSHI YOSHIDA	*Global rank test with data-driven optimal weights in clinical trials with multiple endpoints*
TP19	Angela Alibrandi	*The non parametric combination methodology to analyze influence of food regimes on oxidative stress parameters*
TP20	Emilie Barre	*Exploring the multidimensional benefits of a new treatment with generalized pairwise comparisons*
TP21	Lubomir Stepanek	*Nonparametric comparison of two proportions: le cam's theorem and chernoff bound applied to upper-lower index*
TP22	Urko Aguirre	*Identification of risk factors for wound non-healing: a descriptive study.*
TP23	Alberto Alvarez-Iglesias	*An extension of the numbers-needed-to-treat concept based on net benefit*
TP24	Gregor Buch	*Comparison of selection strategies to identify biostatistical methods - case study on group variable selection*
TP25	Bram Burger	*Optimising sample allocation to batches in laboratory-based observational studies*
TP26	Andrea Corbetta	*Genetic and environmental determinants of drug adherence*
TP27	Marian Mitroiu	*The ich e9(r1) estimand framework adapted in a phase iii equivalence rct conducted during covid-19 pandemic.*
TP28	Giulia Gambini	*Adherence of spirit guidelines in no-profit clinical trials*
TP29	Philip Hougaard	*A new testing strategy for a trial with two doses compared to placebo*
TP30	Wilmar Igl	*Federated analysis of multiple data sources*
TP31	Insu Jang	*3div: a comprehensive database of 3d genome and 3d cancer genome*
TP32	Bartosz Jenner	*Comparison of adverse event burden between treatment groups in clinical trials using mixture models*
TP33	Miyu Kobayashi	*A simultaneous evaluation of superiority and non-inferiority for comparing screening tests.*
TP34	JAEHO LEE	*Korean nucleotide archive as a new data repository for nucleotide sequence data*
TP35	Kim Luijken	*Replicability of simulation studies for the investigation of statistical methods: the replisims project*
TP36	Hugo Luttenauer	*Dealing with sparse networks of interventions: learnings from a simulation study*
TP37	Eirini Pagkalidou	*Diagnostic network meta-analytic methods for pulmonary embolism*
TP38	Minsu Park	*Generalized outlier detection for skewed distributions in biomedical signal*
TP39	Jan Rekowski	*Practical guide on statistical items in the new spirit-define extension for early phase dose-finding trials*
TP40	Laurent Renard Triché	*Sample size estimation in clinical trials using ventilator-free days as primary outcome: a systematic review.*
TP41	Ettore Rocchi	*Myofibril linearity indexes in muscle: a montecarlo-based analysis of performance*
TP42	Myanca Rodrigues	*Variability in primary outcome reporting in clinical trials for older adults with depression*
TP43	Marcia Rueckbeil	*Collaboration opportunities on biostatistics with ema*
TP44	Marzieh Shahmandi	*Evaluating zero-inflated models with a different base distribution in clinical trials and medical research*
TP45	Kanae Takahashi	*Hypothesis testing procedure of f1 scores for binary and multi-class classification in the paired design*
TP46	Sara Urru	*Foreign body injuries recognition and management through text analysis of case reports*
TP47	Kazue Yamaoka	*On the cluster randomised trials where intervention effects are heterogeneous*
TP48	Byoung-Ha Yoon	*Application of graph theory to integrate complex relationships among heterogeneous biological data*
TP49	Yue ZHAI	*Performance comparisons between clustering models for reconstructing ngs results from technical replicates*
TP50	Samuel Zimmermann	*Text classification to automate abstract screening using machine learning*
TP51	Andrea Cappozzo	*Constructing dna methylation biomarkers for cardiovascular diseases by penalized multilevel multitask learning*
TP52	Carolien Maas	*Restricted mean survival time provides intuitive and robust absolute treatment effect estimates in risk strata*
TP53	Alessia Mapelli	*Multi-outcome feature selection via anomaly detection autoencoders for radiogenomic in breast cancer patients*
TP54	Mehran Moazeni	*A personalized algorithm to detect cardiac arrythmia and major bleeding in advanced heart failure patients*
TP55	Teun Petersen	*Personalized scheduling of biomarker measurements for heart failure surveillance programs with competing risks*
TP56	Marika Vezzoli	*The clivus reconstruction: a novel method for identifying the optimal set of scaffolds*
TP57	Nina Deliu	*Enhancing patient outcomes and statistical efficiency in rare-disease phase-ii trials: the stratosphere study*

Poster Sessions

TP1 Methods for combining dependent p-values with side information and application to genetics

Alsulami H.*
Queen Mary University of London, United Kingdom ~ London ~ United Kingdom

In bioinformatics, combining p-values from various statistical tests is a common procedure. Methods for combining multiple dependent p-values that maintain control of the false discovery rate are crucial in hypothesis testing, especially in the analysis of genetic data [2]. This work aims to present and compare different combination methods and study the importance of the correlation among the combined p-values. Also, an overview of the various weighting schemes in the literature and an application of a weighted p-values method utilising prior information are presented. In this work, the impact of the correlation on the significance of the combined p-values is investigated using four existing methods in a simulation study. Data from a multivariate normal distribution are simulated and the properties of these methods are compared with respect to type 1 error rate and power. Methods based on parametric estimation of the covariance matrix maintain the nominal type 1 error rate. However, the nonparametric approach has higher type 1 error rates, especially for small sample sizes. As expected, ignoring the correlations and using methods that combine independent p- values gives conservative or liberal type 1 error rates. In general, all methods have comparable power where it is affected by the sample size, effect size and the sparseness of the signals. Prior information is a valuable component that could be incorporated into the analysis through weights which give different levels of importance to the hypotheses of the combined test. It has been shown that weighted p-values substantially improve the power and accuracy of the combined p-value method. For this reason, we demonstrate an application to genome data from a cancer study using this technique and exploit the advantage of the available biological data. Assigning the optimal weights is still challenging, yet plausible choices would improve the power to detect significant effect sizes [1]. Our results show that the power to detect an effect size may be markedly affected when the combined p-value test does not appropriately account for the correlation among the individual p-values. Moreover, the availability of information resources, either in the literature or biological databases, has a significant impact on the power.

[1] G. Alves, YK. Yu, PloS one, 9, 2014, e91225.
[2]Y. Liu, J. Xie, Journal of the American Statistical Association,115, 2020, 393-402.

TP2 Deriving interpretable thresholds for variable importance in random forests by permutation

Blanco M.*[1], Schlieker L.[1], Mueller T.[1], Ott A.[2], Buchner H[1]
[1]*Staburo GmbH ~ Munich ~ Germany,* [2]*Roche Diagnostics GmbH ~ Penzberg ~ Germany*

In the context of clinical research and in particular precision medicine the identification of predictive or prognostic biomarkers is of utmost importance. Especially when dealing with high-dimensional data discriminating between informative and uninformative variables plays a crucial role. Machine Learning approaches and especially Random Forests are promising approaches in this situation as the Variable Importance (VIMP) of a Random Forest can serve as a decision guidance for the identification of potentially relevant variables. Many different approaches for Random Forest VIMP have been proposed and evaluated (e.g., Speiser et al. 2019). One of the algorithms is the well-performing Boruta method (Kursa and Rudnicki 2010), which adds permutated - and thus uninformative - versions of each variable (so-called shadow variables) to the set of predictors. Based on that, it classifies the covariables into three possible importance categories: confirmed, tentative or rejected. We propose a variation of the Boruta method and evaluate the relevance of the variables based on different criteria. Our method is independent of the simulations runs and compares the VIMP of each covariate directly with its permuted version. In addition, the uninformative versions are generated by permutating the rows of the dataset, which preserves the relationship between the original variables. For evaluating the importance of the features we use different criteria, e.g., proportion of positive difference in paired VIMP, descriptive statistics of the shadow VIMPs and distance metrics between paired distributions. We examine our method on real public available datasets of varying sizes and compare its performance to the Boruta algorithm. In our approach, the user is guided by several criteria summarized in one visual presentation and has therefore the flexibility to be more conservative by picking all important covariates or more permissive by taking only the most important ones, depending on the nature of their problem.

[1] Speiser JL, Miller ME, Tooze J, Ip E. A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. Expert Syst Appl. 2019;134:93-101. doi:10.1016/j.eswa.2019.05.028
[2]Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. Journal of Statistical Software, 36(11), 1–13. https://doi.org/10.18637/jss.v036.i11

TP3 Fast and precise survival testing for genome-wide association studies

Yu T., Benner A., Edelmann D.*
German Cancer Research Center ~ Heidelberg ~ Germany

Genome-wide association studies (GWAS) typically involve analyzing millions of single-nucleotide polymorphisms for their association with survival responses. To adjust for multiple testing, a stringent genome-wide significance level of α=5x10-8 is commonly used. In this context, standard survival statistics based on the Cox model such as the Wald or the Score test are unable to reliably control the type I error rate. On the other hand, more reliable alternatives, such as the Firth correction, are computationally expensive, making them impractical for large datasets. The objective of this study is to compare the type I error rate, power, and runtime of various Cox model-based survival tests. We compared the type I error rate, power, and runtime of various Cox model-based survival tests, including the Score test, Wald test, Likelihood ratio test, Firth correction, and a saddle-point approximation based Score test (SPACOX) using simulations and real data from the UK Biobank. Our findings reveal that the Wald and Score tests are highly anti-conservative for low minor allele frequencies (MAFs) and/or event rates, whereas SPACOX is substantially conservative in some settings. Furthermore, these tests exhibit different behavior depending on the direction of the effect. Except for score test-based procedures, the runtime of all tests is prohibitively high, particularly for the Firth correction. To address this challenge, we propose a fast and precise testing procedure for GWAS based on prescreening via an extremely efficient version of the Score test, followed by testing of the screened subset of genes using Firth correction or Likelihood ratio test. We demonstrate the performance of our procedure using simulations and real data from the UK Biobank. In conclusion, standard survival statistics based on the Cox model such as the Wald or the Score test are unable to reliably control the type I error rate in GWAS. We propose a fast and precise testing procedure based on prescreening via an extremely efficient version of the Score test, followed by testing of the screened subset of genes using Firth correction or Likelihood ratio test. This method provides a practical and accurate alternative for GWAS that can be applied to large datasets.
Bi, Wenjian, et al. "A fast and accurate method for genome-wide time-to-event data analysis and its application to UK Biobank." The American Journal of Human Genetics 107.2 (2020): 222-233

TP4 Hierarchical negative-binomial model for analysis of correlated sequencing data: practical implementations

Górczak K.*
Hasselt University ~ Hasselt ~ Belgium

High-throughput techniques for biological and (bio)medical sciences often result in read counts used in downstream analysis. Nowadays, complex experimental designs in combination with these high- throughput methods are regularly applied and lead to correlated count-data measured from matched samples or taken from the same subject under multiple treatment conditions. Additionally, as is common with biological data, the variance is often larger than the mean, leading to overdispersed count data. Hierarchical models [1] have been proposed to analyze overdispersed, correlated data from paired, longitudinal or clustered experiments. We shortly discuss the use of the negative-binomial model with normally-distributed random effects for differential analysis of sequencing data. We also compare different implementations of the model in R, Python, SAS and STATA.
We focus on the multivariate negative-binomial model with random effects for the analysis of correlated, overdispersed sequencing data [2]. We consider several implementations in R (GLMMadaptive, lme4, MNB and glmmTMB), Python, SAS (PROC NLMIXED) and STATA (menbreg command), using both simulated and real-life datasets. We compare the results obtained for the different implementations in function of numerical solutions (such as the Laplace approximation or (adaptive) Gaussian-Hermite quadrature) used to compute and maximize the marginal likelihood. The majority of the available implementations (in R, STATA, Python) allow fitting the model with random intercepts. In SAS, using more complex random-effect structures is possible. Several of the implementations apply adaptive Gauss-Hermite quadratures that are known to be preferred from a numerical-accuracy point of view. The multivariate negative-binomial model with random effects offers a flexible approach to the analysis of correlated sequencing data obtained from complex (for instance, longitudinal) experimental designs. There are several software implementations that facilitate the routine use of the model in practice.

[1] G. Molenberghs, G. Verbeke, C.G.B. Demetrio, et al. A family of generalized linear models for repeated measures with normal and conjugate random effects. Stat Sci, 25, 2010, 325-347.
[2] D. Kazakiewicz, J. Claesen, K. Górczak, D. Plewczyński, T. Burzykowski. A multivariate negative- binomial model with random effects for differential gene-expression analysis of correlated mRNA sequencing data. Journal of Computational Biology, 26, 2019, 1-10.

Poster Sessions

TP5 Contributions and challenges of data fission for post-clustering differential analysis

Hivert B.*[1], Agniel D.[2], Thiébaut R.[1], Hejblum B.[1]
[1]Univ. Bordeaux, INSERM, INRIA, SISTM team, BPH, U1219, F-33000 Bordeaux, France ~ Bordeaux ~ France, [2]RAND Corporation, Santa Monica, CA 90401, USA ~ Santa Monica ~ United States of America

Gene expression data analysis is often organized around two successive steps: first a clustering using all genes to construct homogeneous and separate subgroups of observations, followed by a differential analysis step to identify the genes that are differentially expressed between those estimated clusters. This violates the traditional framework for statistical inference where hypothesis testing must be fixed before the data analysis. Good properties of tests (e.g., control of the type I error rate) are no longer guaranteed. Recently, a new approach has been proposed to address broader problems of selective inference: the data fission [1]. By decomposing the information contained in each single observation into two parts, it allows each of the two analysis steps to be applied on independent datasets, ensuring all good statistical properties of traditional tests. This approach relies on a compromise of information from the original data which is kept in each of the two parts that is tuned by a single hyper-parameter τ. We propose to use data fission for post-clustering inference of Gaussian data. We performed numerical simulations to evaluate the performance of data fission combined with dearseq [2], a variance component score test for transcriptomics analysis in the case of post- clustering inference. Since the quality of the clustering also impacts the power of the test, we showed that it was more judicious to retain the majority of the information for the clustering step, with both analytical and practical results. We applied this approach to real log2-cpm normalized RNA-seq data on 54 patients from a COVID19 study to identify gene expression discriminating between clusters linked to COVID severity.
The compromise of information required by data fission to generate the two new datasets used in the analysis impacts both the quality of the clustering and the statistical power of the test. The selection of this parameter is thus crucial for the practical implementation of the approach on real data and its performance.

[1] Leiner, J., Duan, B., Wasserman, L., & Ramdas, A. (2022). Data fission: splitting a single data point. arXiv preprint arXiv:2112.11079.
[2] Gauthier, M., Agniel, D., Thiébaut, R., & Hejblum, B. P. (2020). dearseq: a variance component score test for RNA-seq differential analysis that effectively controls the false discovery rate. NAR genomics and bioinformatics, 2(4), lqaa093.

TP6 Comparison of standard linear and ridge regression models used in the shape index calculation

Katina S.*[1], Šindlár V.[2]
[1]Department of Mathematics and Statistics, Masaryk University and Institute of Computer Science of the Czech Academy of Sciences ~ Brno, Prague ~ Czech Republic, [2]Department of Mathematics and Statistics, Masaryk University ~ Brno ~ Czech Republic

A human face, captured by stereo-photogrammetry, is characterised by about 100,000 points [1]. To perform further statistical analysis of the surface of the human face, based on the automatic identification of several anatomical and geodesic curves, we need to calculate the shape index (SI), a measure of local surface topology, which is done by using several different standard linear statistical models (LM) and ridge regression models (RM) of local z on x and y coordinates [2]. Our goal is to compare the initial results given by LM and RM and smoothed results produced by additional smoothing of SI. LM and RM are applied on a sufficiently large neighbourhood of all surface points. In both cases, the models of different order were used. The estimates of regression coefficients related to the quadratic terms and their interaction are elements of the Weingarten matrix from which the principal curvatures, and then also SI, are calculated. The goodness of fit is measured by (adjusted) coefficients of determination. Further smoothing of SI is carried out by a penalised regression model. In the case of the initial (non-smoothed) models, the models of lower order give us better results regarding the smoothness of SI, despite an unsatisfactory fit in problematic areas on the human face (hair, brows, ears, wrinkles, presence of facial palsy symptoms). Moreover, the additional smoothing improve the spatial distribution of SI in higher-order models. Once SI is precisely determined, the curves can be automatically identified and used in the subsequent multivariate statistical analyses.

[1] L. Vittert, A.W. Bowman, S. Katina. A hierarchical curve-based approach to the analysis of manifold data. The Annals of Applied Statistics 13, 4, 2019, 2539–2563.
K[2] atina S, et al. The definitions of three-dimensional landmarks on the human face: an interdisciplinary view. Journal of Anatomy 228, 3, 2016, 355–365.

TP7 A statistical methodology to select and combine covariates for disease classification in high-dimensional data

Salaroli C.J.*, Pardo M.D.C.
University Complutense of Madrid ~ Madrid ~ Spain

The Youden index is a measure of the performance of diagnostic markers that also identifies the optimal cut-off point. This measure has also been proposed in disease classification as the objective function to be maximized to combine biomarkers, even in high-dimensional contexts [1], i.e. with thousands of features and only dozens of observations. Once the single or a combination of multiple relevant biomarkers has been identified, a step forward to further improve disease diagnosis is to consider covariates, i.e. specific patient information like physical attributes, working and social habits, alternative "-omics" data - also high-dimensional -, and so on, often correlated with the recognized biomarkers. In case these additional dimensions are not known to be associated with the phenomenon, not only combining but also selecting features at the same time is a task that can significantly improve classification performance, considering only the relevant regressors and excluding the variables that cause noise.
We introduce a new method which, given the combination of biomarkers, estimates a patient-specific optimal cut-off point as a function of the covariates information, in case of both low and high dimensions, optimizing the penalized Youden index. In other words, our proposal adjusts, for each patient, the threshold from which he is considered a case, looking for the best selection and combination of covariates. The first results of the new method we propose show encouraging performance in selecting and combining covariates, improving disease classification.

[1] C. J. Salaroli, M. C. Pardo, PYE: A Penalized Youden Index Estimator for selecting and combining biomarkers in high-dimensional data, Chemometrics and Intelligent Laboratory Systems, Volume 236, 2023, 104786.

TP8 Reduced visits at emergency departments for cardiac conditions and cardiac mortality in the covid-19 pandemic

Katsoulis M.*, Gomes M., Lai A.
UCL ~ London ~ United Kingdom

In this study, we wanted to highlight some of the indirect effects of the COVID-19 pandemic by estimating the effect of reduced ED visits on cardiac mortality in England. To explore how the reduction of attendances at emergency departments for suspected cardiac disease affected cardiac mortality, we utilized an instrumented difference-in-differences design[1,2]. We used the COVID-19 pandemic as the instrument (selected date: March 12, when the UK Chief Medical Officers raised the UK risk from COVID-19 to high). We estimated the relationship between daily ED visits and cardiac deaths by using the 2-stage least squares method. First, we regressed the exposure (ED visits for cardiac diseases) on the instrument (COVID-19 pandemic), assigning the value of 0 if the ED visit occurred before March 12, 2020, or 1 if on or after this date. Second, we regressed the outcome (number of cardiac deaths) on the predicted exposure. In both steps, the following 4 terms were included to adjust for seasonality: (1) period (0 for previous years, 1 for this year [December 18, 2019 to April 15, 2020]), (2) t (time in days from 12/18), (3) squared t, and (4) cubic t. The mortality associated with untreated acute cardiac disease may occur immediately or after a delay, or lag period, the length of which is determined by the type and severity of the presenting disease. We, therefore, estimated the effects of reduced ED visits on cardiac deaths for a range of time lag periods (delays), between nonpresentation and the associated mortality, from 0 to 20 days. We estimated that every 100 non-attendances at EDs for suspected cardiac disease were associated with between 3.1 (95% CI, 1.5–4.6) and 8.4 (95% CI, 7.0–9.8) excess cardiac deaths. We found evidence of reduced ED attendances of patients with suspected cardiac disease during the COVID19 pandemic peak in England and an associated time-lagged increase in cardiac mortality. An increase in weekly non-COVID-19 cardiac mortality of up to 18%, compared with the previous 5 years was observed, and implies that one cardiac death could have been prevented or delayed for every 12 ED visits with suspected cardiac disease

[1] Katsoulis M, Gomes M, Lai AG, et al. Circ Cardiovasc Qual Outcomes. 2021;14(1):e007085.
[2] Ye T, Ertefaie A, Flory J, et al. Instrumented difference-in-differences. Biometrics 2022.

Poster Sessions

TP9 Applying mnlfa to examine consistency in dsm-5 opioid use disorder between pain patients and people who inject

Bruno R.*[1], Peacock A.[2], Campbell G.[3], Larance B.[4], Lintzeris N.[5], Nielsen S.[6], Hall W.[3], Cohen M.[2], Degenhardt L.[2]
[1]University of Tasmania ~ Tasmania ~ Australia, [2]University of New South Wales ~ Sydney ~ Australia, [3]University of Queensland ~ Brisbane ~ Australia, [4]University of Wollongong ~ Wollongong ~ Australia, [5]South Eastern Sydney Local Health District ~ Sydney ~ Australia, [6]Monash University ~ Melbourne ~ Australia

There has been little work examining the syndrome structure of DSM-5 opioid use disorder among consumers of pharmaceutical opioids, and that which has been conducted has yielded conflicting information about which symptoms are the most and least severe. With some controversy, DSM-5 use disorder criteria restrict the presence of physiological signs to situations where individuals are using opioids outside of the boundaries specified by their prescriber. We aimed to examine the consistency of DSM-5 opioid use disorder among chronic pain patients and people who inject pharmaceutical opioids. DSM-5 pharmaceutical opioid use disorder was assessed using the Composite International Diagnostic Interview in two cohorts: 1422 people prescribed strong pharmaceutical opioids for chronic non-cancer pain and 606 people tampering with pharmaceutical opioids by injecting. Moderated non- linear factor analysis (MNLFA) was used to simultaneously test whether opioid use disorder items were invariant across cohort type. Rates of DSM-5 opioid use disorder varied greatly across samples: 20.8% in the pain cohort; 96.5% in the injecting sample. MNLFA demonstrated differential functioning at the item level for physiological symptoms (more readily identified among injectors) and desire to reduce use (more readily identified in pain patients). However, there was no meaningful cohort moderation of factor loadings, supporting the consistency of the latent structure of the opioid use disorder construct. MNLFA offered advantages over the more piecemeal traditional approach to identification of differential item functioning. While the use of restrictions on physiological symptoms does not adversely affect the unifactorial nature of the syndrome in patients prescribed opioids, the restriction on DSM-5 criteria effectively requires use disorder to be more severe before these are regarded as symptoms. However, chronic pain patients appear to interpret symptom items relating to desire to reduce use fundamentally differently to injectors, likely intertwined with their experience of their pain condition.

TP10 Dyadic approaches for patient self-care and caregiver contribution to self- care assessment in type 2 diabetes

Fabrizi D.*, Ausili D., Rebora P.
University of Milano Bicocca ~ Monza ~ Italy

Patient self-care and caregiver contribution to self-care in chronic illnesses should be considered together as a dyadic phenomenon called “dyadic engagement in illness care”. In Type 2 Diabetes Mellitus (T2DM), there is a lack of studies using a dyadic approach. The possibility of classifying dyadic engagement in T2DM care may uncover patterns of behavior useful for improving T2DM management. Mixed effects models (MM) have been used to obtain dyadic scores to be used as input of latent class analysis (LCA)[1]. However, the advantages of this approach over simpler synthetic dyadic measures are not clear. This study aimed at identifying distinct patterns of dyadic engagement in T2DM care comparing two methods of dyadic data analysis. This cross-sectional study involved 251 patients with T2DM and their caregivers. Patient self-care and caregiver contribution to self-care were measured respectively by the Self-Care of Diabetes Inventory and the Caregiver Contribution to Self-Care in Diabetes Inventory, each one consisting in three scales scored 0-100. To assess the dyadic engagement, as first approach we adopted MM with random intercept and slope, obtaining the average of dyadic engagement, and the incongruence in dyadic engagement within the dyad[2]. Then, we used the MM coefficients to perform a LCA able to identify patterns of dyadic engagement. As alternative approach, we estimated dyadic average and incongruence by the raw mean and by the difference between patient and caregiver scores. Then, we used them as input to perform the LCA. Interestingly, the LCA clustered the same dyads in the same classes in both approaches, with identical fit indices. The model with three classes showed the best performances both in terms of fit and clinical characterization of the dyads. Consequently, the class trends were similar in the two approaches. MM accounts for the interdependence within the dyad and for the measurement error, returning predicted measures shrunk towards the overall mean. However it yielded the same clusters of the simpler approach using the raw data[8]. The results of this study highlight the need to further explore dyadic data analysis.

[1] C. Lee, E. Vellone, K. Lyons et al., Int. J. Nurs. Stud., vol. 52(2), 2015, pp. 588–597.
[2] K. Lyons, C. Lee, Eur. J. Cardiovasc. Nurs., vol. 19(2), 2020, pp. 178–184.

TP11 Latent markov model for profiling heart failure patients’ adherence to drugs

Savaré L.*, Ieva F., Fontana N.
Politecnico di Milano ~ Milano ~ Italy

Heart failure (HF) is a complex clinical syndrome associated with increased healthcare costs and a high burden of mortality and morbidity [1]. Despite this, no attempts have been made to jointly explore the association between clinical outcomes and adherence to the polytherapy administered to these patients. To explore an individual trait of interest that is not readily apparent, we focus on the pharmaceutical path followed by HF patients using Latent Markov models. This study also seeks to detect and model patients' latent state of polytherapy adherence over time. By considering changes in behaviour about dispensed drugs, this technique enables us to profile patients based on a range of levels of adherence throughout time. We define our latent Markov model using multivariate response variables that represent three levels of adherence to the three drugs commonly administered to HF patients each month throughout the first year of observation. Age, gender, and the multisource comorbidity score are examples of fixed and time-varying covariates that will affect the initial and transition probabilities of the latent process. We fit different LMM with a different number of latent states, and the final model is chosen according to different measures and its interpretability, such as BIC and AIC, in our case with three latent states [2]. Looking at the estimated conditional response probabilities, we can interpret the three latent states as three levels of adherence to polytherapy. Once the model has been estimated, a decoding approach is used to produce a path prediction of the latent state for each subject. We establish eight separate "latent profiles" based on this progression of latent states, which describe the patients' latent states over time and, subsequently, the patients' adherence over time. Finally, we analyse through prognostic models the association between the different adherence profiles over time and the probability of survival and hospital readmission. We observe that different adherence behaviours lead to a different probability of survival and hospital readmission. Our approach yields a useful tool to profile heart failure patients’ adherence to the overall drug treatment planned for these patients.

[1] PS. Jhund, K. Macintyre, CR. Simpson, JD. Lewsey, S. Stewart, A. Redpath, et al. Long-term trends in first hospitalization for heart failure and subsequent survival between 1986 and 2003: a population study of 5.1 million people. Circulation 2009; 119:515–23. https://doi.org/10.1161/CIRCULATIONAHA.108.812172
[2] F. Bartolucci, G.E. Montanari, and S. Pandolfi. Three-step estimation of latent Markov models with covariates. Computational Statistics & Data Analysis, 83:287–301, 2015.

TP12 Challenges in group sequential designs in rare diseases

Bodden D.*[1], Heussen N.[2], Hilgers R.[1]
[1]RWTH Aachen University ~ Aachen ~ Germany, [2]Sigmund Freud Private University ~ Vienna ~ Austria

Group sequential designs are promising in rare diseases, as they can reduce the average needed sample size. Due to the interim analyses, the significance level per look has to be adjusted to yield an overall significance level ⍺ at the end of the study. This is usually done under the assumption that the control and experimental group are stage wise balanced. However, only a few special randomization procedures guarantee this. Even when a randomization procedure with stage wise balancing is used, the withdrawal of patients or invalid data can still result in an unbalanced allocation ratio which can lead to an inflated type-I-error. This is especially relevant in small sample sizes, as a slight deviation from a balanced allocation is more impactful. The objective is to measure the impact on the type-I- error for different randomization procedures that violate the terminal balance assumption. Furthermore, to propose a solution to strictly control the type-I-error. We will look at a two-sided z-test for comparing two treatments with normal responses of known variance. To calculate the critical values, we will use an error spending approach with alpha-spending functions suggested by Lan & DeMets [1]. We simulate randomization sequences from different randomization procedures to evaluate the type-I-error. We also propose a solution for the inflated type- I-error by recalculating the critical values according to the updated information from the allocation ratio after each stage. Our findings underscore the importance of carefully selecting a randomization procedure in group sequential designs. It is common practice, that the critical values are being updated when the actual timing of an interim analysis deviates from the pre-defined timing. We propose that this should also be done for deviations from the expected allocation ratio to prevent an inflation of the type-I-error.

[1] Gordon Lan, K. K., and David L. DeMets. "Discrete sequential boundaries for clinical trials." Biometrika 70.3 (1983): 659-663.

TP13 Multiple regression models for predicting rare outcomes in hypertrophic cardiomyopathy

Žebrauskiene D.[1], Puronaite R.[4], Sadauskiene E.[2], Masiuliene R.[3], Preikšaitiene E.[1], Jakaitiene A.*[5]
[1]Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [2]Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [3]Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania, [4]Center of Informatics and Development, Vilnius University Hospital Santaros Klinikos, Clinic of Cardiac and Vascular Diseases, Institute of Clinical Medicine, Faculty of Medicine, Institute of Data Science and Digital Technologies, Faculty of Mathematics, [5]Institute of Data Science and Digital Technologies, Faculty of Mathematics and Informatics, Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University ~ Vilnius ~ Lithuania

Contemporary treatment of hypertrophic cardiomyopathy (HCM) has significantly reduced HCM- related mortality over the last 60 years. However, the outcomes associated with sudden death in HCM patients significantly impair quality of life and current risk scores still lack sensitivity [1]. The purpose of our research is to develop prognostic models of rare outcomes related to sudden cardiac death in HCM patients examined and treated at the Vilnius University Hospital Santaros klinikos (VUHSK). Clinical data from paediatric and adult HCM patients treated at the VUHSK between 2005 and 2022 were analysed. Patients or their legal representatives gave consent to participate in the study approved by the Academic Ethics Commission of the Faculty of Medicine of Vilnius University (no. 2020/1-1182-669). For analysis, we have selected six outcomes related to sudden cardiac death (4 binary and 2 continuous). For binary outcome we set up binary multiple logistic and for continuous – multiple regression models. As explanatory factors we have 126 phenotypical variables and genotype data for genetic variants associated with HCM in our database. The large share of independent phenotype and genetic variables are factor variables that enter the regression model as k-1 dummy variables (k number of layers in the factor variable). This might double the number of independent variables, which might exceed the sample size. Furthermore, our outcome variable is rare, that is, 2- 12% of the sample size. We apply various strategies for variable selection (univariate analysis, step- wise selection, LASSO and Ridge regressions, and other). We obtain that the most conservative selection of explanatory variables for most outcomes is LASSO, and the Ridge regression remains the most variable. We calculate confidence intervals for the estimated parameters using the bootstrapping technique. To build prediction models when the number of predictors exceeds the sample size, the problem of variable selection must first be solved. For variable selection to predict rare outcomes related to sudden cardiac death in patients with HCM, it is suggested to combine several strategies in order to achieve the high precision of the model.

[1] Maron BJ, Desai MY, Nishimura RA, et al., Management of Hypertrophic Cardiomyopathy: JACC State-of-the-Art Review. J Am Coll Cardiol. 2022 Feb 1;79(4):390-414.

Poster Sessions

TP14 A review of the design of clinical trials in rare cancers

Rane P.*, Kannan S.
The Advanced Centre for Treatment, Research and Education in Cancer (ACTREC) ~ Navi Mumbai ~ India

Clinical research for patients with rare cancers has been very challenging because of a lack of clinical expertise, good study design requires large numbers of patients, and recruitment goals are rarely feasible. Our objective is to carry out a systematic review to find out different approaches to clinical trial designs that have been developed that can be used to study new treatments for patients with rare cancers. We performed a systematic review of the literature on rare cancer clinical trials from 2000-2022 in PUBMED using the search strategy clinical trial and ("rare cancer"). Clinical trial articles on humans were included. Review and the methodological articles were excluded. Study characteristics and other clinical and design-related data like study type, cancer type, randomization, sample size, type of design used, the primary endpoint, the objective for using a particular design, and the result of the attainment of the objective were extracted from the selected articles. The search retrieved 78 results from oncology journals in the PubMed database of which 25 were eligible for analysis. Response rates were more commonly used as the primary endpoint (64%). The majority of trials had Phase II studies (72%) which 22% were randomized studies and 33% were terminated because of slow accrual or safety concerns. The average sample size in phase II studies were 33 patients (range 8-73). The most commonly used design in phase II was Simon's two-stage design (50%), Not providing information about (38%), only one study used the Bayesian framework for a two- stage design, and one study used a double endpoint study design. The major reason for choosing the Bayesian framework was focus on estimation. The adaptive decision rule for the efficacy conclusion without fixing the sample size was easily obtained using the Bayesian design. Most of the trials has been used Simon's two-stage designs although Bayesian design can also be one of the alternative design for adaptive decision rule. Our review showed that improvement is still needed while designing the clinical trials in rare diseases specifically in oncology. Develop standard guidelines for designing clinical trials in rare diseases is recommended.

Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. Biometrics. 1994 Jun;50(2):337-49. PMID: 7980801.

TP15 Clinical trial design and estimation methods that can yield additional information than single-arm trials

Sato T.*, Hida E.
Osaka University ~ Suita ~ Japan

Conventional randomized controlled trial designs are difficult to implement in small populations, such as in rare disease and pediatric disease areas. Various methodological and statistical considerations have been reported for such small clinical trials [1, 2]. Many single-arm trials of only a test treatment are still often conducted due to feasibility, allowing within-patient comparisons to be assessed. In single-arm trials, the efficacy of a test drug is evaluated based on a pre-specified threshold. However, it is well known that even if a treatment effect is better than the threshold in a well-controlled single- arm trial, the estimate of the treatment effect is subject to bias. Therefore, simple estimators from single-arm studies may make it difficult to draw valid conclusions about efficacy. Under these circumstances, it is also desirable to be able to estimate the true effect size of the test drug without being affected by bias. In this study, we propose a clinical trial design and treatment effect estimation method that can be used in SCTs. We propose a new method of estimating treatment effects using a delayed-start design. In the delayed-start design, a randomized controlled trial is conducted in the first period and a single-arm trial is conducted in the second period, allowing for the estimation of the true treatment effect. Although various factors, such as disease and treatment characteristics, determine the estimand and change the modelling, we have given model-specific estimation methods and interpretations. We show that the appropriate use of delayed-start design enables the estimation of the true treatment effect in addition to the evaluation of efficacy by comparison with a pre-determined threshold as in single-arm trials. By conducting numerical study, we evaluate performance and give model-specific interpretations. Delayed-start designs with modelling appropriate for the main estimand may be more effective for practical clinical trials in the rare disease and pediatric disease areas than single-arm trials.

[1] IOM. Small clinical trials. issues and challenges (2001).
[2] CHMP. Guideline on clinical trials in small populations (2006).

TP16 The impact of allocation bias on the test decisions in clinical trials with multiple endpoints

Schoenen S.*[1], Hilgers R.[1], Heussen N.[2]
[1]RWTH Aachen University ~ Aachen ~ Germany, [2]Medical School-Sigmund Freud Private University ~ Vienna ~ Austria

Context: Regulatory guidelines advise researchers to focus on a single primary endpoint in clinical trials. However, in some scenarios, consideration of multiple endpoints can provide a more comprehensive understanding of treatment effects and may lead to increased power or reduced sample size, respectively. This is especially relevant in rare diseases where only a limited number of patients are available. To ensure correct test decisions in trials with multiple endpoints, special testing procedures such as the Bonferroni adjustment or the All-or-None decision rule are commonly employed. However, in two-arm parallel group trials biased test decisions can also be introduced by the allocation process.
Objectives: We aim to investigate the effect of allocation bias on test decisions for different randomization procedures (RPs) in the planning phase of a clinical trial. Therefore, we introduce a model for two-arm parallel group trials with multiple continuous endpoints, that assesses the impact of allocation bias on test decisions when the Bonferroni adjustment or the All-or-None decision rule is applied.
Method: We generalize Proschan's biasing policy [1] to continuous, normally distributed multiple endpoints so that endpoint-specific bias effects on the test decisions of the Bonferroni or All-or-None procedure can be quantified. Based on this model, we derive a formula to compute the biased type-I-error rate for the All-or-None procedure and the biased family-wise error rate for the Bonferroni procedure conditional on a randomization sequence. Then, the impact of allocation bias on the test decision is analyzed by comparing these errors for different clinical settings and RPs in a simulation study.
Results: We show that, for the Bonferroni procedure as well as for the All-or-None procedure, allocation bias leads to inflation of the mean family-wise error and mean type-I-error, respectively. The strength of this inflation is affected by the choice of the RP, whereas the number of endpoints or small sample sizes, demonstrate no influence.
Designing a valid and meaningful clinical trial with multiple endpoints requires considering the impact of allocation bias on the test decision during the planning phase. Therefore, the presented model can be used to provide the mandatory choice of an appropriate RP.

[1] Proschan M. Influence of selection bias on type I error rate under random permuted block designs. Stat Sin 1994; 4: 219–231.

Poster Sessions

TP17 Addressing biases in odds ratios for rare chronic skin conditions using real-world prescription drug exposures

Swiderski M.*[1], Vinogradova Y.[1], Knaggs R.[1], Harman K.[2], Harwood R.[1], Prasad V.[3], Persson M.S.[4], Figueredo G.[1], Layfield C.[1], Gran S.[1]
[1]University of Nottingham ~ Nottingham ~ United Kingdom, [2]University Hospitals of Leicester NHS Trust ~ Leicester ~ United Kingdom, [3]University of Nottingham, King's College London ~ Nottingham, London ~ United Kingdom, [4]Swedish Rheumatism Association ~ Stockholm ~ Sweden

Bullous pemphigoid (BP) is an autoimmune skin disease presenting with itching and blisters of the skin. It is a rare condition whose cause is unknown, with increasing incidence and high mortality in older people[1]. Antibiotic use could be one of the potential risk factors for developing BP[2]. Because BP's early symptoms resemble skin infections and may result in extra antibiotic prescribing, the association between BP risk and antibiotics needs careful investigation by performing clinically relevant and informed sensitivity analyses. We conducted a population-based, nested case-control study using the UK Clinical Practice Research Datalink. Patients diagnosed with BP (cases) between 1998-2021 were matched to up to five controls by birth year, sex, and general practice. Multivariable conditional logistic regression model adjusted for ethnicity, comorbidities and other medications commonly prescribed for older people estimated the risk of developing BP associated with at least one prescribed penicillinase-resistant penicillin (antibiotics for treating skin infections) within a year before diagnosis (0). Sensitivity analyses were conducted to account for: (i) longer drug exposure by changing the definition of exposure from one to two years before BP diagnosis; (ii) for treating undiagnosed BP symptoms (protopathic bias) by excluding six months of prescription data before diagnosis ; (iii) latest skin infection within six months before diagnosis which could initiate antibiotics prescriptions; (iv) health-seeking behaviour by controlling for the number of consultations in the multivariable model. The main analysis demonstrated a 7-times increased BP risk associated with penicillinase-resistant penicillin prescribing (0; AOR: 7.28, 95%-CI: 6.88-7.70), which was higher than the sensitivity analyses. The most considerable difference was for excluding prescriptions in the last six months before BP diagnosis (ii; 2.57, 2.39-2.78) followed by including prescriptions in the last 2 years (i; 5.86, 5.56 - 6.18), adjusting for the latest skin infection (iii; 6.34, 5.98-7.61), and adjusting for the number of consultations (iv; 6.37, 6.02-6.75). The results indicate an association between penicillinase-resistant penicllins and BP. Sensitivity analyses reveal the size of the association changes considerably after taking into account diagnosis delay and/or misdiagnosis of BP. Clinically-informed sensitivity analyses are necessary when using real-world data.

[1]Persson, M. S. M., et al., Br J Dermatol, 184, 2021, 68-77 [2]Verheyden, M. J., et al., Acta Derm Venereol, 100, 2020, adv00224

TP18 Global rank test with data-driven optimal weights in clinical trials with multiple endpoints

Yoshida S.*[1], Yamaguchi Y.[2], Maruo K.[3], Gosho M.[3]
[1]Astellas Pharma Inc. ~ Tokyo ~ Japan, [2]Astellas Pharma Global Development Inc. ~ Chicago ~ United States of America, [3]University of Tsukuba ~ Tsukuba ~ Japan

Multiple efficacy endpoints are generally assessed in clinical trials and the ICH E-9 guidance recommends selecting a single endpoint as a primary endpoint of the study. However, there are some cases where more than one single primary endpoint is desirable to figure out the characteristics of the new drug and some clinical studies for rare diseases assess multiple endpoints as primary endpoints due to the heterogeneity of symptoms or lack of disease information. The global test is often used in these studies since it does not require multiplicity adjustment and could provide a higher statistical power than the Bonferroni test in a particular setting. The global test uses the weighted sum of statistics for each endpoint for the statistical comparison, so the determination of weights is a key element of this test. Ramchandani et al. [1] proposed to divide the patients in the study into multiple strata and calculate the weights for the next stratum based on the data from the previous stratum and combine the statistics from each stratum for the statistical comparison. We extended the method proposed by Ramchandani [1] to use the entire current study data to determine the weights at the analysis stage. To prevent alpha error inflation, our method is based on the permutation test. The simulation studies were conducted to investigate the performance of the proposed method and compare it with other global tests. The simulation studies show that our proposed method can control the type I error rate under the nominal significance level regardless of the number of primary endpoints and the correlation among the endpoints, and provide higher statistical powers compared with other global tests in several scenarios.
Simulation studies suggest that the proposed method could be an alternative approach to determine the weights to maximize the statistical power in case it is difficult to pre-determine the weights based on the available information before the study initiation. Further investigation would be required using other distributions of endpoints.

[1] Ramchandani R, Schoenfeld DA, Finkelstein DM. Global rank tests for multiple, possibly censored, outcomes. Biometrics. 2016;72:926–935.

TP19 The non parametric combination methodology to analyze influence of food regimes on oxidative stress parameters

Alibrandi A.*, Zirilli A., Campennì A., Ruggeri R.M., Cannavò S.
University ~ Messina ~ Italy

Statistical methodology is an useful and powerful tool in the clinical scientific research and, in this background, permutation tests' applications increased in recent years to solve complex multivariate problems, by virtue of their broad flexibility and adaptability in numerous research fields. The main aim of this paper is to assess the existence of significant differences between two dietary regimes (omnivorous vs vegetarian) with reference to some markers of oxidative stress: SOD, GPX, TRX, GR, ABTS, AGEs, AOPPs, by using the Non Parametric Combination (NPC) methodology, based on permutation test [1]. Two hundred subjects were enrolled by Endocrinology Unit of Messina University Hospital and invited to compile a questionnaire about their dietary habits. None were under any pharmacological treatment [2]. By means of NPC test, all comparisons were performed stratifying for gender (male vs female), age class (≤ 40 vs > 40 years), BMI class (≤25 vs > 25), FT4 (normal vs altered), Hashimoto's Thyroiditis (yes or no) and physical activity (practiced or no). We identified parameters of oxidative stress that discriminate the two examined dietary regimes, omnivorous vs vegetarian. The GPX parameter is significantly lower in subjects who follow an omnivorous diet than in vegetarians, in particular in the female stratum, in both age groups, in subjects with normal weight, not affected by Hashimoto's thyroiditis and in both groups defined in function of physical activity. Also, the TRX parameter is significantly higher in vegetarian subjects than in omnivores; more specifically, we find this result in women, in both age groups, in subjects with BMI less than 25, in both FT4 classes, in subjects not affected by Hashimoto's thyroiditis and in both categories defined by physical activity. Furthermore, NPC test shows that ABTS is significantly higher in omnivores over 40 years of age. Finally, the AGE marker is significantly lower in vegetarians with normal FT4 levels. Thanks to the NPC methodology we can state that food styles exert a significant influence on some oxidative stress parameters.

[1] F.Pesarin, Multivariate Permutation tests: with application in "Biostatistics", John Wiley & Sons, 2001, Chichester, UK.
[2] R.M., Ruggeri et al., Influence of Dietary Habits on Oxidative Stress Markers in Hashimoto's Thyroiditis, Thyroid, 2020, 31(1): 96-105.

TP20 Exploring the multidimensional benefits of a new treatment with generalized pairwise comparisons

Chiem J., Barre E.*, Kosta S., Saad E., Salvaggio S., De Backer M., Deltuvaite--Thomas V., Buyse M.
International Drug Development Institute (IDDI) ~ Louvain-la-Neuve ~ Belgium

A randomized clinical trial was conducted to test the effect of an investigational drug on the incidence, duration and time of onset of severe adverse events (AEs) for patients receiving radiotherapy for solid non-resectable tumors. We aimed at quantifying the overall benefit of the investigational drug on all these outcomes in a single analysis. Method: The Generalized Pairwise Comparisons (GPC) method allows the simultaneous evaluation of several prioritized outcome measures.[1] The method consists of performing all possible pairwise comparisons between one patient from the investigational arm and another patient from the control arm. Each pair is classified as a "win" or a "loss" depending on which of the two patients does better, using the prioritized outcomes. The Net Treatment Benefit (NTB) is the difference between the proportions of wins and losses. The NTB estimates the net probability that a random patient from the experimental arm would have a better outcome than a random patient from the control arm. In this trial, the prioritized outcomes were, in decreasing order of clinical importance: (1) incidence of AEs of WHO grade 4, (2) incidence of AEs of WHO grade 3, (3) total number of days with AEs, (4) time to onset of first AE. Results: This GPC analysis was conducted on 407 patients (Investigational drug arm = 241, Placebo control arm = 166). The overall probability of a better outcome was 56.2% in the investigational group vs. 33.1% in the control group, resulting in NTB = 23.1% (P=0.00019). The additive contributions of the prioritized outcomes on NTB were 12.6% for incidence of grade 4 AEs, 4.2% for incidence of grade 3 AEs, 3.2% for total number of days with AEs and 3.1% for time to onset of first AE. The analysis using GPC is more statistically sensitive as well as more clinically relevant to detect patient-relevant treatment benefits than a simple comparison of incidences between the treatment arms. The NTB is a natural measure of treatment effect that can be used regardless of the nature of the outcomes considered (binary, categorical, continuous or time to event).

[1] Buyse M (2010). Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med, 29:3245-3257.

TP21 Nonparametric comparison of two proportions: le cam's theorem and chernoff bound applied to upper-lower index

Stepanek L.*[1], Habarta F.[2], Mala I.[2], Marek L.[2]
[1]Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business & Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University ~ Prague ~ Czech Republic, [2]Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business ~ Prague ~ Czech Republic

There are well-established methods, e.g., z-test for two proportions or nonparametric permutation alternatives, when comparing two proportions of a dichotomous event in two populations, given that each individual in each population has an identical probability of success and counts of successes follow binomial distributions. However, if individuals of one population do not have an identical probability of success, then the counts of successes in each population follow the Poisson-binomial distribution. Grand total averaging of all non-identical probabilities of success across a population and application of traditional methods for two populations' comparison would bias results, though. Although Poisson-binomial distribution modeling reflects the situation adequately, calculations behind it could be computationally greedy. In this work, we address the introduced issues and apply a proposed methodology on an upper-lower index, which is used for quick comparison of populations' proportions, e.g., probabilities of correct answers to items between two groups of students in test evaluation. The upper-lower index lacks feasible inference since it averages not necessarily identical probabilities within each population. Assuming only a dichotomous outcome, each individual's probability of success follows the Bernoulli distribution; however, the distributions are not identically distributed. Thus, concerning individual probabilities of success within each population, counts of successes follow Poisson-binomial distribution rather than a binomial. Calculations with derived expected values and variances of Poisson-binomial distribution are computationally exhaustive; thus, using Le Cam's theorem and Chernoff bound, we derived a lower bound for the difference of counts of successes in both populations. So, the real difference in proportions of successes in populations is with high probability ("almost-sure") even greater than the lower bound, which is computationally cheap. Finally, we apply the proposed technique to simulated data of test items in two groups and compare the proportions of correct answers to items, estimating the statistical properties of the introduced nonparametric approach. Le Cam's theorem and Chernoff bound enable us to computationally quickly estimate a lower bound of difference in proportions of successes in two populations. Since success rates in populations follow Poisson-binomial distribution, it is a feasible approach compared to grand total averaging and using traditional tests for proportions.

[1] Chernoff, H. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations. The Annals of Mathematical Statistics, 23, 1952, 493–507.
[2] Le Cam, L. An Approximation Theorem for the Poisson Binomial Distribution. Pacific Journal of Mathematics, 10, 1960, 1181–1197.

TP22 Identification of risk factors for wound non-healing: a descriptive study.

Aguirre Larracoechea U.*[1], Burzako Perez M.A.[2]
[1]Research Unit, Osakidetza Basque Health Service, Barrualde-Galdakao Integrated Health Organisation, Galdakao- Usansolo Hospital. Kronikgune Institute for Health Services Research. Network for Research on Chronicity, Primary Care, and Health Promotion (RICA, [2]Healthcare Management Unit, Osakidetza Basque Health Service, Barrualde-Galdakao Integrated Health Organisation, Galdakao-Usansolo Hospital. ~ Galdakao ~ Spain

Chronic injuries are a serious problem for individuals, society and the Health System itself, and have serious consequences for the quality of life, nursing activity and material resources of those who suffer from them and their environment. The objective of this study is to identify the risk of non-healing depending on the type of wound. Retrospective longitudinal study performed during the 2013-2020 period, where the information on chronic injuries received in the nursing care management programme was extracted through Bussiness Intelligence platforms. A descriptive analysis of the information collected during the period of study has been carried out: frequencies and percentages for categorical and mean and standard deviations (or interquartile range) for continuous variables. competing risk models were used in order to determine the risk and cumulative incidence of chronic wound failure by type and subtype. Statistical procedures have been implemented through SAS System v9.4 and R Studio 4.4. The statistical meaning was taken when the p-value is < 0.05. The sample of the study consists of 17828 chronic injuries associated with 8109 patients, of which 4215 (51.98%) are women. Taking pressure ulcers (PU) as a reference, vascular ulcers are 1.55 times more likely not to heal the wound (sHR (95% CI): 1.55 (1.38, 1.74), p < 0.001). Diabetic ulcers increase by 88% (sHR (95% IC): 1.88 (1.38, 2.55), p < 0.001) the likelihood of not healing the wound in relation to UPPs. Current information systems, through predictive standards, allow patients to be stratified according to different levels of risk, evolving, orienting the therapeutic process and optimizing the resources of the health services of the public network.

[1] J. Beyersmann, A. Allignol, M. Schumacher, Competing Risks and Multistate Models with R, Springer 2012.
[2] D.G. Kleinbaum, M. Klein, Survival Analysis, Springer 2005.

Poster Sessions

TP23 An extension of the numbers-needed-to-treat concept based on net benefit

Alberto A.*, John F.
School of Medicine, HRB Clinical Research Facility University of Galway. ~ Galway ~ Ireland

Net Benefit (NB) is used to estimate the effect of a new intervention by contrasting the probabilities of a favourable pair and unfavourable pair, for any pair of randomly selected individuals, one on treatment and the other one on control. In favourable pairs, the treated individual has the better outcome of the two, with the reverse being true in unfavourable pairs. The advantage of this type of summary is that it can be calculated and interpreted in a very similar way for any type of response variable including categorical, continuous and time to event outcomes. It is known that the NB is equal to the absolute risk reduction (ARR) for binary outcomes, which facilitates an alternative interpretation of the Numbers Needed to Treat (NNT), calculated as 1/ARR=1/NB, as follows: number of individuals we need to treat to see one additional improved outcome, in comparison with a similarly sized set of randomly selected controls. In this presentation we extend this novel interpretation of the NNT to continuous and survival outcomes providing examples in the context of clinical trials. We also show formulae for the estimation of NB using marginal cumulative distribution functions that considerably simplify previously published calculations. Finally, we explore a connection between the NB and the Area Under the Curve (AUC) in an ROC analysis, thereby providing a novel graphical representation for the NB. This natural extension of the calculation of the numbers needed to treat for continuous and time to event responses will facilitate the interpretation of the effect of treatments in a manner that is easily understood by clinicians for a wide variety of outcomes. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Statistics in medicine. 2010 Dec 30;29(30):3245-57.

Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. New England journal of medicine. 1988 Jun 30;318(26):1728-33.

TP24 Comparison of selection strategies to identify biostatistical methods - case study on group variable selection

Buch G.*, Wild P.S.
Preventive Cardiology and Preventive Medicine, Department of Cardiology, University Medical Center of the Johannes Gutenberg University Mainz ~ 55131 Mainz ~ Germany

Current literature emphasizes the need for neutral comparisons for objective evaluation of biostatistical methods.[1] In this context, guidelines have been developed to aid in the planning and reporting of simulation studies.[2] However, there is still no consensus on the decision path that should be used to select statistical methods for comparison. In worst case, a comparison is not fair and comprehensive, because relevant techniques have not been considered. Structured approaches prevent this and can be used to identify appropriate methods. Several such strategies were examined in this work. Four different selection strategies were considered: a systematic literature review, a systematic software review, a selective literature review and a selective software review. The results of these strategies were compared by using them to identify approaches implemented in R for selecting groups of variables associated with a given outcome. The systematic reviews were carried out by one reviewer (using the R-Package packagefinder for the software review and the PRISMA guideline for the literature review), while the results of the selective reviews were taken from existing sources. In total, the four selection strategies identified 18 unique methods implemented in R for selecting groups of variables. The systematic software review found 17 approaches, the systematic literature review identified 14, the selective literature review found eight and the selective software review discovered six. Of all the techniques identified, only five were found by all review strategies. The systematic software review identified three techniques that were not found by any other strategy, and the systematic literature-based review discovered one such method. Systematic reviews are recommended over selective review strategies, as they identified more approaches. The results of the selective reviews mainly included the approaches identified by all strategies, suggesting that a selective review only identifies the most established techniques. Since a systematic software review can be largely automated, it is more efficient than a systematic literature review. However, a software-based review is limited to implemented approaches and should be accompanied by a literature review, for example, to also identify approaches that are too simple to have their own implementation.

[1] Boulesteix, A. L., Binder, H., Abrahamowicz, M., & Sauerbrei, W. (2017). On the necessity and design of studies comparing statistical methods. Biometrical Journal. Biometrische Zeitschrift, 60(1), 216-218.
[2] Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Statistics in medicine, 38(11), 2074-2102.

TP25 Optimising sample allocation to batches in laboratory-based observational studies

Burger B.*, Teare D., Lin N., Nsengimana J.
Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

In observational laboratory-based research samples often need to be processed in batches. This can be due to the limited number of samples that can be processed within a certain time (either by hand or on instruments), due to sample multiplexing limitations, or other factors limiting sample throughput. Additionally, available samples are often not balanced on the variable(s) of interest and/or nuisance variables, e.g. disease state, treatment, sex, age. In these settings it can become challenging to divide samples over batches in an efficient way that minimises the variances of the estimators or contrasts of interest. Current algorithms[1,2] randomly generate a set of allocations, from which they choose one "optimal" allocation based on some form of generalised variance, in effect ignoring any particular research question the researcher might have. Within the least squares framework it is possible to calculate the variances of specific comparisons of interest independent of outcome data. This makes it straightforward to compare different batch allocations based on the variances of those comparisons that are relevant to the research questions. For relatively small experiments it is possible to generate all possible allocations, while in more complex settings we randomly generate batch allocations, softly aiming towards diversity within batches. Subsequently, variances of point estimates and comparisons of interest are calculated to facilitate meaningful comparisons between the generated allocations. This can also make explicit that variances of interest are not influenced by imbalances in nuisance variables if interactions are assumed to be non-significant and thus not included in the statistical model. In this project we have proposed a means for researchers to compare possible allocations based on the variances of point estimators and comparisons of interest. By providing a choice of allocations along with the corresponding variances, the trade-off between different comparisons can be judged by the end-user. Additionally, the researcher can compare an allocation of their choice against ones that have been algorithmically generated. Besides providing an easy to use graphical tool for allocating samples to batches, this can help build intuition and confidence by showing visually the impact different allocations can have on the strength of the experiment.

[1] B. Burger, M. Vaudel, H. Barsnes, Biostatistics, 2022, kxac014
[2] L. Yan, C. Ma, D. Wang, Q. Hu, M. Qin, J.M. Conroy, L.E. Sucheston, C.B. Ambrosone, C.S. Johnson, J. Wang, and S. Liu. BMC Genomics, 13, 2012, 689

Poster Sessions

TP26 Genetic and environmental determinants of drug adherence

Cordioli M.[1], Corbetta A.*[2], Jukarainen S.[1]
[1]Institute for Molecular Medicine Finland, University of Helsinki ~ Helsinki ~ Finland, [2]Health Data Science Center, Human Technopole ~ Milan ~ Italy

Patients drug-taking behaviour is one of the major factors impacting treatments' efficacy. We investigated the determinants of adherence and persistence in the FinnGen study, combining nationwide drug purchase data with genetics. By using data from Finnish health registries, thus the Finnish drug purchase registry (68,826,654 total purchases), and genetic data from FinnGen[2] (N=356,077), we provide a systematic investigation of adherence across multiple medications. We selected medications which were among the most prescribed and of clinical relevance, and whose usage would not impact well-being significantly. For each of the medications, we defined two phenotypes describing two drug purchasing behaviours: (adherence, defined as the medication possession ratio (proportion of time where the medication supply is available) and persistence, defined as purchasing the medication for at least one year vs early discontinuation (after one purchase). For each of the six drugs, we investigated the associations of adherence and persistence with the genetic predisposition (through Polygenic Scores) of 31 clinically relevant traits. We further run a GWAS of adherence and persistence to each drug. A higher genetic predisposition to participate in follow-up health questionnaires is associated with higher adherence to all six medications, similarly for educational attainment and adherence to statins and antiplatelets. Genetic liability to psychiatric traits like schizophrenia and neuroticism is correlated negatively with adherence. Predisposition for risk factors such as higher BMI and diabetes positively correlates with adherence to statins and blood pressure medications. Notably, genetically higher blood pressure is associated with higher adherence to blood pressure medications and statins, while LDL-cholesterol shows the opposite effect on both. Adherence varies across medications, with more general preventive treatment (e.g. statins and blood pressure medications) having lower median adherence and more spread distributions, and treatment for more severe conditions (breast cancer medication) resulting in higher adherence and narrowed distribution. Patterns for associations between PGSs and persistence replicate those observed for adherence. A GWAS of adherence and persistence identified no genetic variation associated with the two phenotypes (p values under the Bonferroni threshold).

[1] Simpson, Scot H et al. "A meta-analysis of the association between adherence to drug therapy and mortality." BMJ (Clinical research ed.) vol. 333,7557 (2006): 15. doi:10.1136/bmj.38875.675486.55
[2] Kurki, Mitjal., et al. "FinnGen: Unique genetic insights from combining isolated population and national health register data." medRxiv(2022)

TP27 The ICH e9(r1) estimand framework adapted in a phase III equivalence RCT conducted during COVID-19 pandemic.

Mitroiu M.*[1], Ebbers H.[2], Zhou Y.[3], Yang X.[3], Dong Q.[3], Rezk M.F.[1], Addison J.[4]
[1]Evidence Generation Biosimilars, Biogen International GmbH ~ Baar ~ Switzerland, [2]Clinical Research Biosimilars, Biogen International GmbH ~ Baar ~ Switzerland, [3]Research and Development, Bio-Thera Solutions Ltd. ~ Guangzhou ~ China, [4]Evidence Generation Biosimilars, Biogen Idec. ~ Maidenhead ~ United Kingdom

A multiregional Phase III equivalence study of proposed biosimilar BAT1806/BIIB800 vs reference tocilizumab, was conducted partially during the COVID-19 pandemic. As a consequence of COVID-19 pandemic and lockdowns, intercurrent events related or not to COVID-19 were identified. ICH E9(R1) estimand framework[1] with strategies for addressing these intercurrent events was implemented for efficacy evaluation following Regulatory recommendations[2]. Primary estimand for EMA assessed the treatment effect at Week 12 where the clinical question of interest was: "What is the treatment effect had no subject discontinued the treatment, nor missed a study treatment infusion, for any reason, had rescue medication not been available (hypothetical), and death considered a non-response (composite)?". Primary estimand for FDA/NMPA assessed the treatment effect at Week 24 where the clinical question of interest was: "What is the treatment effect regardless of any treatment discontinuation or missed study treatment infusion, regardless of any rescue medication need within protocolled window (treatment policy), and death considered a non-response (composite)?". Secondary estimands for Regulatory Agencies assessed the treatment effect at Week 12/Week 24 where the clinical question of interest was: "What is the treatment effect regardless of any treatment discontinuation or missed study treatment infusion not related to COVID-19 (treatment policy), and had no subject discontinued treatment or missed a study treatment infusion related to COVID-19 (hypothetical), and regardless of rescue medication need within protocolled window (treatment policy), and death considered a non- response (composite)?". Faced with COVID-19, E9(R1) was helpful to define treatment effects using the estimand attributes, especially the strategies for addressing intercurrent events. The implementation of the estimand framework resulted in three different estimand constructs, primary or secondary, depending on recommendations from each Regulatory Agency. The ICH E9(R1) estimand framework was implemented for a pivotal study following guidance and methodological considerations for trials possibly affected by COVID-19. Regulatory Agencies prioritised different estimands as primary, perhaps leading to giving different weight to outcomes more relevant to clinical practice, as opposed to sensitivity to detect differences between candidate biosimilar and reference in this equivalence study.

[1] ICH E9(R1) EWG. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1) [Internet]. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
[2] EMA. Implications of coronavirus disease (COVID-19) on methodological aspects ongoing clinical trials - Scientific guideline [Internet]. European Medicines Agency. 2020 [cited 2023 Feb 22]. Available from: https://www.ema.europa.eu/en/implications-coronavirus-disease-covid-19-methodological-aspects-ongoing-clinical-trials-scientific

Poster Sessions

TP28 Adherence of spirit guidelines in no-profit clinical trials

Gambini G.*[1], De Silvestri A.[1], Ferretti V.V.[1], Musella V.[1], Scotti V.[2], Klersy C.[1]
[1]SSD Biostatistica e Clinical Trial Center, Fondazione IRCCS Policlinico San Matteo ~ Pavia ~ Italy, [2]SSD Servizio di Documentazione Scientifica, Fondazione IRCCS Policlinico San Matteo ~ Pavia ~ Italy

The SPIRIT 2013 Statement provides evidence-based recommendations for the minimum content of a clinical trial protocol. SPIRIT is widely endorsed as an international standard [1]. It consists of a checklist of 33 items and a figure, and is accompanied by an explanation and elaboration document containing relevant details for each item. Following such guidelines ensures clarity, quality and feasibility of clinical trials [2]. After 10 years from its publication, many are still unaware of its usefulness when preparing protocols of clinical trials to be submitted to the Ethical Board (EB). We aimed at analysing such protocols to assess the rate of adherence to the SPIRIT guidelines and its potential correlates. We retrieved information on design and methodology of no-profit protocol submitted in years 2021- 2022 to the local EB. We used the SPIRIT checklist to identify the proportion of items that were satisfied for each study. We report the results as the median with interquartile range (IQR). We plan to use Poisson regression to model correlates of non adherence. We use the Stata software (release 17), for all analyses. Seventy clinical trials are analysed, either coordinated by our Institution or by others. The Fondazione IRCCS Policlinico San Matteo sponsored 20 trials (29%); other Italian Centres sponsored 45 (64%) studies while 5 studies (7%) had an international sponsor. Twelve protocols (17%) were single centre, the remaining 58 being multicentre (60% national and 7% international). Forty-four studies (63%) were designed as randomized trials. The median percentage of satisfied items in methodology (items 9 to 21b), was 72% (IQR 50%-83%). However, for collection, management and statistical analysis (items 18-21b), the median percentage of correctly satisfied items was 56% (IQR 50%-83%). We will explore, present and discuss potential correlates of the lack of adherence to SPIRIT by modelling the number of unsatisfied items over the applicable number of items. Although many of the studies analysed followed the SPIRIT guidelines at least partially, more attention is needed to identify potential modifiable factors in order to increase the adherence to the guidelines for quality research.

[1] An-Wen Chan, MD, DPhil, Jennifer M. Tetzlaff, MSc, Douglas G. Altman, DSc, Andreas Laupacis, MD, Peter C. Gøtzsche, MD, DrMedSci, Karmela Krleža-Jerić, MD, DSc, Asbjørn Hróbjartsson, PhD, Howard Mann, MD, Kay Dickersin, PhD, Jesse A. Berlin, ScD, Caroline J. Doré, BSc, Wendy R. Parulekar, MD, William S.M. Summerskill, MBBS, Trish Groves, MBBS, Kenneth F. Schulz, PhD, Harold C. Sox, MD, Frank W. Rockhold, PhD, Drummond Rennie, MD, and David Moher, PhD, SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials, Ann Intern Med. 2013 February 05; 158(3): 200–207.
[2] Grech V., Write a Scientific Paper (WASP): Guidelines for reporting medical research, Elsevier, Early Human Development 134 (2019) 55–57.

TP29 A new testing strategy for a trial with two doses compared to placebo

Hougaard P.*
Lundbeck ~ Copenhagen ~ Denmark

The setting for this talk is a clinical trial covering two doses, which are to be compared to placebo. One traditional multiple testing procedure is the hierarchical test, starting with the high dose and only if this test is significant continue with the low dose. Compared to other multiple testing procedures, this one has its merits in case, the low dose has no or little effect. An alternative is to compare the average result of the two doses with placebo. This has best performance in case, the low and high doses have similar effect. This can be used to show study success; that is, effect of the drug, but it is not a closed testing procedure and therefore does not directly address, which dose(s) has(have) an effect. The suggested test is a compromise of these two approaches and uses computations similar to the first step of the MCP-Mod approach of Bretz et al (2005), but instead of the second step, it is made into a closed testing procedure, giving a conclusion for each dose. A simulation study shows that this approach has higher power to show difference than the hierarchical, when the low dose has an effect that is above 0.65 relative to the high dose effect and better than the average approach, when the effect is below 0.8 relative to the high dose.
The suggested approach has better performance than both traditional approaches when the relative effect is in the range 0.65 to 0.8, and this is a highly relevant range for many such clinical trials.
Bretz et al (2005) Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies, Biometrics 61, 738-748

TP30 Federated analysis of multiple data sources

Igl W.[1], Viaggi S.*[2]
[1]ICON, PLC ~ Uppsala ~ Sweden, [2]ICON, PLC ~ Milano ~ Italy

Background. Federated Analysis has received rapidly increasing attention over the last three years in the scientific literature. It describes the centralized analysis of de-centralized databases while preserving the privacy of personal data. Federated Analysis was proposed for the identification of rare adverse events in international post-marketing studies to overcome legal barriers [1]. Here, we will explain theoretical concepts and demonstrate the application of Federated Analysis to medical data using the DataSHIELD software to lower the technical barrier for its application. Concepts. A system of federated databases integrates heterogenous data sources with a single standard query interface. This distributed architecture allows the implementation of control mechanisms blocking the transfer of personal data, while at the same time the summation of local statistical loss functions over databases can be used to estimate parameter values in a global model. For example, Generalized Linear Models (eg Linear, Logistic) can be decomposed to allow Federated Analysis of horizontally partitioned data without loss of information. Horizontally partitioned data contains different sets of individuals, but the same sets of variables in the federated databases. Application. DataSHIELD is a component of the OBiBa software suite which supports the entire data management lifecycle including collection, harmonization, sharing, and analysis of data. DataSHIELD implements a multi-component client-server architecture (including the statistical analysis system R) to perform Federated Analysis. Recommendations are given on how to easily build a Federated Analysis system using virtualized software images (based on Docker technology) and an example of a Federated Analysis of a medical real-world evidence dataset (VigiBase) is presented. VigiBase is the WHO global database of reported potential side effects of medicinal products. Complete, detailed, step-by-step technical instructions on installation, data import, and data analysis of a Federated Analysis system including multiple databases are given in a related tutorial [2]. Conclusion. Federated Analysis as implemented in the DataSHIELD software offers a technical solution to overcome current institutional, legal, and ethical barriers for utilizing personal data while preserving privacy. The available virtualized software images of the DataSHIELD software dramatically lowers the technical barrier for setting up such a client-server architecture in academic or cross- industry scientific collaborations.

[1] R. Gedeborg, W. Igl, B. Svennblad, P. Wilén, B. Delcoigne, K. Michaëlsson, R. Ljung, & N. Feltelius (2022). Federated analyses of multiple data sources in drug safety studies. Pharmacoepidemiology and Drug Safety. https://doi.org/10.1002/pds.5587, Supporting Information (Original Technical Reports, .zip): https://tinyurl.com/bwb75efu
[2] W. Igl (2023). Federated Analysis with R/DataSHIELD – Installation, Data Import, Data Analysis [Tutorial]. https://wilmarigl.de/?p=424

TP31 3DIV: a comprehensive database of 3d genome and 3d cancer genome

Kim K.[2], Jang I.*[1], Kim M.[2], Lee B.[1], Jung I.[2]
[1]Korea Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology ~ Daejeon ~ Korea, Republic of, [2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST) ~ Daejeon ~ Korea, Republic of

Hi-C(High-throughput conformation capture) technology can capture genome-wide all-to-all interaction of genomic loci. It enables studying three-dimensional(3D) genome organization through an interaction contact map. Recent studies have revealed that genomic rearrangements in the cancer genome often disorganize higher-order chromatin structures, which can be pathogenic. However, unveiling the pathogenicity of the 3D genome is challenging due to the lack of appropriate tools specialized for disorganized higher-order chromatin structures. Here, we updated a 3D-genome Interaction Viewer and database(3DIV) by uniformly processing ~230 billion raw Hi-C reads to expand our contents to the 3D cancer genome. The updates of 3DIV are listed as follows: 1) the collection of 401 samples, including 220 cancer cell line/tumor Hi-C data and 28 promoters capture Hi-C data, 2) the live interactive manipulation of the 3D cancer genome to simulate the impact of structural variations and 3) the reconstruction of Hi-C contact maps by user-defined chromosome order to investigate the 3D genome of the complex genomic rearrangement. In summary, the updated 3DIV will be the most comprehensive resource to explore the gene regulatory effects of both the normal and cancer 3D genome. 3DIV is freely available at http://3div.kr

[1] Kim et al, 3DIV update for 2021: a comprehensive resource of 3D genome and 3D cancer genome, Nucleic Acids Research, 49 Database issue, 2021, 38-46
[2] Yang et al, 3DIV: A 3D-genome Interaction Viewer and database, Nucleic Acids Research, 46 Database issue, 2018, 52-57

TP32 Comparison of adverse event burden between treatment groups in clinical trials using mixture models

Jenner B.*[1], Chen M.[2], Wang Z.[3], Hsu Schmitz S.[1]
[1]*Statistics and Decision Sciences, Janssen Pharmaceutical ~ Allschwil ~ Switzerland,* [2]*Statistics and Decision Sciences, Janssen Research & Development ~ Shanghai ~ China,* [3]*IQVIA ~ Beijing ~ China*

Often in clinical trials (CT) a fraction of patients doesn't experience adverse events (AE). Within treatment, patient populations may consist of two subsets: with and without AEs. Treatment effect on AE burden can be assessed using mixture models (MMs) estimating the proportion of patients with AEs (binomial component) and the average AE burden among those who experienced AEs (non-zero component). If the models correctly reflect the mechanism generating data, they could offer a better insight into between-treatment difference. For illustration, data from two randomized double-blind placebo controlled CTs (Trials 1 and 2) were explored. Treatment-specific AEs in the first 12 weeks are of interest. The AE burden score (AEBS) is a continuous, exposure adjusted metric incorporating duration, severity, and frequency of AEs (Le-Rademacher et al. 2020). For patients without AEs, the AEBS is zero. With excess zero-values the distribution of AEBS is very skewed, therefore the binary-gamma (B-Gamma) MM is applied. Acknowledging incomplete information in severity or/and duration of some AE episodes, AE rate (AE frequency divided by exposure duration) is also evaluated. The binary-truncated negative binomial (B- tNB) MM is applied with exposure duration as offset. The ability of a model to discriminate AE burden (AEBS or AE rate) between treatment groups is assessed based on the ratio of treatment effect and its standard error. The discriminative ability of the MMs is compared to that of the reference models: Tweedie regression (for AEBS) and the negative binomial (NB) model (for AE rate), respectively. In Trial 1, for B-Gamma/B-tNB the discriminative ability was 11.7/10.8 for the non-zero component and 12.0/12.0 for the binomial component. When the MM components were combined, the overall discriminative ability was 14.9/14.0, slightly worse than 17.1/16.4 from Tweedie and NB. In Trial 2, the discriminative ability of the B-Gamma/B-tNB components was 4.8/5.6 for the non-zero component and 3.7/3.7 for the binomial component. The overall discriminative ability was 5.8/6.1, similar to 6.1/6.9 from Tweedie and NB. Based on the limited data above, the use of MMs suggests a trade-off with more insightful between- treatment comparisons but reduced ability to discriminate AE burden between treatment groups.

TP33 A simultaneous evaluation of superiority and non-inferiority for comparing screening tests.

Kobayashi M.*, Shinoda S., Saigusa Y., Yamamoto K.
Yokohama City University ~ Yokohama ~ Japan

In medicine, screening tests are particularly important for highly invasive tests. For example, Kobayashi et al. (2022) researched the diagnostic performance of novel approach for predicting advanced fibrosis in nonalcoholic fatty liver disease. They compared the performance of the conventional screening test and the novel screening test which is inspection items of conventional screening test and an extra inspection item. When the performances of those screening tests are compared, it is important that the number of patients undergoing highly invasive test are reduced by a novel test, and whether missing prevalent patients are clinically acceptable. That is, it is important to show both 1) superiority for a novel test to a conventional test in at least one in specificity (SP) or positive predictive value (PPV), and 2) non-inferiority for it in SP, PPV and sensitivity. For this, we can perform the analysis by using existing multiple comparison procedures. In this study, we propose a new test to improve the power of existing procedures. We evaluate the performance of the proposed method in terms of the actual type 1 error rate and power by simulations. In addition, we applicate the proposed method to the real data in Kobayashi et al. (2022). The results of the simulation studies indicated that the proposed method achieved higher power and could control the type I error rate. Furthermore, the real data analysis also showed the usefulness of the proposed method.

Kobayashi T, Ogawa Y, Shinoda S, Iwaki M, Nogami A, Honda Y, Kessoku T, Imajo K, Yoneda M, Saito S, Yamamoto K, Oeda S, Takahashi H, Sumida Y, Nakajima A. (2022). A 3-step approach to predict advanced fibrosis in nonalcoholic fatty liver disease: impact on diagnosis, patient burden, and medical costs. Scientific Reports, 12, doi: 10.1038/s41598-022-22767-z.

TP34 Korean nucleotide archive as a new data repository for nucleotide sequence data

Lee J.H.*, Park J., Kim S., Lee B.
Korea Research Institute of Bioscience & Biotechnology ~ Daejeon ~ Korea, Republic of

K-BDS (Korea BioData Station) collects all biological data generated from national R&D projects. To encourage the reuse of big data, database resources are provided to support research activities in academia and industry. Introducing the Korean Nucleotide Archive (KoNA), a repository of extensive sequencing data organized into multiple databases for each data type. KoNA has Korean Read Archive (KRA) recently collected more than 581TB of next-generation sequencing data from various genome projects, including the Korea Post Genome Project. To ensure data quality and interoperability, the database adopts the Standard Operating Procedures (SOPs) of the International Nucleotide Sequence Database Collaboration (INSDC). The SOP includes a continuous quality control process and manual inspection of submitted data and metadata using an automated pipeline. Users use GBoX, a high-speed transmission system, for fast and stable data transmission. In addition, data downloaded from KoNA can be processed directly through the cloud service Bio-Express, allowing sequencing data to be submitted and reused as an all-in-one service. To help users understand the database, we briefly introduce the current services and contents of KoNA (https://www.kobic.re.kr/kona).
The KoNA web service is implemented with Spring Boot, an application framework and inversion of control container (http://www.springsource.org, version 2.2.4), and Thymeleaf, a server-side Jave view template engine (version 3.0.12). All codes are developed using Eclipse (http://www.eclipse.org), an integrated development environment (IDE) that features the rapid development of Java-based web applications. To provide stable web services, the KoNA is hosted on a CentOS7 operating system with three servers: Spring Boot-based web server, an HBase server for database management, and a Gbox server for fast file upload and download. KBDS aims to collect all biological information generated from bio R&D. we plan to improve the structure of KGA, which currently has some functions to request and approve access to human data deposited with KRA. We believe that our relentless commitment to KoNA not only meets the unmet need of Korea's national sequencing repository but also contributes to the advancement of current genomics by providing valuable data sets to global researchers.

Korean Nucleotide Archive (KoNA). https://www.kobic.re.kr/kona

TP35 Replicability of simulation studies for the investigation of statistical methods: the replisims project

Luijken K.*[1], Lohmann A.[2], Alter U.[3], Claramunt Gonzalez J.[4], Clouth F.[5], Fossum J.[6], Hesen L.[2], Huizing A.[7], Ketelaar J.[2], Montoya A.[6], Nab L.[2], Nijman R.[2], Penning De Vries B.[2], Tibbe T.[6], Wang A.[8], Groenwold R.[2]
[1]*University Medical Center Utrecht ~ Utrecht ~ Netherlands,* [2]*Leiden University Medical Center ~ Leiden ~ Netherlands,* [3]*York University ~ Toronto ~ Canada,* [4]*Leiden University ~ Leiden ~ Netherlands,* [5]*Tilburg University ~ Tilburg ~ Netherlands,* [6]*University of California ~ Los Angelos ~ United States of America,* [7]*Netherlands Organization for Applied Scientific Research ~ Leiden ~ Netherlands,* [8]*University of Toronto ~ Toronto ~ Canada*

Results of statistical simulation studies evaluating the performance of statistical methods are often considered actionable and thus can have a major impact on the way empirical research is implemented. However, so far there is limited evidence about the reproducibility and replicability of statistical simulation studies [1, 2]. Therefore, eight highly cited statistical simulation studies were selected, and their reproducibility and replicability was investigated by teams of replicators with formal training in quantitative methodology. The replicator teams retrieved relevant information from the original publications and used it to write simulation code with the aim of replicating the results. The primary outcome was the feasibility of replicability based on reported information in original publications. Additionally, results of the replication were compared to those reported in the original publication. No simulation code was openly available for any of the identified simulation studies. Replicability varied greatly: Some original studies provided detailed information leading to almost perfect replication of results, whereas other studies did not provide enough information to implement any of the reported simulations. Replicators had to make choices regarding missing or ambiguous information in the original studies, error handling, and software environment. Factors that facilitated replication included descriptions of the data-generating procedure in graphs, formulas, structured text, and publicly accessible code and additional resources such as technical reports. Replicability of statistical simulation studies was mainly impeded by lack of information and sustainability of information sources. Reproducibility could be easily achieved for simulations studies by providing open code and data as a supplement to the publication. In addition, simulation studies should be transparently reported with all relevant information either in the research paper itself or in easily accessible supplementary material to allow for replicability.

[1] Boulesteix, A.-L., Hoffmann, S., Charlton, A., & Seibold, H. (2020). A replication crisis in methodological research? Significance, 17(5), 18-21.
[2] Lohmann, A., Astivia, O. L. O., Morris, T., & Groenwold, R. H. (2021). It's time! 10+ 1 reasons we should start replicating simulation studies.

Poster Sessions

TP36 Dealing with sparse networks of interventions: learnings from a simulation study

Luttenauer H.*[1], Arnould C.[2], Papadimitropoulou K.[3], Le Nouveau P.[4], Gauthier A.[5]
[1]Health Economics and Market Access, Amaris Consulting ~ London ~ United Kingdom, [2]Independent researcher ~ Paris ~ France, [3]Health Economics and Market Access, Amaris Consulting ~ Lyon ~ France, [4]Health Economics and Market Access, Amaris Consulting ~ Nantes ~ France, [5]Health Economics and Market Access, Amaris Consulting ~ Barcelona ~ Spain

Network meta-analysis (NMA) is a widely used approach to estimate the relative effect between treatments that have not been directly compared. However, in the era of personalised medicine, sparse disconnected networks, constructed on narrow target populations, can limit the implementation of the NMA. We previously conducted a simulation study to explore whether extending the network evidence, while increasing heterogeneity, could improve precision in sparse networks [1]. This work aims to extend the previous simulation study to generalize the findings. We initially simulated a star-shaped network consisting of three treatments, and gradually introduced complexity by adding more studies per direct comparison, and increasing heterogeneity, in six simulation scenarios assessing binary data. We performed 100 simulations per scenario, and conducted standard NMAs and network-meta regressions to determine the least biased scenario for estimating relative treatment effects. Bias was quantified for each scenario using the mean relative difference (MRD) to provide a ranking of least to most biased, across the different scenarios. Our results suggest that adding more studies while increasing heterogeneity can improve precision, even in the presence of high levels of heterogeneity. The scenario associated with the lowest bias was scenario 4, followed by scenario 3 (with MRD of 18.8% and 18.9% respectively, versus 35.0% and 25.3% for scenarios 1 and 2 respectively). Rankings showed that the scenario with the smallest number of trials had the highest probability of ranking first, but also the highest probability of ranking last. The scenario with the highest number of trials had the highest probability of ranking at least third. Meta-regressions improved the estimates' accuracy for scenarios 3 and 4, but the impact of meta- regressions heavily depends on the proportion of observed heterogeneity that has been arbitrarily chosen. Our findings suggest that extending network evidence, even with increased heterogeneity, can improve precision in sparse networks. However, the quality of included trials and the threshold for adding heterogeneity must also be considered. These findings have important implications for personalized medicine, where small and disconnected networks are increasingly common.

[1] Luttenauer, H., Le Nouveau, P., & Gauthier, A. (2021). CE4 Expanding Evidence Base VS Introducing Heterogeneity in Networks for Network Meta-Analyses: A Simulation Study. Value in Health, 24, S2-S3.

TP37 Diagnostic network meta-analytic methods for pulmonary embolism

Pagkalidou E.*[1], Doundoulakis I.[2], Apostolidou--Kiouti F.[1], Bougioukas K.[1], Papadopoulos K.[1], Tsapas A.[1], Farmakis I.[1], Antonopoulos A.[2], Giannakoulas G.[1], Haidich A.[1]
[1]Department of Hygiene, Social-Preventive Medicine and Medical Statistics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki ~ Thessaloniki ~ Greece, [2]First Department of Cardiology, Hippokration Hospital, National and Kapodistrian University ~ Athens ~ Greece

Systematic Reviews (SRs) of diagnostic test accuracy (DTA) studies comparing the accuracy of multiple index tests are increasingly being published, but they can be methodologically challenging. Eleven DTA network meta-analysis (NMA) models and two hierarchical meta-regression methods have been proposed to compare the performance of multiple tests according to their diagnostic accuracy [1,2]. Our objective was to apply different methods of DTA-NMA in studies reporting results of five imaging tests for the diagnosis of suspected pulmonary embolism (PE): pulmonary angiography (PA), computed tomography angiography (CTPA), magnetic resonance angiography (MRA), planar ventilation/perfusion (V/Q) scintigraphy and single photon emission computed tomography ventilation/perfusion (SPECT V/Q). We searched four databases PubMed, Cochrane Central, Scopus and Epistemonikos from inception until the end of December 2021 to identify SRs describing diagnostic accuracy of PA, CTPA, MRA, V/Q Scan and SPECT V/Q for suspected PE. The searches were last updated in June 2022. Study- level data were extracted from them and pooled using one hierarchical meta-regression approach (HSROC) and two DTA-NMA models to compare accuracy estimates of different imaging tests. Risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool and certainty of evidence using the Grading of Recommendations Assessment, Development and Evaluation framework. We identified 13 SRs, synthesizing data from 33 primary studies and for 4 imaging tests (PA, CTPA, MRA and V/Q Scan). HSROC meta-regression model using PA as reference standard showed that MRA had the best overall diagnostic performance with sensitivity=0.93 (95% confidence interval (CI): 0.76, 1.00) and specificity=0.94 (95% CI: 0.84, 0.99). However, NMA-DTA models indicated that V/Q scan had the highest sensitivity while CTPA was most specific. Selecting a different DTA-NMA method to assess multiple diagnostic tests can potentially affect the estimates of diagnostic accuracy. There is no established method, but the choice depends on the data and familiarity with Bayesian setting.

[1] Veroniki AA, Tsokani S, Paraskevaidis E, Mavridis D. Evaluating multiple diagnostic tests: An application to cervical cancer. Hell J Obstet Gynecol 2021;20:11–24. https://doi.org/10.33574/hjog.2161.
[2] Veroniki AA, Tsokani S, Agarwal R, Pagkalidou E, Rücker G, Mavridis D, et al. Diagnostic test accuracy network meta-analysis methods: A scoping review and empirical assessment. J Clin Epidemiol 2022. https://doi.org/10.1016/j.jclinepi.2022.02.001.

TP38 Generalized outlier detection for skewed distributions in biomedical signal

Park M.*, Jeong S.
Chungnam National University ~ Daejeon ~ Korea, Republic of

In medical signal research, removing noise is crucial for more accurate diagnoses and for identifying the cause of the noise. Outlier removal considering the overall trend is necessary for signals such as electrocardiograms (ECGs), which have fluctuations in measurement values over time. In addition, outlier removal for skewed data can cause a problem of reducing the length of the fence depending on the position of the median, which differs from the original intention.
To solve these problems, we propose an algorithm that applies the decomposition techniques based on the statistical empirical mode decomposition first, and then applies a generalized outlier detection method that can detect skewed outliers in the intrinsic mode functions. This proposed algorithm, which detects skewed influence values while removing time-series trends, was applied to simulation data, and its performance was demonstrated by applying it to ECG data with motion artifacts. As simulation data, samples following a beta distribution using various parameters were added to a sine function with a time trend, and the proposed algorithm showed more stable results than conventional skewed methods. In addition, in the real data analysis, we used ECG data with motion artifacts to which noise was added, and any added noise was detected along with other influence values.
Therefore, the proposed algorithm not only decomposed time trends but also effectively detected outlier detection with a skewed distribution. Through this study, we expect an appropriate detection of influential values in biomedical signal data containing extreme values and clearer identification of their causes.
D. Kim, M. Park, H.-S. Oh, Bidimensional statistical empirical mode decomposition, IEEE Signal Processing Letters, 19, 2012, 191-194.

TP39 Practical guide on statistical items in the new SPIRIT-DEFINE extension for early phase dose-finding trials

Rekowski J.*, Yap C.
The Institute of Cancer Research ~ London ~ United Kingdom

Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) 2013 provides guidance on the development of clinical trial protocols[1]. Although largely applicable to many types of trial designs, some trials using specialised designs may require additional protocol considerations. Early phase dose-finding trials are typically highly adaptive and usually employed as phase I or seamless phase I/II trials. Using dose escalation/de-escalation strategies, they aim to recommend a tolerable dose range for subsequent trials based on safety and other information such as pharmacokinetics, pharmacodynamics, biomarker activity, and clinical activity. They may study any intervention that can be given in different doses and/or schedules, e.g., drugs, vaccines, cell therapies, gene therapies, digital therapeutics, rehabilitation, or radiotherapy, and may either involve healthy volunteers or participants with a condition of interest. The overall quality of early phase dose-finding protocols from ClinicalTrials.gov in 2017-2023 was reported to be markedly variable and poor, with insufficient reporting in many applicable SPIRIT 2013 items[2]. SPIRIT 2013 has recently been extended to address the special features of early phase dose-finding trial protocols – in a process following the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) methodological framework for guideline development. The resulting consensus-driven SPIRIT-DEFINE (SPIRIT – DosE-FIndiNg Extension) statement recommends essential items to be included in early phase dose-finding trial protocols to promote transparency, completeness, and reproducibility of methods. In this presentation, we will emphasise the importance of a widespread implementation of this SPIRIT extension. We will focus on the statistical aspects of new items in SPIRIT-DEFINE and items modified from SPIRIT 2013. Such items cover a detailed elaboration of the trial design (e.g., adaptive features, timing of interim analyses, planned dose range with starting dose(s), dose allocation method, interim decision-making criteria, expansion cohort(s), operating characteristics, and dose transition pathways), clear definitions of analysis populations, and plans for handling intercurrent events and missing data. We will highlight the importance of including specific items and will provide practical advice of addressing them in discussing good examples. We envision that SPIRIT-DEFINE will ultimately improve participant safety and reduce research inefficiencies in early phase dose-finding trials through more standardised and transparent protocols.

[1] Chan, A.-W., et al., SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200-207.
[2] Villacampa, G., et al., Assessing the reporting quality of early phase dose-finding trial protocols: A methodological study. 2023; https://dx.doi.org/10.2139/ssrn.4399146.

Poster Sessions

TP40 Sample size estimation in clinical trials using ventilator-free days as primary outcome: a systematic review.

Renard Triche L.*[1], Futier E.[2], De Carvalho M.[3]
[1]Department of Perioperative Medicine, CHU Clermont-Ferrand ~ Clermont-Ferrand ~ France, [2]Department of Perioperative Medicine, CHU Clermont-Ferrand / iGReD, CNRS, INSERM, Université Clermont Auvergne ~ Clermont-Ferrand ~ France, [3]Université Clermont Auvergne, Health Library ~ Clermont-Ferrand ~ France

Ventilator-free days (VFDs) are a composite endpoint increasingly used as the primary outcome in critical care trials. However, because of the skewed distribution and competitive risk between components, sample size estimation remains challenging[1]. To systematically assess whether the sample size, as calculated to evaluate VFDs in trials, was congruent with VFDs' distribution and the impact of alternative methods on sample size estimation. A systematic literature search was conducted within the PubMed and Embase databases for randomized trials in adults with VFDs as the primary outcome until December 2021. After reviewing definitions of VFDs, we extracted the sample size and methods used for its estimation in each study. The data were collected by two independent investigators and recorded in a standardized, pilot-tested forms tool. Sample sizes were calculated using alternative statistical approaches, and risks of bias were assessed with the Cochrane risk-of-bias tool. Of the 26 clinical trials included, 19 (73%) raised "some concerns" when assessing risks of bias. Twenty-four (92%) trials were two-arm superiority trials, and 23 (89%) were conducted at multiple sites. Almost all the trials (96%) were unable to consider the unique distribution of VFDs and death as a competitive risk, features for which more complex models might be useful to provide the most reasonable sample size. Moreover, significant heterogeneity was found in the definitions of VFDs, especially regarding varying start time and type of respiratory support. Methods for sample size estimation were also heterogeneous and simple models, such as the Mann–Whitney–Wilcoxon rank- sum test, were used in 14 (54%) trials. Finally, the sample sizes calculated varied by a factor of 1.6 to 17.4. A standardized definition and methodology of VFDs, including the use of a core outcome set, seems to be required. Indeed, this could facilitate the interpretation of findings in clinical trials, as well as their construction, especially the sample size estimation which is a trade-off between cost, ethics, and statistical power[2].

[1] N. Yehya, M.O. Harhay, M.A.Q. Curley, et al., American Journal of Respiratory and Critical Care Medicine, 200, 2019, 828–836.
[2] S.-C. Chow, J. Shao, H. Wang, et al., Sample Size Calculations in Clinical Research, Chapman and Hall/CRC, 2017, 1-510.

TP41 Myofibril linearity indexes in muscle: a montecarlo-based analysis of performance

Rocchi E.*, Peluso S., Amatori S., Sisti D.
Dept. of Biomolecular Sciences, Unit of Biomathematics and Biostatistics, Urbino University ~ Urbino ~ Italy

An interesting problem in analysing the image of muscle tissue is represented by the attempt to evaluate the linearity of myofibrillar structures and their possible deviation from a straight line. We have recently proposed two different indices for estimating the linearity of myofibrils: the first is based on the reciprocal of the ratio between the sum of the lengths of the sarcomeres and the length between the head of the first sarcomere and the tail of the last one. The second is based on the length of the mean vector, a concept mutualized by circular statistics [1]. In this contribution, we wanted to verify the performance of these two indices, by means of simulations using Montecarlo methods. Sequences of central axes of sarcomeres have been generated in which the angle formed by the new axis with respect to the previous one is randomly generated according to a Von Mises distribution with variable concentration parameter k (ranging from 0 to 50). Also, the length of the sarcomere (which however can only influence the first method) was randomly generated according to a normal Gaussian distribution (mean 1, standard deviation 0.25). The simulations carried out at various concentration parameters k, show different behaviors as the length of the myofibril varies (ie the number of sarcomeres that compose it). In the first method, the estimate varies significantly as the number of sarcomeres varies, while in the second (based on circular statistics) the estimate remains substantially stable (even if obviously the standard error varies). In conclusion, the simulations suggest using the mean value length method, even though it is computationally more demanding.

[1] E. Rocchi, S. Peluso, S. Amatori, D. Sisti. European Journal of Translational Myology, 32, 2022, 10736.

TP42 Variability in primary outcome reporting in clinical trials for older adults with depression

Rodrigues M.*[1], Oprea A.[1], Johnson K.[1], Dufort A.[1], Sanger N.[1], Ghiassi P.[2], Sanger S.[1], Syed Z.[1], Panesar B.[1], D"Elia A.[1], Parpia S.[1], Samaan Z.[1], Thabane L.[1]
[1]McMaster University ~ Hamilton ~ Canada, [2]Canadian Partnership Against Cancer ~ Toronto ~ Canada

Findings from randomized controlled trials (RCTs) are synthesized in meta-analyses to inform evidence-based health care. However, unclear specification of outcomes in published RCTs impedes synthesis efforts, and hinders knowledge translation and clinical decision-making. The variability in outcomes measured in trials for older adults with major depressive disorder (MDD) has been described; however, the comprehensiveness of outcome reporting remains unknown. The objective of our study was to assess the reporting of primary outcomes in trials evaluating treatments for geriatric MDD. Eligible studies for outcome reporting assessment included RCTs which assessed interventions for older adults with MDD, and were published between 2011-2021, specifying a single, discernable primary outcome. Outcome reporting was assessed independently and in duplicate using a pre- defined checklist of 58 reporting items. Information for primary outcomes in each published trial were scored as "fully reported", "partially reported", or "not reported" for each item on the checklist. Thirty-one of 49 identified publications had a single, discernable primary outcome, and were eligible for outcome reporting assessment. The majority of RCTs (60%) did not fully report over half the 58 checklist items. Less frequently-reported items included: outcome measurement instrument properties (range: 5% to 50%) and justification of criteria used to define clinically meaningful change (9%).
Limitations: Our study did not included trials which assessed multiple primary outcomes, or RCTs which were unclear in identification of their primary outcomes. The state of primary outcome reporting in geriatric depression trials may potentially differ from our study findings. There is variability in reporting of geriatric depression RCTs, with frequent omission of key details regarding primary outcomes. Omission of key details may impede interpretability of study findings, and hinder synthesis efforts which inform clinical guidelines and evidence-based decision-making. Consensus on the minimal criteria for outcome reporting in geriatric MDD trials is required.
Rodrigues M, Sanger N, Dufort A, Sanger S, Panesar B, D'Elia A, et al. Outcomes reported in randomised controlled trials of major depressive disorder in older adults: protocol for a methodological review. BMJ Open. 2021;11:e054777.
Rodrigues M, Syed Z, Dufort A, Sanger N, Ghiassi P, Sanger S, et al. Heterogeneity across outcomes reported in randomized controlled trials for older adults with major depressive disorder: findings from a systematic survey. Under Rev J Clin Epidemiol. 2023.

TP43 Collaboration opportunities on biostatistics with ema

Rueckbeil M.*
European Medicines Agency ~ Amsterdam ~ Netherlands

The European Medicines Agency (EMA) operates at the heart of the European Medicines Regulatory Network. Its mission is to foster scientific excellence in the evaluation and supervision of medicines, for the benefit of public and animal health in the European Union (EU). This includes advice and evaluation with regard to methodology for clinical trials. Maintaining a good dialogue with researchers, including biostatisticians, directly supports EMA's mission. Several mechanisms are in place to foster exchange and collaboration between biostatistical researchers and EMA. These include support and potential involvement from EMA in externally funded regulatory science and public health research projects. In addition, there are opportunities for biostatistical researchers to contribute to EMA's work through studies procured by EMA, working with EMA as a collaborating expert, and providing comments on regulatory guidance documents under public consultation. Selected examples are presented to illustrate opportunities for methodological exchange and collaboration between biostatistical researchers and EMA. This includes information about an EMA reflection paper on single-arm trials which has been published for public consultation in 2023. Exchanges and collaborations between biostatistical researchers and EMA create mutual benefits by leveraging expertise to support EMA fulfill its mission while enhancing the impact and utility of research outcomes. Increasing the awareness and strengthening of existing opportunities will further benefit public and animal health in the EU.
Collaborating expert: https://careers.ema.europa.eu/content/Collaborating-Expert/?locale=en_GB Information for Academia: https://www.ema.europa.eu/en/partners-networks/academia
EMA's scientific guidelines on biostatistics: https://www.ema.europa.eu/en/human-regulatory/research- development/scientific-guidelines/clinical-efficacy-safety/biostatistics

TP44 Evaluating zero-inflated models with a different base distribution in clinical trials and medical research

Shahmandi M.*, Mossop H., Chadwick T., Wason J.
Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

In the field of medical science and clinical trials, it is often observed that the outcome variable exhibits a high number of zeros and a strong right skewness. This can lead to poor fit of standard distributions and can result in unjustified conclusions when assessing the impact of independent variables or risk factors on the outcome variable. Currently, there is no universally accepted statistical distribution for data with these characteristics. Zero-inflated models using Negative Binomial (ZINB) as the base distribution are commonly used, but replacing the base distribution with a more suitable one like discretised lognormal distribution (DLN) could improve the models. DLN provides a zero-inflated discretised lognormal distribution (ZIDLN), where the location parameter μ primarily influences the occurrence of very low values, while the scale parameter σ affects how high the count data are likely to be. The present study aims to demonstrate the applicability and usefulness of ZIDLN versus ZINB in clinical and medical research using simulation and a real dataset available in the "MixAll" package in R. To conduct the simulation, samples of count data with different sizes were generated from a discrete Weibull distribution, allowing for control over the proportion of count zeros generated and the skewness of the data. The generated data was then modified to have varying degrees of zero- inflation. The simulation results demonstrate that both ZIDLN and ZINB models have the ability to accommodate different levels of zero-inflation in the data. While both models exhibit comparable estimated probability of zero-inflation and Akaike information criterion (AIC), ZIDLN produces more precise estimates by providing narrower confidence intervals, particularly for smaller sample sizes. Zero-inflated models applying DLN instead of NB as a base distribution, could offer more precise estimates and conclusions in clinical trial studies that typically have smaller sample sizes.

[1] M. Shahmandi, P. Wilson, M. Thelwall, A new algorithm for zero-modified models applied to citation counts. Scientometrics, 125, 2020, 993–1010.
[2] M. Thelwall, The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. Journal of Informetrics, 10, 2016, 336-346.

TP45 Hypothesis testing procedure of f1 scores for binary and multi-class classification in the paired design

Takahashi K.*[1], Yamamoto K.[2], Kuchiba A.[3], Shintani A.[4], Koyama T.[5]
[1]Hyogo Medical University ~ Nishinomiya ~ Japan, [2]Yokohama City University ~ Yokohama ~ Japan, [3]Kanagawa University of Human Services ~ Yokosuka ~ Japan, [4]Osaka Metropolitan University Graduate School of Medicine ~ Osaka ~ Japan, [5]Vanderbilt University Medical Center ~ Nashville ~ United States of America

In modern medicine, medical tests are used for various purposes including diagnosis, disease screening, prognosis, and risk prediction. Some measures exist to quantify the test performance, and sensitivity, specificity, and positive and negative predictive values which are commonly used for binary tests. Additionally, the F1-score for binary tests (biF1), has also come to be used in the medical field. Most of measures for performance of medical tests are only applicable to binary classification data, and multi-class classification data need to be dichotomized to compute these measures. As measures of multi-class classification performance, two types of measures have been proposed: a micro- averaged F1-score (miF1) and a macro-averaged F1-score (maF1)[1]. The miF1 pools per-sample classifications across classes and then calculates the overall F1-score. Contrarily, the maF1 computes an unweighted mean of the F1-scores for each label. Some statistical methods for inference have been proposed for the biF1, and the methods for computing confidence intervals of the miF1 and maF1 were proposed recently[2]. However, statistical testing procedures of F1 scores for binary and multi-class classification in the paired designs have not been proposed yet. Thus, we aim to provide the methods for comparing the biF1s, miF1s, and maF1s. Using the delta-method and the multivariate central limit theorem, we developed Wald and score statistics for comparing the biF1s, miF1s, and maF1s. A simulation study was conducted to evaluate the performance of the test statistics. Simulation results showed that the empirical type I error rates tended to be larger than nominal type I error rate (0.05) for Wald statistics when the sample size is relatively small. Contrarily, the empirical type I error rates with score statistics are close to 0.05 for all sample sizes. The empirical powers of Wald statistics and score statistics were similar, especially when the sample size is large. Through the simulation study, we conclude that the proposed hypothesis testing procedure based on the score statistics is useful for comparing the biF1s, miF1s, and maF1s in paired study design.

[1] CD. Manning, P. Raghavan, H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008, 234-265.
[2] K. Takahashi, K. Yamamoto, A. Kuchiba, T. Koyama, Applied intelligence (Dordrecht, Netherlands), 52, 2022, 4961-4972.

TP46 Foreign body injuries recognition and management through text analysis of case reports

Urru S.*[1], Sciannameo V.[2], Ahsani--Nasab S.[1], Lorenzoni G.[1], Azzolina D.[3], Gregori D.[1], Berchialla P.[2]
[1]Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova ~ Padova ~ Italy, [2]Department of Clinical and Biological Sciences, University of Torino ~ Torino ~ Italy, [3]Department of Environmental Sciences and Prevention, University of Ferrara ~ Ferrara ~ Italy

Foreign body (FB) injuries are a severe public health problem in children, which may cause significant morbidity depending on the characteristics of the FB, its anatomical location, the child's age, and delays in diagnosis, and occasionally lead to a fatal outcome [1]. The annual incidence rate of FB injuries in children presenting in a pediatric emergency department is increasing. Given the problem's relevance and potential adverse complications, both prevention and timely recognition are important to improve patient outcomes. Delays in diagnosis of FB injuries still occur due to insufficient time to understand the history behind a critical presentation, lack of information on the event, which a supervising adult may not witnessed, or the specific symptoms. The purpose of this study is to analyze FB injuries case reports to acquire a better understanding of the mechanism of injuries. A total of 299 case reports on FB injuries from 2017 to 2021 were collected. Documents were pre- processed by filtering out non-English texts, the reference section and all words not included in the main text. A corpus of 193 single case reports was created and the Document Term Matrix (DTM) was built. DTM contains the word frequencies over the documents which were used to perform Topic Modelling (TM) and Generalized Linear Mixed Model (GLMM). TM [2] is a Natural Language Processing technique for the latent structure (i.e., topics) discovering in a collection of documents; it is a generative process of word counts in which words generate the topic distribution which in turn generate the document distribution. The identified topics highlight the relation between body part and object which determine the injury type. GLMM was applied to the most frequent words to detect the symptoms of each injury type; topics were included in the model as random effects to improve the estimation. The innovation of this work is the combined use of TM and GLMM on FB injuries case reports which can lead to the construction of guidelines for pediatric emergency department to facilitate the injury recognition and to intervene promptly.

[1] D. Passali, D. Gregori, G. Lorenzoni, S. Cocca, M. Loglisci, FM. Passali, L., Bellussi, Foreign body injuries in children: a review, Acta Otorhinolaryngologica Italica, 35(4), 2015, 265-71
[2] D.M. Blei, A.Y. Ng, M.I., Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research, 3(2), 2003, 993-1022

TP47 On the cluster randomised trials where intervention effects are heterogeneous

Yamaoka K.*[1], Watanabe J.[2], Adachi M.[3], Watanabe M.[4], Suzuki A.[1], Tango T.[5]
[1]Teikyo University Graduate School of Public Health ~ Tokyo ~ Japan, [2]The Department of Nutrition Management, Minami Kyushu University ~ Miyazaki ~ Japan, [3]Nutrition Support Network LLC ~ Sagamihara ~ Japan, [4]Showa Women's University ~ Tokyo ~ Japan, [5]Center for Medical Statistics ~ Tokyo ~ Japan

Cluster randomised trials (CRTs) to assess school-based interventions to improve adolescents' health are often conducted. In these CRTs, healthy adolescent populations are generally assessed, and no overall effect was observed due to heterogeneous effects. Usually, when a significant interaction term has been observed, a relevant subgroup analysis is performed. For example, in our formerly performed CRT [1], we examined the effects of a parent-involved school-based lifestyle education program to reduce adolescents' subjective psychosomatic symptoms and to improve dietary intakes in the sixth month. In that study, the change from the baseline of energy intake was analysed, and the result showed a statistically significant baseline*treatment interaction (p=0.02). The subgroup analysis showed a significant effect only in the lower-weight group (p<0.001). However, to perform an intention-to-treat (ITT) analysis, new outcome measures taking baseline heterogeneity into account should be devised according to the educational objective. In Japan, the Ministry of Health and Welfare recommends "Dietary Reference Intakes for Japanese" by age group [2] to prevent lifestyle-related diseases. In this presentation, we shall propose the following outcome measures based on that reference value. The proposed outcome measure was a dichotomous variable indicating whether the energy intake in the sixth month was within 20% of the individual reference value. The proportion was analysed using a baseline-adjusted mixed-effects logistic model, and a significant effect was observed (p<0.008). The proposed method was based on ITT and is useful when the intervention effects are expected to be heterogeneous. In school-based CRTs that include healthy subjects, it is necessary to conduct an analysis that pays attention to the meaning of the changes. In the presentation, we will also show the results of the other outcomes.

[1] Watanabe J, et al. BMC Public Health 2022;22:461. https://doi.org/10.1186/s12889-022-12832-7.
[2] Ministry of Health, Labour, and Welfare. Overview of the Dietary Reference Intakes for Japanese (2020). https://www.mhlw.go.jp/content/10900000/000862500.pdf.

Poster Sessions

TP48 Application of graph theory to integrate complex relationships among heterogeneous biological data

Yoon B.*
Korea Research Institute of Bioscience and Biotechnology ~ Daejeon ~ Korea, Republic of

The goal of precision medicine is to provide personalized treatment for each patient. However, the diversity and volume of biomedical data and the proliferation of clinical knowledge derived from myriad biological databases and publications make this difficult. Therefore, understanding the relationship between complex and heterogeneous biological data has become very important[1]. To solve this, we collected various biological data (protein-protein interaction, drug-target, gene- disease, etc) from various existing sources and pre-processed them such as deduplication to construct a graph database containing about 150,000 nodes and 100 million relationships. We compare performance between Neo4j, a popular graph database, and MySQL to show that graph- based databases can be used for complex biological relationships. MySQL, a traditional relational DB, showed latent or incomplete responses to complex queries with multiple join statements, whereas Neo4j showed very fast responses to the same queries. These results demonstrate that using a graph-based database is an efficient way to store and retrieve complex biological relationships. And this biological inter-connectivity enables the use of clinical biostatistics techniques and network analysis techniques to reveal hidden patterns and to infer new knowledge.

[1] da Silva, W. M., Wercelens, P., Walter, M. E. M., Holanda, M., & Brígido, M. (2018). Graph databases in molecular biology. In Advances in Bioinformatics and Computational Biology: 11th Brazilian Symposium on Bioinformatics, BSB 2018, Niterói, Brazil, October 30–November 1, 2018, Proceedings 11 (pp. 50-57). Springer International Publishing.

TP49 Performance comparisons between clustering models for reconstructing ngs results from technical replicates

Zhai Y.*[2], Roy P.[1], Bardel C.[3]
[1]Service de Biostatistique-Bioinformatique, Hospices Civils de Lyon ~ Lyon ~ France, [2]Laboratoire de Biométrie et Biologie Évolutive, CNRS UMR 5558 ~ Villeurbanne ~ France, [3]Service de Génétique, Groupement Hospitalier Est, Hospices Civils de Lyon ~ Lyon ~ France

To improve the performance of an individual's DNA sequencing results, a number of researchers use replicates of DNA sequencing results from the same individual or from monozygotic twins. With such replicates, several statistical methods allow obtaining sequencing results with better performance. However, though usual, these models have been rarely compared, their comparison results often unclear, and their final conclusions controversial. In this work, several clustering models were compared regarding their abilities to reconstruct a new callset with improved performance from several genome sequencing replicates Three technical replicates of genome NA12878 were considered and five model types were compared (consensus, latent class, Gaussian mixture, Kamila-adapted k-means, and random forest) regarding four performance indicators: sensitivity, precision, accuracy, and F1-score. The consensus model improved the precision by 0.1%, whereas the latent class model brought 1% precision improvement (97% to 98%) without compromising sensitivity (= 98.9%). The Gaussian mixture model and random forest provided callsets with higher precision (both > 99%) but at the price of lower sensitivity vs. no use of combination model. Kamila increased precision (> 99%) while keeping a high sensitivity (98.8%) and proved to ensure the best overall performance. According to precision and F1-score indicators, the compared non-supervised clustering models that combine multiple callsets are able to improve sequencing performance vs. supervised models tested elsewhere. Among the models compared, the Gaussian mixture model and Kamila offered non- negligible precision and F1-score improvements and may be recommended for new callset reconstruction from replicates.

TP50 Text classification to automate abstract screening using machine learning

Zimmermann S.*[1], Vey J.[1], Pilz M.[2]
[1]*Institute of Medical Biometry, Heidelberg University Hospital ~ Heidelberg ~ Germany,* [2]*Department Optimization, Fraunhofer Institute for Industrial Mathematics (ITWM) ~ Kaiserslautern ~ Germany*

Systematic reviews synthesize all available evidence on a specific research question. A paramount task in this is the comprehensive literature search, which should be as extensive as possible to identify all relevant studies and reduce the risk of reporting bias. The identified studies need to be screened, which is time-consuming and resource intensive. In the first stage of this process, the title-abstract screening (TIAB), abstracts of all initially identified studies are screened and classified regarding their inclusion or exclusion for full-text screening. Conventionally, this is accomplished by two independent human reviewers. In the last years, there has been research to automate this process[1,2]. We present a semi-automated approach to TIAB screening using natural language processing (NLP) and machine learning (ML) based classification that was applied within a systematic review project. The total 4460 identified abstracts and their titles were split into a training (1/3) and test set (2/3) after which they were preprocessed into numerical matrices using NLP. Based on the processed training data, variable selection was performed, subsequently, different ML algorithms were trained using 5- fold cross-validation and grid search for the respective tuning parameters. The AUC value was used as an optimization criterion and the decision of the two human reviewers was used as reference. The algorithms were evaluated on the test set. The Random Forest showed the best performance (AUC: 96%). Choosing a cut-off to avoid missing any relevant abstract (n=136) resulted in only 755 false positives (FP rate: 26.5%). Conversely, 2089 abstracts were correctly classified as to be excluded (FN rate: 0%). In our case study, the manual TIAB screening workload for the second reviewer could be reduced by about 70%. We propose an approach where a ML model can replace one human reviewer after being trained on a sufficient number of abstracts. The second reviewer only needs to get involved in cases of discrepancies between the decision of the first reviewer and the classification model. Our ML-based text classification approach proved to be powerful, adaptable, and it considerably reduced human workload in creating systematic reviews.

1. O'Mara-Eves, A., Thomas, J., McNaught, J. et al. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 4, 5 (2015). https://doi.org/10.1186/2046-4053-4-5
2. Lange, T, Schwarzer, G, Datzmann, T, Binder, H. Machine learning for identifying relevant publications in updates of systematic reviews of diagnostic test studies. Res Syn Meth. 2021; 12: 506–515. https://doi.org/10.1002/jrsm.1486

Poster Sessions

TP51 Constructing dna methylation biomarkers for cardiovascular diseases by penalized multilevel multitask learning

Cappozzo A.*[1], Ieva F.[1], Giorito G.[2]
[1]*Mox Lab, Department of Mathematics, Politecnico di Milano ~ Milano ~ Italy,* [2]*IRCCS Istituto Giannina Gaslini ~ Genova ~ Italy*

DNA methylation (DNAm) is an epigenetic process that regulates gene expression, typically occurring in cytosine within CpG sites (CpGs) in the DNA sequence. DNAm variability is related to lifestyle and environmental risk factors, providing an unbiased proxy of an individual state of health. By regressing blood measured quantities (response variables) on methylation levels it is possible to construct DNA methylation biomarkers [2]. Such surrogates possess extensive advantages over their blood- measured counterparts, as they can directly take into account genetic susceptibility and subject specific response to risk exposure. This work aims to propose a framework for mixed-effects multitask learning to create a multivariate DNAm biomarker for cardiovascular diseases when multiple risk factors are to be considered and a structured dependence pattern exists among the study samples. We develop a penalized estimation scheme using an expectation-maximization algorithm for mixed- effects multitask learning. The procedure allows for the inclusion of any penalty criteria previously developed for fixed-effects models, extending results recently proposed in the literature [1]. Profiting from the Italian multi-center branch of the European Prospective Investigation into Cancer and Nutrition (EPIC) study we construct a multivariate DNAm surrogate biomarkers for multiple correlated risk factors for cardiovascular diseases. We show that our approach, modeling multiple outcomes together and coupled with a multivariate group-lasso penalty, outperforms state-of-the-art alternatives both in predictive power and bio-molecular interpretation of the results. Our proposed framework for mixed-effects multitask learning provides a general approach to develop multivariate DNAm biomarkers from high-dimensional predictors in multi-center studies. The devised DNAm surrogates yield patient-specific risk-indicators appropriate for prediction tasks such as CVD prevention. We further find significant enrichment in DNA regions of molecular pathways regulating apoptosis, oxidative stress, and the immune system. The analysis suggests that our approach can contribute to future research aimed at identifying novel DNAm biomarkers for non-communicable diseases.

[1] F. Rohart, M. San Cristobal, B. Laurent. "Selection of fixed effects in high dimensional linear mixed models using a multicycle ECM algorithm." Computational Statistics & Data Analysis 80 (2014): 209- 222.
[2] J. Zhong, G. Agha, A. A. Baccarelli. "The role of DNA methylation in cardiovascular risk and disease: methodological aspects, study design, and data analysis for epidemiological studies." Circulation research 118, no. 1 (2016): 119-131.

Poster Sessions

TP52 Restricted mean survival time provides intuitive and robust absolute treatment effect estimates in risk strata

Maas C.C.*[1], Dinmohamed A.G.[2], Kent D.M.[3], Kersten M.J.[4], Van Klaveren D.[1]
[1]Erasmus University Medical Center ~ Rotterdam ~ Netherlands, [2]Netherlands Comprehensive Cancer Organisation (IKNL) ~ Utrecht ~ Netherlands, [3]Tufts Medical Center ~ Boston ~ United States of America, [4]Cancer Center Amsterdam ~ Amsterdam ~ Netherlands

The absolute effect of a treatment on a survival outcome is generally assessed by the difference in Kaplan-Meier estimates of the survival probability at a certain time point. To quantify the heterogeneity of absolute treatment effects, these estimates are stratified by patient subgroups[1]. However, Kaplan- Meier estimates at a specific time point are difficult to interpret in case of non-proportional treatment effects and are imprecise in small strata, especially with substantial censoring. In contrast, the difference in restricted mean survival time (RMST)–the average time-to-event over a fixed follow-up period–is an absolute treatment effect measure that is better interpretable and more robust to censoring since it is based on the entire Kaplan-Meier curve[2]. We aimed to study the use of RMST for estimating absolute treatment effects in risk-based patient subgroups. From the nationwide, population-based Netherlands Cancer Registry, we selected 1,626 adult patients diagnosed with advanced-stage diffuse large B-cell lymphoma in the Netherlands between 2014 and 2018. We compared overall survival (OS) between patients treated with six cycles of chemotherapy every 21 days (6xR-CHOP21) or the same regimen with two additional cycles of rituximab (6xR- CHOP21+2R). To assess treatment heterogeneity, we compared OS across four risk strata defined by the International Prognostic Index (IPI). To reduce confounding, we balanced the characteristics in the two treatment groups using stabilized inverse propensity score weights. In four groups ranging from low to high IPI, the difference in 5-year RMST in years showed a clear pattern of increasing OS benefit when receiving 6xR-CHOP21+2R versus 6xR-CHOP21: 0.20 [95% confidence interval: -0.05–0.44]; 0.19 [-0.11–0.48]; 0.37 [0.06–0.68]; 0.63 [0.23–1.03]. These RMST differences are more precise and easier to interpret than the differences in survival probability estimates at 5 years: -0.05 [95% confidence interval -0.17–0.06], 0.05 [-0.06–0.16], 0.10 [-0.03–0.23], 0.12 [-0.01–0.26]. Absolute treatment effect estimates based on RMST were easily interpretable and presented a more robust pattern across risk strata. The RMST helps assess treatment effect heterogeneity in data with randomized treatment allocation and in observational data with negligible unmeasured confounding. This approach needs to be validated systematically in multiple studies.

[1] Kent, D.M., E.W. Steyerberg, and D. van Klaveren, Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. The BMJ, 2018. 363.
[2] Uno, H., et al., Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol, 2014. 32(22): p. 2380-5.

TP53 Multi-outcome feature selection via anomaly detection autoencoders for radiogenomic in breast cancer patients

Mapelli A.*[1], Massi M.C.[1], Rancati T.[3], Ieva F.[2]
[1]HDS – Health Data Science Center, Human Technopole ~ Milano ~ Italy, [2]MOX – Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano ~ Milano ~ Italy, [3]Data Science Unit, Fondazione IRCCS – Istituto Nazionale dei Tumori ~ Milano ~ Italy

Severe long-term side effects occur in approximately 5% of cancer patients receiving radiotherapy, significantly affecting their quality of life. Subjects' radiosensitivity can determine various types of Late Toxicities (LTs). The identification of genetic biomarkers predictive of LTs' development, and their inclusion in clinical risk models, may improve personalized treatment planning. To this aim, a Feature Selection (FS) method for biomarker discovery can be employed. In this setting, the FS should be robust to class imbalance (especially for rare symptoms) and feature noise from genome data imputation, while effectively account for gene-gene interactions to capture the true generative mechanism of toxicity endpoints. Moreover, simultaneously modelling several LTs is important to identify predisposing factors associated with radiosensitivity without explicitly defining it. In this work we propose a Deep Learning-based multivariate FS method for high-dimensional data in the presence of high-order interaction and noise. The proposed method expands on the Deep Sparse AutoEncoder Ensemble (DSAEE) proposed in [1] by introducing multi-endpoint FS and controlling imputation noise with a tailored denoising training procedure. In the original DSAEE, an ensemble of autoencoders is trained to reconstruct the control group, and tested on a mixed population of controls and cases. The most discriminant features are selected comparing the distribution of reconstruction errors between groups. Multivariate FS is accomplished by redefining the control group (i.e. patients without any LT), and training an ensemble for each endpoint, to eventually merge the sets of selected features. We confirm via simulations that our methodology outperforms the univariate one in identifying highly informative features for class separation. We then test it on the RADPrecise Breast Cancer Cohort, and incorporate the selected variants into clinical risk models through the interaction-aware method for polygenic risk scoring (PRSi) [2]. PRSi integration determined a significant increase in models' performance, highlighting the predictive and descriptive qualities of the variants selected by our multivariate DSAEE. This work introduces a methodology for FS in a multi-outcome binary framework that effectively addresses the challenges of genomic studies. The model's clinical applicability is a crucial aspect of its significance.

[1] M. C. Massi, F. Gasperoni, F. Ieva et al.: "Feature selection for imbalanced data with deep sparse autoencoders ensemble", Statistical Analysis and Data Mining: The ASA Data Science Journal, 15(3), June 2022, 376-395.
[2] N. R. Franco, M.C. Massi, F. Ieva et al.: "Development of a method for generating SNP interaction-aware polygenic risk scores for radiotherapy toxicity", Radiotherapy and Oncology, 159, June 2021, 241-248.

TP54 A personalized algorithm to detect cardiac arrythmia and major bleeding in advanced heart failure patients

Moazeni M.*[1], Numan L.[2], Szymanski M.[2], Vander Kaaij N.[3], Asselbergs F.[3], Van Laake L.[2], Emmeke A.[1]
[1]Utrecht University ~ Utrecht ~ Netherlands, [2]Department of Cardiology, University Medical Centre Utrecht ~ Utrecht ~ Netherlands, [3]Department of Cardiothoracic Surgery, University Medical Centre Utrecht, University of Utrecht ~ Utrecht ~ Netherlands

Advanced heart failure patients usually require a donor heart for destination therapy. However, since donor hearts are scarce, left ventricular assistant devices (LVADs) have become a popular alternative. Unfortunately, patients often experience complications with this treatment. Telemonitoring LVAD parameters such as power and flow may improve outcomes by detecting early signs of deterioration. Currently, LVADs use a simple monitoring approach with a fixed threshold for all patients, resulting in delayed detection of adverse events and false alarms. To address this, we developed a personalized algorithm that can detect unscheduled admissions caused by common complications like cardiac arrhythmia and major bleeding. This tailored algorithm allows for better detection and management of complications in individual patients. The algorithm (PRECISION-LVAD) uses patient-tailored thresholds to identify abnormal power and flow observations. It employs a linear mixed-effects (LME) model that considers pump parameters of a group of stable patients without any admission and the longitudinal data of each individual patient. This results in a personalized mean pump value that is flexible and reflects the patient's stable historical baseline. The patient-specific mean is then subtracted from real-time measurements to obtain residuals, which are smoothed with an exponentially weighted moving average (EWMA) statistical process control chart, and compared to upper and lower control limits determined by the EWMA control chart. If the smoothed residuals exceed these control limits, the algorithm triggers an alarm. Our findings indicate that PRECISION-LVAD was capable of detecting 59% and 79% of cases related to CA and MB, with a low false alarm rate (FAR) of 2%. However, the FAR varied between patients. Furthermore, the median number of days between the first alarm and admission due to CA or MB was 6.5 and 7.0 days, respectively. Although PRECISION-LVAD shows promise as a powerful tool for detecting CA and MB, some events were still not detected by the algorithm. Therefore, continuous refinement of the algorithm using data streams is necessary. One possible approach is to use latent variable models such as Hidden-Markov models to monitor patients based on their switching hidden states. This would allow for more accurate and timely detection of CA and MB.

1. Numan L, Moazeni M, Oerlemans MIFJ, Aarts E, Van Der Kaaij NP, Asselbergs FW, et al. Data- driven monitoring in patients on left ventricular assist device support. Expert Rev Med Devices. 2022;19(9):677–85.

TP55 Personalized scheduling of biomarker measurements for heart failure surveillance programs with competing risks

Petersen T.*, Kardys I., Rizopoulos D.
Erasmus MC ~ Rotterdam ~ Netherlands

Patient surveillance is a form of personalized medicine where periodically collected biomarkers are used to provide patient-specific risk estimates for better informed treatment decisions in patients at- risk of an adverse event. The time interval between the periodic measurements is often fixed in existing surveillance programs. However, recent studies have shown that the efficiency and efficacy of these programs can be improved by incorporating patient-specific information into scheduling decisions.[1] Until now, these optimizations have only been studied using single biomarkers and adverse events. An example of patient surveillance is found in heart failure, a heterogeneous condition linked to various biological processes, where accounting for multiple biomarkers may provide complementary information, and where various clinical endpoints are relevant. In this work, we extend personalized scheduling strategies to clinical settings with multiple biomarkers and competing events, and assess its benefits in a population of stable heart failure patients using the Bio-SHiFT study. Consider a patient in a hypothetical surveillance program in whom multiple biomarkers have been measured at various past time points. Using joint models for longitudinal and time-to-event data, dynamic personalized risk predictions can be estimated for all competing events.[2] If the predicted risk of the adverse event of interest occurring within a certain time span exceeds a pre-determined level, an intervention is initiated. Otherwise, the patient remains to be followed and a new biomarker measurement is scheduled. We incorporate the patient's longitudinal history in the scheduling process by imposing a limit to the acceptable predicted risk of the event of interest occurring before the new measurement. We examined the effectiveness of this strategy versus fixed schedules in a simulation study. Preliminary results using simulations based on the heart failure cohort showed that this approach leads to fewer measurements and more accurate risk estimates. Our study demonstrates the benefits of integrating patient-specific information into scheduling decisions for surveillance programs with multiple biomarkers and competing outcomes. This approach can contribute to more timely interventions, and thus, better outcomes, in clinical care and more efficient use of resources.

[1] Rizopoulos, D., Taylor, J. M., Van Rosmalen, J., Steyerberg, E. W., & Takkenberg, J. J. (2016). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. Biostatistics, 17(1), 149-164.
[2] Andrinopoulou, E. R., Rizopoulos, D., Takkenberg, J. J., & Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. Statistical methods in medical research, 26(4), 1787-1801.

TP56 The clivus reconstruction: a novel method for identifying the optimal set of scaffolds

Vezzoli M.*[1], Sandri M.[1], Renzetti S.[1], Doglietto F.[2], Calza S.[1]
[1]University of Brescia ~ Brescia ~ Italy, [2]Catholic University School of Medicine ~ Rome ~ Italy

In the last two decades, endoscopic transnasal surgery has revolutionized the treatment of sinonasal and skull base diseases by providing access to deep regions of the skull base, such as the clivus, with minimal brain retraction and cranial nerve manipulation. However, skull base reconstruction remains a challenge [1], which can be addressed through 3D printing of patient-specific scaffolds. Identifying a limited number of clivus variations could offer an alternative to creating tailored scaffolds, reducing production costs and waiting times. This study aims to propose a new method to identify a small number of adaptable clivus prototypes that require minimal reduction of their shape measures. Surgeons can quickly adjust these prototypes in the operatory room by reducing key measurements by a maximum of 1 mm. Four clivus measurements identified by neurosurgeons as determinant of its shape were evaluated on 163 MR Images (MRI) belonging to Italian adults with unknown age and sex collected at Ospedale Civile of Brescia (Italy). These measures are distributed as a multivariate normal random variable. Our interest lies in the hyper-ellipsoid defined by the multivariate normal distribution with mean and covariance matrix estimated from the sample, which encloses 95% of the data. For computational efficiency, this hyper-ellipsoid was approximated by a set of points on a regular grid with a step size of 0.15 mm. The clivus prototypes were represented by hyper-cubes with side 1 mm and centres ci. We used the Hooke-Jeeves algorithm [2], an optimization procedure designed for non-smooth functions, to calculate the optimal placement of k hyper-cubes for maximum coverage of the hyper-ellipsoid. The number of prototypes k was determined based on ensuring a coverage of the hyper-ellipsoid greater than 99%. For each patient, a single prototype is assigned, chosen to minimize the amount of reduction across the four dimensions required by the surgeon in the operatory room. The described procedure has the potential to be easily extended to other bone prototypes and could serve as a valuable tool for biomedical engineers in the design of prostheses.

[1] Mattavelli, D. et al. Additive Manufacturing for Personalized Skull Base Reconstruction in Endoscopic Transclival Surgery: A Proof-of-Concept Study. World Neurosurg 155, e439–e452 (2021).
[2] Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). Introduction to derivative-free optimization. Society for Industrial and Applied Mathematics.

TP57 Enhancing patient outcomes and statistical efficiency in rare-disease phase- ii trials: the stratosphere study

Deliu N.*[1], Jones R.J.[2], Toshner M.[3], Villar S.S.[4]
[1]MRC – Biostatistical Unit, University of Cambridge; MEMOTEF, University of Rome La Sapienza ~ Rome ~ Italy, [2]Victor Phillip Dahdaleh Heart and Lung Research Institute (HLRI) ~ Cambridge ~ United Kingdom, [3]Victor Phillip Dahdaleh Heart and Lung Research Institute (HLRI); Royal Papworth Hospital ~ Cambridge ~ United Kingdom, [4]MRC – Biostatistical Unit, University of Cambridge ~ Cambridge ~ United Kingdom

Designing early- and late-phase clinical trials for rare conditions pose unique challenges, ranging from ethical and practical enrolment considerations to generating robust evidence from limited sample sizes. Pulmonary arterial hypertension (PAH), for example, is a life-limiting progressive disorder characterised by high blood pressure in the arteries of the lungs that affects 15 to 50 cases per million in the US and Europe. Although treatable, PAH has no available cure. In such a setting, conducting trials to primarily learn about treatment effectiveness (as in traditional fixed randomised designs) may result in an unacceptably low expected outcome for patients in presence of superior treatments. Further, meeting conventional power requirements, while controlling for type-I error, is infeasible in inherently small and heterogeneous populations. StratosPHere 2 is the first-ever precision-medicine trial of treatments targeting common genetic causes of the disease: BMPR2 mutations [1]. Using this case study, this work outlines the potential of innovative designs for addressing the above challenges in small populations. The proposed design builds on a response-adaptive randomised framework and incorporates Bayesian principles for utilising the continuously accrued responses (transcriptomic biomarker expressions) to adapt the allocation probabilities [2]. It involves three stages, each of which enrolling and allocating from 6 to 8 patients to one of the three study arms (two active and one control arm). The adaptive design is stratified by two BMPR2 mutation subgroups and performed independently in each stratum. In addition to varying the allocation probabilities from one stage to another, the design is allowed to drop one of the two active arms in case of early evidence of inferiority. Extensive simulations, using prior PAH patient data, are carried out to perform sample size evaluation with 20 patients per mutation stratum. Compared to a traditional fixed design, the flexibility of the proposed framework results in substantial gains in both statistical power and a higher chance for patients to receive the superior arm. The use of innovative adaptive designs holds great potential for assessing treatment effects in rare disease settings, achieving a near-optimal balance between study patient benefits and statistical guarantees.

[1] Deliu, N., Mark, T., Jones, R. J., Villar, S. S., et al. StratosPHere 2: A response-adaptive randomised placebo-controlled Phase II trial to evaluate hydroxychloroquine and phenylbutyrate in pulmonary arterial hypertension caused by mutations in BMPR2. Study protocol in submission to Trials.
[2] Wason, J., Trippa, L. A comparison of Bayesian adaptive randomization and multi-stage designs for multi-arm clinical trials. Trials 14 (Suppl 1), P40 (2013). https://doi.org/10.1186/1745-6215-14-S1-P40

WEDNESDAY 30 AUGUST 2023

POSTER SESSIONS 3

WP1	Giulia Capitoli	*Image segmentation via bayesian mixtures to detect latent spatial patterns in mass spectrometry imaging*
WP2	Yefeng Fan	*A novel approach to modelling weighted fmri networks using a bayesian weightedexponential random graph model*
WP3	Thibaut Galvain	*Anticoagulants in patients with atrial fibrillation and risk of falls: a bayesian network meta-analysis.*
WP4	Laura Quinn	*Re-analysing frequentist systematic reviews using bayesian methods*
WP5	Janharpreet Singh	*Multi-indication meta-analysis for sharing of information on treatment effectiveness across cancer subtypes*
WP6	Valeria Vitelli	*Rank-based bayesian joint variable selection and clustering of genome-wide transcriptomic data from pan-cancer*
WP7	Miriam Isola	*Circulating extracellular vesicles as biomarkers for covid-19 severity*
WP8	Nikola Dordevic	*Influence of hormonal contraceptives on plasma metabolite concentrations in a large cohort*
WP9	Mi Mi Ko	*Association of serum resistin with blood stasis syndrome in korean medicine for metabolic diseases*
WP10	Vittorio Simeon	*Prognostic value of inflammatory indexes in patients with hepatocellular carcinoma: results from the ita.li.Ca*
WP11	Anna Tkachev	*Shotgun lipidomics for the development of a diagnostic test for schizophrenia*
WP12	Md. Abdul Basit	*A calibrated sensitivity model for observational studies with multivalued treatments*
WP13	Matthew Boyton	*Genetically proxied effects of plasma proteins on risk of ms identified via mendelian randomization*
WP14	Florie Brion Bouvier	*A comparison on real data of methods to develop individualized treatment rules*
WP15	Kate Edgar	*Transporting a 'hospital at home' rct treatment effect to a target population with differing delirium measures*
WP16	Teresa Greco	*Spasticity in patients with multiple sclerosis: a bayesian network application*
WP17	Gustav Jonzon	*Causaloptim: a user-friendly r-package for computation of symbolic sharp bounds on counterfactual queries*
WP18	Andriana Kostouraki	*A note on the latent unconfoundedness assumption for hierarchical observational data: a simulation study*
WP19	Emanuele Koumantakis	*Tyrosine kinase inhibitors discontinuation in chronic myeloid leukemia: a retrospective cohort study*
WP20	Kim Luijken	*Tell me what you want, what you really really want: estimands in observational pharmacoepidemiologic studies*
WP21	Leah Pirondini	*Handling informative patient monitoring in routinely-collected data used to estimate treatment effects*
WP22	Ilaria Prosepe	*Estimating the effect of treatment delay: g-computation versus clone-censor-reweight approach*
WP23	Priom Saha	*Sensitivity analysis with matched pairs from observational studies*
WP24	Matthew Smith	*On causal inference for the relative survival setting*
WP25	Ellie Van Vogt	*Missing data handling and model-agnostic tests for treatment effect heterogeneity in randomised trials*
WP26	Alessandra Carobbio	*Multistate models for the analysis of time-to-event data: the case of myeloproliferative neoplasms*
WP27	Thibaud Charrier	*Increased cardiac risk after a second malignant neoplasm among childhood cancer survivors, a fccss study*
WP28	Maarten Coemans	*A semi-parametric, non-proportional competing risks model predicting outcomes after kidney transplantation*
WP29	Stacy Cyrille	*Analyzing duration of response in phase ii oncology trials*
WP30	Annalisa Orenti	*A multi-state model evaluating the association of oxygen therapy with the course of cystic fibrosis in europe*
WP31	Ivonne Solis-Trapala	*Graphical and multistate modelling of home dialysis uptake in patients who need kidney replacement therapy*
WP32	Amarit Tansawet	*Sglt2i better slow down chronic kidney disease progression in type 2 diabetes patients: a multistate analysis.*
WP33	Federica Bellerba	*Sars-cov-2 trends in italy, germany and school opening during the omicron variant: a quasi-experimental study*
WP34	Anikó Lovik	*Mental health of family members and friends of covid-19 patients: an observational cohort study*
WP35	Andrea Nova	*Life expectancy changes and covid-19 vaccination campaign impact on all-cause mortality in italy*
WP36	Pascal Roy	*Agent-based models for the covid-19 epidemic: understanding the effects of determinants of disease spread*
WP37	M. Iftakhar Alam	*A combined criterion for dose finding in phase i clinical trials*
WP38	Abeer Althobety	*High-order asymptotic intervals for the toxicity probability in phase i clinical trials*
WP39	Drifa Belhadi	*Bayesian decision analysis for clinical trial design with binary outcome: illustration for ebola virus disease*
WP40	Laura Etfer	*Optimal allocation of clusters to sequences in stepped wedge trials with binary outcome data*
WP41	Anneke Grobler	*Trial estimands: using the composite strategy for intercurrent events when the outcome variable is continuous*
WP42	Xiaoran Lai	*Efficient dose insertion in phase 1 trials: an adaptive approach using model-based design*
WP43	Wenyue Li	*Incorporation of historical data for basket trials in the early phase*
WP44	Federico Rotolo	*Joint modeling of pharmacodynamic biomarker and safety in ph1 clinical trials in oncology: sanofi experience*
WP45	Keiichiro Seno	*Active level set estimation in phase i dose-finding clinical trials*
WP46	Katie Stocking	*The use of basket trials in gynaecology*
WP47	Dominic Stringer	*Why do mental health trials only analyse a single primary outcome? We can do better.*
WP48	Amin Yarahmadi	*On controlling the type i error rate in an interrupted group sequential trial with interim analyses*
WP49	John Andrew	*Longitudinal functional data analysis with applications in biomechanics*
WP50	Francesca Graziano	*Growth mixture modelling to identify sodium patterns. Application to icu data in traumatic brain injury (tbi)*
WP51	Mélanie Guhl	*Uncertainty computation at finite distance in nonlinear mixed models: evaluation of a new bayesian method*
WP52	Ester Luconi	*Time series analysis on historical death data of milan from the milan registers (1452-1845)*
WP53	Giuseppe Occhino	*The efficacy of a motivational interview in heart failure patients and their caregivers: a dyadic analysis*
WP54	Chiara Oriecuia	*Meta-analysis of patient-reported outcomes: methodological proposal and application to rcts in heart failure*
WP55	Pierre-Emmanuel Poulet	*Modeling a disease-modifying treatment with a piecewise geodesic mixed-effect model*
WP56	Alma Revers	*Safety signal detection in biochemical measures in randomized clinical trials, a bayesian mixed effect model*
WP57	Inês Sousa	*Joint model for multiple longitudinal responses with informative time measurements*
WP58	Alexandre Bailly	*Multimodal prediction of echocardiography prescription*
WP59	Alan Balendran	*Diverse concepts of machine learning robustness in healthcare: a scoping review*
WP60	Gregor Buch	*A pragmatic regularization strategy for collinearity-tolerant selection of nonlinear relations*

WEDNESDAY 30 AUGUST 2023

WP61	Chiara Chiavenna	*Predicting attitudes towards covid-19 vaccination for children: a machine learning-based approach*
WP62	Aurora Gaeta	*Breast lesion malignancy prediction using machine learning and deep learning approaches on ultrasound images*
WP63	Rana jreich	*Evaluation of eligibility criteria impact on study outcome and patient count*
WP64	Naser Kamyari	*Prediction of polypharmacy in half a million adults iranian population: comparing machine learning algorithms*
WP65	Lan Kelly	*Signal detection in medical devices in spontaneous reports using natural language processing*
WP66	Corrado Lanera	*Deep learning-based prediction of major arrhythmic events in dilated cardiomyopathy*
WP67	Federica Luppino	*Demag predicts the effects of variants in actionable genes with structural and evolutionary features*
WP68	Wanchana Ponthongmak	*Detection of recurrent cancer from emr using natural language processing: a systematic review*
WP69	Nanaka Seiwa	*Outlier detection for tree-structured model in regression problem*
WP70	Vittorio Torri	*Weakly-supervised classification of clinical documents: a case study on italian discharge letters*
WP71	Rossella Miglio	*Machine learning regression models for predicting hospital stay for general medicine-specific patients*
WP72	Elena Albu	*Missforest v2 – missing data imputation for prediction settings*
WP73	Anca Chis Ster	*Evaluating bias in causal mediation effects with non-adherence and missing data: better simulate than never*
WP74	Imad EL BADISY	*Comparison of imputation methods in the presence of time-varying and non-linear covariates effects*
WP75	Minna Genbäck	*Sensitivity bounds in multilevel modeling of longitudinal data with data missing not at random*
WP76	Tetiana Gorbach	*Practical approach for missing data sensitivity analyses in joint modelling of cognition and dementia risk*
WP77	Alfred Kipyegon Keter	*Bayesian latent class analysis correcting for verification and reference standard bias in tb prevalence*
WP78	Rheanna Mainzer	*The handling of missing data with multiple imputation in observational studies that address causal questions*
WP79	Rocío Aznar-Gimeno	*Min-max-median/iqr approach. Comparison with min-max, logistic regression and xgboost.*
WP80	Rosanna Comoretto	*Mortality prediction: one size may not fit all. The italian experience on the validity of the pim 3 score*
WP81	Paula Dhiman	*Reporting quality of protocols for studies developing and validating a prediction model: a systematic review*
WP82	Paula Dhiman	*A systematic review of sample size calculation for machine learning prediction model studies in oncology*
WP83	Shaun Hiu	*Baseline prediction model of time-to-flare in rheumatoid arthritis after dmard cessation: the bio-flare study*
WP84	Alexandra Hunt	*A systematic review of prediction models for transition to psychosis in individuals meeting arms criteria*
WP85	Sadhana Kannan	*Clinical utility of predcit a prognostic tool in breast cancer: decision curve analysis*
WP86	Dariush Khaleghi Hashemian	*Building centile charts of ecg parameters in children from 0 to 16 years old: an italian cross-sectional study*
WP87	Ashok Krishnamurthy	*Uncertainty quantification in the predictions of a ebola outbreak using bayesian data assimilation*
WP88	Zewen Lu	*Do researchers consider the time-dependent interplay between time to assessment and severity in acute stroke?*
WP89	Thitiya Lukkunaprasit	*Individual treatment effects of sodium-glucose cotransporter-2 inhibitors on chronic kidney disease risk*
WP90	Ryunosuke Machida	*Predicting number of events using joint model in clinical trials with a time-to-event endpoint*

WP91	Kazumi Omata	*Model predictions of the effects of rapid art on hiv and aids*
WP92	Carmen Petitjean	*Quantifying the added predictive value of prs in cvd risk prediction tools in individuals with morbidity*
WP93	Sergio Sabroso-Lasa	*How correlations between markers influence the net benefit increase of a predicitve model*
WP94	Simon Schwab	*Development of a risk calculator and web application for kidney transplantation with r/shiny*
WP95	Daniel Stahl	*Impact of sample size and events per predictors on performance & stability of psychosis risk prediction models*
WP96	Persefoni Talimtzi	*Evaluation of systematic reviews of prognostic models for covid-19: an overview of systematic reviews.*
WP97	Junfeng Wang	*Evaluating calibration at moderate-strong level using patient subgroups identified with clustering analysis*
WP98	Junfeng Wang	*Assessing the reporting of image and ai based cvd diagnostic models: evaluation and implementation of claim*
WP99	Calin Avram	*The chi square test and the need to apply corrections for big data*
WP100	Giulia Barbati	*Evaluation of the effectiveness of pcsk9-i in a target trial emulation framework based on ehr*
WP101	Lauren Coan	*Spatial statistical multilevel modelling of optic nerve deformation in multiple disease groups of glaucoma*
WP102	Elsa COZ	*Quantifying the risk of treatment-related adverse events in oncology: data, issues and models*
WP103	Mireya Diaz	*Balance of time-dependent covariates in real-world data via the fréchet distance*
WP104	Arianna Galotta	*Big data and statistical inferences problems: is effect size measure an alternative to p-value?*
WP105	Giota Touloumi	*Projections of cardiovascular disease deaths using a microsimulation model in the absence of cohort data*
WP106	Cinzia Maria Papappicco	*A new statistical index for evaluating variability in physical state index during pediatric anesthesia.*
WP107	Sukanya Siriyotha	*Quality of life in patients after 1-year percutaneous coronary intervention: real-world data analysis*
WP108	Philippe Wagner	*"Doing the doctors work" - usefulness of a simple prediction model applied to health register data in sweden*
WP109	Eva Andersson	*Have air pollution levels decreased more in high income areas – a 19-year follow-up in gothenburg, sweden*
WP110	Susana Diaz Coto	*Comparison of clinic versus home devices for spirometry assessment*
WP111	Dariush Ghasemi	*Systematic mediation analysis of genetic loci associated with kidney function in a population-based study*
WP112	Timothy Grant	*Using autoregression to model novel viral outbreaks in human populations to manage medical resources.*
WP113	Péter Hársfalvi	*Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test*
WP114	Donghwan Lee	*Sensitivity analysis for multiple exposure effects in generalized linear models with unmeasured confounders*
WP115	Evangelos Kritsotakis	*Analytic error in national-level prevalence surveys of healthcare-associated infections: a systematic review*
WP116	Farzana Osman	*Clinically versus cytokine-defined genital inflammation and the risk of hiv infection*
WP117	Luigi Palla	*Diurnal eating patterns and their effect on body mass index in the italian population (inran-scai 2005-2006).*
WP118	Jenő Reiczigel	*Logistic regression with covariate-dependent probability of misclassification*
WP119	Davide Bernasconi	*Comparison of propensity score methods with multiple treatments: simulations and clinical application*
WP120	Mirko Bonetti	*30 Day mortality in hospitalised covid-19 patients: a retrospective study for the province of bolzano (italy)*

WP121	Eirini Chrysanthou	*Importance of sex-stratification in biomarker identification of early-stage melanoma survival analyses*
WP122	Ömer Faruk Dadaş	*Comparison of statistical methods for survival data with time-dependent covariates and competing risks*
WP123	Paola De Lorenzo	*Analysis of treatment outcome in an observational study in pediatric leukemia with a propensity score approach*
WP124	Patrick Djidel	*Assessment of impact on the survival outcome due to disclosure of immature survival data in oncology trials*
WP125	Luca Genetti	*Efficient estimation of the marginal mean of recurrent events in randomized clinical trials*
WP126	Ryusei Kimura	*Comparison of asymptotic and re-randomization tests under non-proportional hazards scenarios with minimization*
WP127	Jun Ma	*Stratified cox models under partly interval censoring*
WP128	Tomomi Nishikawa	*The performance of overlap weighting ps method for survival analysis when interaction in a subgroup exist*
WP129	Eleonora Pagan	*Evaluating surrogate endpoints for overall survival in rcts testing immune checkpoint inhibitors*
WP130	Matteo Petrosino	*Outcome prognostication in acute brain injury using the neurological pupil index (orange) study.*
WP131	Dikshyanta Rana	*Impact of censoring mechanisms in assessment of prognostic technologies*
WP132	Tomohiro Shinozaki	*Pairwise cox modeling approach for causally interpretable average hazard ratio under nonproportional hazards*
WP133	Achilleas Stamoulopoulos	*Joint modelling of longitudinal biomarkers and of risk of serious non-aids event in people living with hiv*
WP134	Lubomir Stepanek	*Adjusted kaplan-meier estimate for prediction of a decrease of covid-19 antibodies below laboratory cut-off*
WP135	Sook-Young Woo	*Estimation of the cutoff value for continuous prognostic factors in survival data with competing risk*
WP136	Asanao Shimokawa	*Construction of survival tree based on restricted mean survival time*

WP1 Image segmentation via bayesian mixtures to detect latent spatial patterns in mass spectrometry imaging

Capitoli G.*[3], Colombara S.[2], Cotroneo A.[2], De Caro F.[2], Morandi R.[2], Schembri C.[2], Zapiola A.G.[2], Denti V.[3], Smith A.[3], Denti F.[1]
[1]Department of Statistics, Università Cattolica del Sacro Cuore ~ Milano ~ Italy, [2]Department of Mathematics, Politecnico of Milano ~ Milano ~ Italy, [3]Department of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy

Mass spectrometry records molecular mass abundance for a broad set of different molecules e.g., lipids, peptides, and glycan, given a sample of a specific biological tissue. In particular, the MALDI-MSI technique produces imaging data where, for each pixel, a mass spectrum is recorded for each molecule. However, the standard statistical methods, do not fully address the pixel spatial dependencies and the networks of different molecules. Here, we investigate the use of Bayesian mixture models to segment these multilayer biomedical images with the aim to detect groups of pixels that present similar patterns to extract interesting insights, such as anomalies that one cannot capture from the morphological original tissue. To perform model-based clustering, we compare the standard K-means, the Gaussian mixture model (GMM), which does not consider the spatial correlation between neighbouring pixels, and the Hidden Potts Model (HPM), which uses a hidden Markov random field to account for the spatial nature of the dataset. Given the large dimensionality of the data, as the first step, we investigate a one-dimensional approach considering only the first functional principal component scores (fPCs) value for each pixel. The shape of the distribution of fPCs suggests using GMM to segment the pixels of the MALDI-MSI image. Second, in order to integrate the different molecular layers, we extend the likelihood specification to handle multivariate measurements extending the Gibbs sampler implemented by Moores et al. [1] to the multivariate case for both GMM and HPM. Moreover, with HPM, we enhance the separation of clusters by adding the spatial information, and the number of pixels flagged as uncertain decrease from K-means and GMM to HPM. Based on the proposed approach, we are able to integrate data of various types of molecules extracted from the same tissue, highlighting uncover hidden patterns resulting from the spatial correlation among pixels and the relationship between multiple molecular layers.

[1] T. Moores, D. Feng, and K. Mengersen. bayesImageS: Bayesian Methods for Image Segmentation using a Potts Model. R package (v0.6-1), 2021.

WP2 A novel approach to modelling weighted fmri networks using a bayesian weightedexponential random graph model

Fan Y.*, White S.
University of Cambridge ~ Cambridge ~ United Kingdom

The Exponential Random Graph Model (ERGM) is a statistical model commonly used for inference on networks. The inference is typically performed on a single bi-nary network. We proposed a new method Multi-network Multi-layered Bayesian Exponential Random Graph Model that extends inference in two areas: weighted edges and multiple networks. These extensions are necessary to model weighted functional magnetic resonance imaging (fMRI) correlation networks from multiple individuals. A binary network has edges that can either be present or absent, whereas in a weighted network, edges are associated with weight values; as appropriate for correlation-derived edges. The standard ERGM framework can only be fitted on an individual network, while inference is more often required across groups of individuals. Furthermore, the standard ERGM framework is unable to fit on weighted networks, inspiring the need for frameworks to model fMRI correlation networks. Our Multi-network Multi-layered Bayesian Exponential Random Graph Model adopts a multi-layered framework such that a weighted network is considered as layers of binary networks each associated with an increasing threshold across the domain of edge weights, and the transition processes across different layers are modelled as an ERGM. This multi-layer structure is further fitted into a Bayesian hierarchical structure to enable the joint modelling of multiple networks, enabling an analysis of group- level network topological features. We can then model brain function across multiple individuals using fMRI-derived correlation networks (brain regions are represented as vertices and the activation between those regions are represented as edges). We will present an analysis of simulated fMRI networks based on the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study to illustrate the model performance. We have implemented this framework on a set of simulated fMRI networks. The simulations are generated from a specifically designed on-set combination such that we understand the underlying 'truth' of the simulated networks. The inference of the Multi-network Multi-layered Bayesian Exponential Random Graph Model when fitting on the simulated networks agrees with the 'truth' of the simulation both on the individual network level as well as on the group level. Lehmann B.C.L, Henson R.N., Cam-CAN Geerligs L, and White S.R. "Characterising group-level brain connectivity: a framework using Bayesian exponential random graph models". In: bioRxiv (2019). DOI: 10 . 1101 / 665398. eprint: https : / / www . biorxiv.org/content/early/2019/06/10/665398.full.pdf. URL: https://www.biorxiv.org/content/early/2019/06/10/665398.
Caimo A and Gollini I. A multilayer exponential random graph modelling approach for weighted networks. 2018. DOI: 10.48550/ARXIV.1811.07025. URL: https: //arxiv.org/abs/1811.07025.

WP3 Anticoagulants in patients with atrial fibrillation and risk of falls: a bayesian network meta-analysis.

Galvain T.*[2], Ruaraidh H.[1], Donegan S.[3], Wilkinson R.[2], Lip G.[4], Czanner G.[2]
[1]Liverpool Reviews And Implementation Group, Health Data Science, University of Liverpool and The Royal Liverpool and Broadgreen University Hospitals, Liverpool Health Partners ~ Liverpool ~ United Kingdom, [2]School of Computer Science and Mathematics, Liverpool John Moores University ~ Liverpool ~ United Kingdom, [3]Department of Health Data Science, University of Liverpool ~ Liverpool ~ United Kingdom, [4]Liverpool Centre for Cardiovascular Science, Liverpool Heart and Chest Hospital ~ Liverpool ~ United Kingdom

Direct oral anticoagulants (DOACs) are now the preferred treatment option over Vitamin-K antagonists (VKA) in patients with atrial fibrillation (AF).1,2 However, AF patients at risk of falls or with history of falls often do not receive anticoagulants due to the risk of bleeding. This study intends to assess the efficacy and safety of DOACs in patients with AF and history of falls or risk of falls. A systematic literature review was conducted until December 2022 in CENTRAL, CINAHL, ClinicalTrials.gov, EMBASE, MEDLINE, Scopus and Web of Science to identify studies evaluating safety and efficacy comparing VKA to DOACs in patients with AF and history or risk of falls. Primary outcomes were stroke/systemic embolism (SSE) and major bleeding (MB). Two reviewers identified studies, extracted data and assessed the risk of bias. Bayesian network meta-analyses were conducted. Hazard ratios (HRs) and 95% credible intervals (CrIs) were used to assess the effect of drugs while the cumulative ranking curves (SUCRA) were used to reflect their hierarchy. 887 articles were identified after removing duplicates. 160 were screened for full text and 10 articles were retained for final quantitative synthesis. Risk of bias was moderate to serious in included studies. Apixaban (HR: 0.72, 95% CrI: 0.59–0.96) and Rivaroxaban (0.80, 0.63–0.99) reduced the risk of SSE compared with VKA while Dagibatran (0.92, 0.68-1.30) and Edoxaban (0.95, 0.48-1.90) did not. Apixaban was ranked first (SUCRA 0.87), followed by Rivaroxaban (0.68), Edoxaban (0.38), Dagibatran (0.37) and VKA (0.19). In reducing the risk of MB, Edoxaban was found to be the best DOAC compared to VKA (HR, 0.66; 95% CI: 0.49–0.94; SUCRA 0.85) followed by Apixaban (0.67; 0.54–0.87; 0.83), Dagibatran (0.79; 0.63–1.04; 0.54), VKA (SUCRA 0.16) and Rivaroxaban (1.02; 0.83–1.28; 0.12). DOACs have different efficacy and safety profiles. Apixaban should be the preferred treatment as it showed the best efficacy and excellent safety in patients with AF and history or risk of falls. Due to limited sample size and limits in designs, future trials comparing directly DOACs are needed in order to provide conclusive results.

1. Hindricks G, Potpara T, Dagres N, et al.: 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). European Heart Journal 2021; 42:373–498.
2. January CT, Wann LS, Calkins H, et al.: 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society of Thoracic Surgeons. Circulation [Internet] 2019 [cited 2022 Aug 2]; 140. Available from: https://www.ahajournals.org/doi/10.1161/CIR.0000000000000665

Poster Sessions

WP4 RE-ANALYSING FREQUENTIST SYSTEMATIC REVIEWS USING BAYESIAN METHODS

Quinn L.*[1], Veenith T.[2], Bion J.[1], Hemming K.[1], Whitehouse T.[2], Lilford R.[1]
[1]University of Birmingham ~ Birmingham ~ United Kingdom, [2]Queen Elizabeth Hospital Birmingham ~ Birmingham ~ United Kingdom

Systematic reviews and meta-analyses of randomised controlled trials (RCTs) are typically performed using a frequentist approach. Although confidence intervals are reported, interpretation is usually based on statistical significance alone. For example, a review of RCTs comparing early versus late tracheostomy on clinical outcomes concluded that early tracheostomy reduces the length of intensive care unit stay and duration of mechanical ventilation, but does not reduce short-term mortality or ventilator-associated pneumonia (VAP) [1]. We report results from a case study, updating the review and re-analysing the results using Bayesian methods, to illustrate how a Bayesian approach can provide a more robust basis for clinical decisions [2]. The results from the systematic review and meta-analysis, quantifying the effect of early versus late tracheostomy on clinical outcomes, were updated and re-analysed using Bayesian methods with uninformative priors. Risk ratios (RRs) and standardised mean differences (SMDs) were calculated with 95% confidence intervals for the frequentist approach and 95% credible intervals for the Bayesian approach. Posterior probabilities were also reported for any benefit (RR<1;SMD<0), a small benefit (number needed to treat≤200;SMD<0.5), or modest benefit (number needed to treat≤100,SMD<1). 19 RCTs with 3,508 patients were included. For 16 RCTs, VAP was measured as a clinical outcome. From the frequentist approach, the risk of VAP was 10% lower (RR=0.90, 95% confidence intervals:0.78 to 1.02) for early compared to late tracheostomy. From the Bayesian approach, the risk of VAP was 11% lower (RR=0.89, 95% credible intervals:0.73 to 1.05). In addition, the posterior probabilities for any benefit (RR<1) was 94%, a small beneficial benefit (number needed to treat ≤200) was 78% and a modest beneficial effect (number needed to treat ≤100) was 51%. For VAP, which had a non-statistically significant results from the frequentist approach, the posterior probability for any benefit was high at 94% but dropped to 78% for a small benefit and 51% for a modest benefit. Bayesian re-analysis of the systematic review and meta-analysis allows for clinicians to have more robust evidence for decision-making and suggests a high probability that early tracheostomy compared with a delayed tracheostomy has at least some benefit for all clinical outcomes considered.

1. Deng, H., et al., Early versus late tracheotomy in ICU patients: A meta-analysis of randomized controlled trials. Medicine, 2021. 100(3).
2. Quinn, L., et al., Bayesian analysis of a systematic review of early versus late tracheostomy in ICU patients. British Journal of Anaesthesia, 2022.

WP5 Multi-indication meta-analysis for sharing of information on treatment effectiveness across cancer subtypes

Singh J.*[2], Anwer S.[1], Dias S.[1], Palmer S.[3], Saramago P.[3], Thomas A.[4], Soares M.[3], Bujkiewicz S.[2]
[1]Centre for Reviews and Dissemination, University of York ~ York ~ United Kingdom, [2]Biostatistics Research Group, Department of Population Health Sciences, University of Leicester ~ Leicester ~ United Kingdom, [3]Centre for Health Economics, University of York ~ York ~ United Kingdom, [4]Leicester Cancer Research Centre, University of Leicester ~ Leicester ~ United Kingdom

It is increasingly common for oncology drugs which initially received licensing approval for a particular patient population (indication) to subsequently have their licenses extended to include additional indications. For example, Bevacizumab (Avastin) first received licensing approval as a treatment for metastatic colorectal cancer but has since received approval for six other cancer subtypes. From an evidence synthesis perspective, focusing on each indication separately can lead to large uncertainty in treatment effect estimates despite the availability of substantial evidence for similar indications. We explore meta-analysis methods for sharing of information across indications to make more efficient use of evidence on treatment effectiveness, using trial-level data on the effectiveness of Bevacizumab for different cancer subtypes as a case-study. A Bayesian hierarchical model proposed by Hemming et al [1], consisting of a within-study level, a between-studies level, and a between-indications level, was applied to perform a meta-analysis of summary data on Overall Survival from randomised controlled trials assessing the effectiveness of Bevacizumab versus comparator cancer treatments. A robust version of the model, based on the approach proposed by Neuenschwander et al [2], which assumes that the indication-level estimates are either exchangeable or independent with one another (i.e., assuming partial exchangeability via mixture prior distributions) was also applied to the data. The pooled treatment effect estimate from applying a random-effects meta-analysis to data from five non-small cell lung cancer (NSCLC) trials was log hazard ratio (HR) = -0.09 (95% CrI: -0.32, 0.16). However, there was an improvement in uncertainty when applying the mixture hierarchical model to data from 29 trials across all indications (i.e., to borrow information from the other indications), where the NSCLC effect estimate was log HR =-0.11 (-0.23, 0.01). Bayesian hierarchical methods can provide a flexible approach to sharing of information across indications, whilst accounting for between-indications heterogeneity, which can lead to improvements in precision. Mixture hierarchical methods can also provide robustness in syntheses where exchangeability does not hold for all indications.

[1] Hemming K, Bowater RJ, Lilford RJ. Pooling systematic reviews of systematic reviews: a Bayesian panoramic meta-analysis. Statistics in Medicine. 2012 Feb 10;31(3):201-16.
[2] Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceutical statistics. 2016 Mar;15(2):123-34.

Poster Sessions

WP6 Rank-based bayesian joint variable selection and clustering of genome-wide transcriptomic data from pan-cancer

Eliseussen E., Muggerud H., Scheel I., Vitelli V.*
University of Oslo ~ Oslo ~ Norway

The use of ranks in genomics is naturally linked to the underlying biological question, since one is often interested in overly-expressed genes in a given pathology. When aiming at analysing transcriptomic patient data for cancer subtype discovery, we have already successfully proposed to use a mixture model-based clustering approach to estimate the cancer subtype based on a Bayesian Mallows rank Model (BMM) [1]. The model is able to process heterogeneous patient data (from different studies / cohorts and different cancers / tissue-of-origin), and to both produce estimates of the consensus ranking of the genes shared among samples in the same cluster, and to fill-in missing data via data augmentation strategies. In the context of patient subtyping, posterior distributions of the unknowns are particularly relevant, since these can provide an evaluation of the uncertainty (and thus reliability) associated to the estimates. However, BMM relies on pre-selecting a reasonable number of genes (e.g., 1000) to be used in the analysis. A lower-dimensional version of BMM (lowBMM) that scales to genome-wide transcriptomic data analysis has also been proposed and used in the context of cancer genomics [2], but no clustering approach was available in this proposal. We propose to perform genome-wide clustering of transcriptomic patient data from a pan-cancer analysis (i.e., combining several cancer types) via a Bayesian mixture of Mallows models that combines clustering via BMM and lowBMM. The model is capable of jointly performing clustering and variable selection, thus successfully selecting the genes best representing the structural patterns of expression characterising each subtype. We study the performance of the novel clustering method in the context of simulations, and we show the results obtained in a pan-cancer application of the model. Genome-wide Bayesian clustering outperforms competitive methods in complex settings (e.g., pan- cancer analyses). The model implementation is efficient as inference is possible genome-wide and on large cohorts; the model does not require any gene preselection, thus improving the reliability and reproducibility of Bayesian clustering; and finally, the model also provides proper uncertainty quantification for all unknowns, which is a clear advantage for personalized medicine.

[1] V. Vitelli, T. Fleischer, J. Ankill, E. Arjas, A. Frigessi, V. N. Kristensen, Zucknick, M. Transcriptomic pan-cancer analysis using rank-based Bayesian inference. Molecular Oncology, 17(4), 2022, 548-563.
[2] E. Eliseussen, T. Fleischer, V. Vitelli. Rank-based Bayesian variable selection for genome-wide transcriptomic analyses. Statistics in Medicine, 41(23), 2022, 4532-4553.

WP7 Circulating extracellular vesicles as biomarkers for covid-19 severity

De Martino M., Beltrami A.P., Caponnetto F., Isola M.*
University of Udine ~ Udine ~ Italy

Assessing biomarkers for COVID-19 outcome is crucial to improve clinical practice in this field. Extracellular vesicles (EVs), found in body fluids, are bioactive nanoparticles released by various type of cells. The aim of this research is investigating the role of circulating EVs in SARS-CoV-2 infection, applying a more classical model, Elastic Net Logistic regression (ENL), and a supervised machine learning model, Random Forest (RF). The study includes 146 patients (≥ 18 years) attending the Infectious Disease Department with a diagnosis of COVID-19. The outcome is defined as death or need for intubation. Clinical, hematological and extracellular vesicles data were collected. Both ELN and RF are trained on a training set and the performance is evaluated on a test set with a 70:30 split ratio. In the case of ELN a regression coefficient beta is estimated, where a positive beta denotes that the event is more likely to happen. On the hand, for the RF model, a feature importance score (FIS) is evaluated, where a higher value corresponds to a higher importance of the feature. Main factors identified by ELN are small fraction of EV expressing CD31 (beta=-0.50), CD140b (beta=-0.26) and CD42b (beta=-0.11); IL6c (beta=0.44), RDW (beta=0.28) and lymphocyte (beta=-0.25). Main factors identified by RF are large fraction of EV expressing CD45 (FIS= 0.10); small fraction of EV expressing CD140b (FIS=0.07), CD42b (FIS=0.06) and CD31 (FIS=0.05); NT-proBNP (FIS=0.08), IL6c (FIS=0.07) and proADM (FIS=0.07). Accuracy evaluated on the test set is 81.3% and 81.8% in the ELN and RF case respectively. The area under the curve is 0.818 and 0.821 regarding ELN and RF respectively. Both ELN and RF show great performance in assessing biomarkers for COVID-19 severity. The results between the two models are comparable and common features were found, showing no predominance of any of the two models. Results suggest that both EVs and their sizes are possible biomarkers per COVID-19 severity.

[1] C. Balbi et al., Circulating extracellular vesicles are endowed with enhanced procoagulant activity in SARS-CoV-2 infection, EBioMedicine, 2021.

WP8 Influence of hormonal contraceptives on plasma metabolite concentrations in a large cohort

Dordevic N.*[1], Hernandes V.V.[2], Rainer J.[1]
[1]Institute for Biomedicine (Affiliated to the University of Lübeck), Eurac Research ~ Bozen ~ Italy, [2]†Current affiliation: Department of Food Chemistry and Toxicology, Faculty of Chemistry, University of Vienna ~ Vienna ~ Austria

The widespread consumption of hormonal contraceptives among young females has a large effect on serum metabolite levels which is usually overlooked. In addition to genetic and environmental factors, hormonal status and usage of hormonal contraceptives [1] appear to play mayor role on plasma metabolites. In this work we investigated the effects of oral contraceptives on serum metabolite levels using the targeted metabolomics data set of the Cooperative Health Research in South Tyrol (CHRIS) study [2, 3], consisting of 175 quantified metabolites in 6,872 participants. To this end, linear regression models were fitted separately to each analyte using the log2 transformed concentration as a response variable and sex, age, BMI, self-reported fasting status, and a binary variable for usage of hormonal contraceptives as covariates. To avoid any confounding effect between sex and sex-hormone medication, the analysis was repeated considering only female samples. Including the use of hormonal contraceptives as a covariate has an impact on the results for sex- associated metabolites, and, to a lesser extent, also on age-related metabolites. The other investigated covariates, however, remain unaffected. The analysis performed on female samples only suggested that almost half of all significant sex-related metabolites are misidentified when these medications are not considered. Consumption of hormonal contraceptives appears to have a strong effect on serum metabolite levels in young females, which is often overlooked. This was demonstrated in the analysis of the targeted metabolomics data set of the Cooperative Health Research in South Tyrol (CHRIS) study.

[1] J.M. Ramsey, J.D. Cooper, B.W. Penninx , S. Bahn. Variation in serum biomarkers with sex and female hormonal status: implications for clinical tests. Sci Rep. (31) 2016 ; 6:26947.
[2] C. Pattaro, M. Gögele, D. Mascalzoni, R. Melotti, C. Schwienbacher, A. De Grandi, L. Foco, Y. D'Elia, B. Linder, C. Fuchsberger, C. Minelli, C. Egger, L.S. Kofink, S. Zanigni, T. Schäfer, M.F. Facheris, S.V. Smárason, A. Rossini, A.A. Hicks, H. Weiss, P.P. Pramstaller. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J Transl Med 13, 2015,348.
[3] V.V. Hernandes, N. Dordevic, E.M. Hantikainen, B.B. Sigurdsson, S.V. Smárason, V. Garcia- Larsen, M. Gögele, G. Caprioli, I. Bozzolan, P.P. Pramstaller, J. Rainer. Age, Sex, Body Mass Index, Diet and Menopause Related Metabolites in a Large Homogeneous Alpine Cohort. Metabolites 12, 2022, 3: 205.

WP9 Association of serum resistin with blood stasis syndrome in korean medicine for metabolic diseases

Mi Mi K.*
KM Science Research Division, Korea Institute of Oriental Medicine ~ Daejeon ~ Korea, Republic of

Blood stasis syndrome (BSS) in Traditional East Asian Medicine is multidisciplinary involving numerous medical specialties. Many studies have attempted to identify objective indicators that aid in the diagnosis of BSS in various fields. Resistin is an adipokine and known to be related to metabolic diseases. In this study, we investigated the levels of serum resistin and other proteins related to metabolic syndrome (MS) and several other diseases categories to identify the association with BSS. Blood stasis syndrome (BSS) in Traditional East Asian Medicine is multidisciplinary involving numerous medical specialties. Many studies have attempted to identify objective indicators that aid in the diagnosis of BSS in various fields. Resistin is an adipokine and known to be related to metabolic diseases. In this study, we investigated the levels of serum resistin and other proteins related to metabolic syndrome (MS) and several other diseases categories to identify the association with BSS. We showed a significant increase in serum resistin levels in patients with BSS with metabolic diseases. These results suggest that that resistin levels are potentially associated with the pathogenesis of BSS in Korean Medicine for metabolic diseases.

[1] Gunter RN, Blood stasis–China's classical concept in modern medicine, New York: Elsevier, 2007.
[2] Al-Daghri N, Chetty R, McTernan PG, Al-Rubean K, Al-Attas O, Jones AF, Kumar S., Cardiovasc Diabetol, vol. 4, no. 10, 2005.

WP10 Prognostic value of inflammatory indexes in patients with hepatocellular carcinoma: results from the ita.Li.Ca

Simeon V.*[1], Sgamato C.[2], Pelizzaro F.[3], Signoriello S.[1], Fordellone M.[1], Farinati F.[3], Nardone G.[2], Trevisani F.[4], Giannini E.G.[5], Rocco A.[2], Chiodini P.[1]
[1]Unit of Medical Statistic, Department of Mental, Physical Health and Preventive Medicine, University of Campania 'L. Vanvitelli' ~ Caserta ~ Italy, [2]Department of Clinical Medicine and Surgery, Gastroenterology, University Federico II ~ Napoli ~ Italy, [3]Department of Surgery, Oncology and Gastroenterology, University of Padova ~ Padova ~ Italy, [4]Unit of Semeiotics, Liver and Alcohol-related diseases, Department of Medical and Surgical Sciences, IRCCS Azienda Ospedaliero- Universitaria di Bologna; Department of Medical and Surgical Sciences, University of Bologna ~ Bologna ~ Italy, [5]Gastroenterology Unit, Department of Internal Medicine, University of Genova, IRCCS Ospedale Policlinico San Martino ~ Genova ~ Italy

Primary liver cancer, mostly hepatocellular carcinoma (HCC), is the third leading cause of cancer- related death worldwide. Currently, prognostic models are based on tumor burden, symptoms, and liver function. The systemic inflammatory response (SIR) is gaining interest as a prognostic biomarker. High levels of SIR markers, such as neutrophil-lymphocyte ratio (NLR) and platelet-lymphocyte ratio (PLR), may reflect increased pro-tumor inflammation and decreased anti-tumor immune function [1]. However, heterogeneity in study design, sample size, and statistical approaches prevent conclusive clinical applicability. The aim of this study was to analyse the prognostic significance of NLR and PLR in a large cohort, named ITA.LI.CA, of Caucasian patients with HCC. The cohort consisted of patients diagnosed with HCC in 24 Italian centres from 2000 to 2018. Univariate and multivariate Cox proportional hazards models were used to assess the prognostic role of each biomarker on overall survival (OS), adjusting for age, Barcelona Clinic Liver Cancer staging and alpha-fetoprotein levels. The clinical value and discrimination properties of SIR biomarkers were evaluated using the basic clinical model with added biomarkers. Of 7882 eligible HCC patients, 1107, with complete laboratory data, were included in the final analysis. Most were males with a median age of 69 years and had cirrhosis. NLR was associated with OS as a logarithmic function (HR 1.61, 95% CI 1.39 - 1.86), while PLR was linearly associated (HR 1.16, 95% CI 1.04 - 1.29). The best cut-off that minimised the p-value of HR was 1.45 and 190.7 for NLR and PLR, respectively. 81.5% of patients had NLR >1.45 and 7.7% had PLR > 190.7. Both biomarkers were significantly associated with poor prognosis in univariate models, and both were significantly associated with OS in the multivariate model adjusted for clinical covariates (NLR: HR 1.71, 95%CI 1.23 - 2.3; PLR: HR 1.74, 95%CI 1.17 - 2.58). However, after applying a shrinkage procedure [2], only NLR remained associated with OS (HR 1.64, 95%CI 1.13-2.39). Both biomarkers had an incremental effect on predicting OS when added to the clinical model. In particular, NLR could be a useful prognostic biomarker for HCC in Caucasian patients.

[1] Yu LX, Ling Y, Wang HY. NPJ Precis Oncol. 2018 Feb 23;2(1):6.
[2] Holländer, N.; Sauerbrei, W.; Schumacher, M. Stat. Med. 2004, 23, 1701–1713.

WP11 Shotgun lipidomics for the development of a diagnostic test for schizophrenia

Tkachev A.*[1], Stekolshchikova E.[1], Morozova A.[2], Andreyuk D.[2], Kostyuk G.[2], Khaitovich P.[1]
[1]Skolkovo Institute of Science and Technology ~ Moscow ~ Russian Federation, [2]Moscow Psychiatric Hospital No. 1, named after N.A. Alekseev ~ Moscow ~ Russian Federation

Researchers and clinicians are in pursuit of a reliable and reproducible molecular signature of psychiatric disorders, for which, currently, no clinically useful diagnostic test exists. With the development of lipidomics methods, the possibilities of using lipids as diagnostic biomarkers are expanding. Previously, we have identified a reproducible signature of lipidome alterations in the blood plasma of individuals with schizophrenia (Tkachev et al, 2023). We proposed a lipid-based predictive model, which high performance in separating patients with the psychiatric disorder from healthy controls suggests the possibility of its further development for potential applications. Here, are moving in this direction by considering a simplified analytical method for assessing the blood plasma lipidome. The direct-infusion mass spectrometry approach we are using enables good coverage of all major lipid classes, while having the advantages of significantly simplified experimental procedures and reliable quantification compared to the liquid chromatography mass spectrometry approach more commonly used for lipidomics. Using this direct-infusion lipidomics approach, we have assessed the blood plasma lipidome of patients with schizophrenia and healthy controls and defined an updated lipid-based diagnostic model separating individuals with psychiatric disorders from controls. To test the possibilities of the diagnostic model for screening purposes, we have further collected blood plasma from a cohort of 500 supposedly healthy volunteers, for which blood plasma lipids will be assessed and the diagnostic lipid-based model would be applied. We suppose that the blood plasma lipidome reflects clinically useful information about the individuals' psychiatric health, and that direct-infusion mass spectrometry provides the appropriate tool for capturing this information for potential practical applications.

Tkachev A, Stekolshchikova E, Vanyushkina A, et al. Lipid Alteration Signature in the Blood Plasma of Individuals With Schizophrenia, Depression, and Bipolar Disorder. JAMA Psychiatry. 2023;80(3):250- 259. doi:10.1001/jamapsychiatry.2022.4350 Supported by Moscow Center for Innovative Technologies in Healthcare, grant №2707-2 .

WP12 A calibrated sensitivity model for observational studies with multivalued treatments

Md Abdul B.*[1], Mahbub Ahm L.[1], Abdus S W.[2]
[1]Institute of Statistical Research and Training ~ Dhaka ~ Bangladesh, [2]University of Rochester Medical Center School of Medicine and Dentistry ~ Rochester ~ United States of America

The identification of causal estimands of interest in observational studies depends on the identifiability assumptions, and one of the most common identifiability assumptions is the no unmeasured confounding (NUC) assumption. However, the NUC assumption is not verifiable from observed data, and violation of this assumption may induce systematic bias in the causal effect estimates. This paper proposes extensions of the marginal sensitivity model for binary treatments proposed by Zhao and colleagues [1]. Firstly, we propose a risk ratio (RR) based sensitivity analysis framework as risk ratios are easier to interpret and are consistent with general intuition. Secondly, observing the scarcity of sensitivity models for multivalued treatments, the proposed RR framework is extended to the multivalued treatment setting using generalized propensity scores. We are also working on two further extensions of our proposed framework– (i) the incorporation of the doubly robust AIPW estimators and (ii) the calibration of the sensitivity parameters to the observed confounders to explain the results of sensitivity analysis in a more meaningful way. We estimate partially identified point estimate intervals of causal effects under any specific sensitivity model by converting the IPW AND AIPW estimators of causal effects to linear fractional programming (LFP) problems using Charnes-Cooper transformation [2], which can be solved very efficiently. Using percentile bootstrap, we also obtain 100(1- α)% confidence intervals for the partially identified causal effects under any specific sensitivity model. Simulation results suggest that the proposed sensitivity models perform well in terms of bias in the point estimate intervals and non-coverage rate of the percentile bootstrap confidence intervals when there is an adequate overlap in the covariate distribution among the treatment groups. Lastly, we demonstrate the implementation of our proposed models by conducting an empirical study that estimates the causal effect of maternal education on women's fertility in Bangladesh. This paper proposes a risk ratio based and calibrated sensitivity analysis framework for observational studies with multivalued treatments. An empirical study was conducted to demonstrate sensitivity analysis using our proposed models, which concluded that maternal education has a significant causal effect on women's fertility in Bangladesh.

[1] Zhao, Q., Small, D. S., & Bhattacharya, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. Journal of the Royal Statistical Society Series B: Statistical Methodology, 81(4), 735-761.
[2] Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. Naval Research logistics quarterly, 9(3-4), 181-186.

WP13 Genetically proxied effects of plasma proteins on risk of ms identified via mendelian randomization

Boyton M.*, Richardson T.G., Walker V., Gaunt T.
University of Bristol ~ Bristol ~ United Kingdom

It is widely reported that multiple sclerosis (MS) is associated with changes in levels of circulating plasma proteins, and several putative blood biomarkers have been suggested. However, it remains unclear whether these changes in protein levels are causal in the disease or artifacts of other processes due to the presence of reverse causation and unmeasured confounding in observational studies. We investigated the potential causal role of blood plasma proteins in MS risk through Mendelian randomization (MR) analysis. In this analysis framework, germline genetic variants that affect modifiable plasma protein levels are applied as instrumental variables to examine their causal effects on disease outcomes. Two-sample inverse-variance weighted MR analysis was conducted. Proteome-wide genetic association data from the deCODE consortium [1] covering 1,831 proteins from 35,559 individuals was used for instrumental exposure variables, and outcome summary statistics were taken from the most recent International Multiple Sclerosis Genetics Consortium GWAS study of 47,429 cases and 68,374 controls. [2]. Our analysis yielded 17 plasma proteins with evidence of a risk increasing or protective role in MS (STAT3, LMAN2, SCGN, AHSG, MAPK3, FCRL3, ITLN1, KLRB1, WARS, HSPA1L, CTRB2, UBASH3B, ICAM5, FGFBP3, INHBC, TNFSF14 & CD59). This list includes several well-established protein markers, such as STAT3 and MAPK3, as well as a number of potentially novel proteins involved in processes such as inflammatory signalling, metabolism and cell-cell adhesion. Our analysis indicates a putative causal role for a panel of blood plasma proteins in MS risk. These results may help to guide the development of therapeutic interventions, identify risk factors, and validate blood biomarkers of disease processes. Further analyses are ongoing with the aim of characterising the role of these proteins within a phenome-wide screening context against a comprehensive set of additional outcome diseases and traits. Early results indicate a potential role for these proteins in Crohn's disease and inflammatory bowel disease highlighting a shared immune aetiology with MS.

[1] Ferkingstad, E., P. Sulem, B. A. Atlason, G. Sveinbjornsson, M. I. Magnusson, E. L. Styrmisdottir, K. Gunnarsdottir, A. Helgason, A. Oddsson, B. V. Halldorsson, B. O. Jensson, F. Zink, G. H. Halldorsson, G. Masson, G. A. Arnadottir, H. Katrinardottir, K. Juliusson, M. K. Magnusson, O. T. Magnusson, R. Fridriksdottir, S. Saevarsdottir, S. A. Gudjonsson, S. N. Stacey, S. Rognvaldsson, T. Eiriksdottir, T. A. Olafsdottir, V. Steinthorsdottir, V. Tragante, M. O. Ulfarsson, H. Stefansson, I. Jonsdottir, H. Holm, T. Rafnar, P. Melsted, J. Saemundsdottir, G. L. Norddahl, S. H. Lund, D. F. Gudbjartsson, U. Thorsteinsdottir, K. Stefansson, Nature Genetics, 53(12), 2021, 1712-1721.
[2] International Multiple Sclerosis Genetics Consortium, Science, 365(6460), 2019.

Poster Sessions

WP14 A comparison on real data of methods to develop individualized treatment rules

Brion Bouvier F.*, Peyrot E., Balendran A., Segalas C., Petit F., Porcher R.
Université Paris Cité ~ Paris ~ France

Identifying subgroups of patients who benefit from a treatment is a key aspect of personalized medicine. Developing individualized treatment rules (ITRs), which map individual characteristics to a treatment, can be achieved by identifying these subgroups. Many machine learning algorithms have been proposed to create such rules. Yet, it is unclear to what extent those algorithms lead to the same ITRs, i.e. recommending the treatment for the same individuals. To this aim, we compared the most common approaches in two randomized control trials: the International Stroke Trial and the CRASH-3 trial.
We distinguish two classes of algorithms to develop an ITR. The first class relies on predicting individualized treatment effects and then deriving an ITR by recommending treatment to those with a predicted benefit. In the second class, algorithms directly estimate the ITR by minimizing a loss function. The majority of the algorithms compared in this project fell under the first class: meta- learners (T-learner, S-learner, X-learner and DR-learner, both with parametric and non-parametric models), generalized random forests, and virtual twins, whereas A-learning, Chen's Weighting, outcome weighted learning and contrast weighted learning fell under the second class. When using non-parametric models, results were compared with and without cross-fitting. For each trial, the performance of ITRs was assessed in terms of value of the rule, average benefit of treatment among people with a positive score and among people with a negative score, population average prescription effect, and c-statistic for benefit. The pairwise agreement between ITRs was also calculated using Cohen's kappa and Matthews correlation coefficients. Results showed that the ITRs obtained by the different algorithms generally had considerable disagreements regarding the individuals to be treated. A better concordance was found among algorithms of the same family (e.g. among all meta-learners with parametric models, or all meta-learners with non-parametric models and cross-fitting). Overall, when evaluating the performance of ITRs in a hold-out validation sample, all algorithms produced ITRs with limited performance, whatever the performance in the training set, which suggests a high potential for overfitting.The methods do not lead to the same ITR which draws some concerns about their practical use.

Janes H, Brown MD, Pepe M, Huang Y. Statistical Methods for Evaluating and Comparing Biomarkers for Patient Treatment Selection ; :35.
Imai K, Li ML. Experimental Evaluation of Individualized Treatment Rules. Journal of the American Statistical Association 2021; :1–15.
van Klaveren D, Steyerberg EW, Serruys PW, Kent DM. The proposed 'concordance-statistic for benefit' provided a useful metric when modeling heterogeneous treatment effects. Journal of Clinical Epidemiology 2018; 94:59–68.
Baldi P, Brunak S, Chauvin Y, Andersen CAF, Nielsen H. Assessing the accuracy of prediction
algorithms for classification: an overview. Bioinformatics 2000; 16:412–424.
Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement
1960; 20:37–46.
Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W. Double/Debiased/Neyman Machine Learning of Treatment Effects 2017. URL http://arxiv.org/abs/1701.08687.
Jacob D. Cross-Fitting and Averaging for Machine Learning Estimation of Heterogeneous Treatment Effects. arXiv:200702852 [stat] 2020; .
Künzel SR, Sekhon JS, Bickel PJ, Yu B. Meta-learners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences U S A 2019; 116:4156– 4165.
Kennedy EH. Optimal doubly robust estimation of heterogeneous causal effects 2020. URL https://arxiv.org/abs/2004.14497.
Athey S, Tibshirani J, Wager S. Generalized Random Forests 2018.
Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika 2020;
:asaa076.
Foster J, Taylor J, Ruberg S. Subgroup identification from randomized clinical trial data. Statistics in Medicine 2011; 30:2867–2880.
Chen S, Tian L, Cai T, Yu M. A general statistical framework for subgroup identification and comparative treatment scoring. Biometrics 2017; 73:1199–1209.18
Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association 2012; 107:1106–1118.
Guo X, Ni A. Contrast weighted learning for robust optimal treatment rule estimation. Statistics in Medicine 2022; :sim.9574.
the International Stroke Trial Collaborative Group, Sandercock PA, Niewada M, Członkowska A. The International Stroke Trial database. Trials 2011; 12:101.
Effects of tranexamic acid on death, disability, vascular occlusive events and other morbidities in patients with acute traumatic brain injury (CRASH-3): a randomised, placebo-controlled trial. The Lancet 2019; 394:1713–1723.

WP15 Transporting a 'hospital at home' rct treatment effect to a target population with differing delirium measures

Edgar K.*[1], Shepperd S.[2], Sharples L.[1], Pendlebury S.[3]
[1]London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom, [2]Nuffield Department of Population Health, University of Oxford ~ Oxford ~ United Kingdom, [3]Centre for Prevention of Stroke and Dementia, Nuffield Department of Clinical Neurosciences, University of Oxford ~ Oxford ~ United Kingdom

Participants in randomised trials are highly selected. There is interest in the potential impact on outcomes when results are transferred to a target population. In 2021, a large (n=1032) RCT comparing Hospital at Home (HAH) with traditional inpatient care for elderly patients with acute illnesses was published. Mortality and length of stay in hospital were important outcomes. We aim to assess the effect of introducing HAH into a less selected target population using ORCHARD, a large observational dataset of patients admitted to 4 hospitals. Additionally, we discuss difficulties in generalising results when different definitions for key confounders, specifically cognitive impairment (CI) and delirium, are used. We applied trial exclusion criteria to the ORCHARD population (n=35,580) to identify potentially eligible patients. Using methods from survey statistics, we applied reweighting methods to estimate the effect of HAH relative to hospital admission in the ORCHARD population. Validation of positivity and consistency assumptions [1] was based on overlap of measured variables and assumed that hospital care in the trial control group and ORCHARD hospitals was similar. Strong ignorability of sample assignment [2] assumes we can sufficiently capture the selection mechanism for trial participation through a set of variables, so that the potential outcomes (for ORCHARD and trial control group) are independent of trial assignment. CI and delirium were measured differently to the trial in ORCHARD, which may lead to a violation. In the RCT, 67% of people in the control group and 66% in the HAH group were alive at 6 months, with estimated OR 0.94, 95% CI (0.67 to 1.33), p=0.72. Transferring trial outcomes to the target population resulted in larger but consistent estimated effect of HAH on mortality at 6 months of OR 0.81, 95% CI (0.57 to 1.16), p=0.25. Post weighting checks revealed 5 patients in the trial (all with delirium) had very large weights, suggesting positivity assumption violations. For these methods, populations having different measures for key variables may result in the ignorability and positivity assumptions being violated.

[1] I. J. Dahabreh, S.E. Robertson, J. A. Steingrimsson, E.A. Stuart, M.A. Hernan. Extending inferences from a randomized trial to a new target population. Statistics in medicine. 2020 Jun 30;39(14):1999-2014.
[2] E. Hartman, R. Grieve, R. Ramsahai, J.S. Sekhon. From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. Journal of the Royal Statistical Society. Series A (Statistics in Society). 2015 Jun 1:757-78.

WP16 Spasticity in patients with multiple sclerosis: a bayesian network application

Greco T.*[1], Poole E.[2], Young A.[2], Alexander J.[2]
[1]Jazz Pharmaceuticals, Inc., Gentium Srl ~ Villa Guardia ~ Italy, [2]Jazz Pharmaceuticals, Inc. ~ Philadelphia ~ United States of America

Approaches to causal inference represent attractive tools for discovering relationships in epidemiology studies but application in clinical trials is still poor. We review and apply Bayesian network theory [1] to data from two enriched clinical trials, including participants with spasticity due to multiple sclerosis and randomized to nabiximols or placebo, to identify conditionalities among clinical factors. A total of 261 participants (106 [n=53 nabiximols vs n=53 placebo] and 155 [n=76 nabiximols vs n=79 placebo] in the training and validation set) and about 40 factors were analysed. Data dimensionality was reduced following an iterative step-by-step regression approach. Naïve Bayesian Network using a structural expectation-maximization algorithm with Bayesian-Dirichlet scoring function was applied to estimate maximum-a-posteriori conditional probabilities [1, 2]. A proposed adherence score, calculated as the reciprocal average of the Euclidean Distances between nodes of the data-driven structure and nodes of the bootstrap-derived structure, was used to compare different data networks. Mediation analysis was conducted to estimate the percentage of the treatment effect, including bootstrap 95% confidence intervals, mediated by the spasticity symptom severity variable. The final data-derived structure was applied as a priori knowledge to the validation set data (propagation).
The final causal network, selected due to the higher adherence score (93%) and parsimonious network structure (4 nodes and 3 edges), included: (i) treatment, (ii) end-of-study spasticity symptom severity, as reflected by Spasticity Numerical Rating Scale (NRS) score, and (iii) mental health or (iv) vitality subscales of the SF-36 questionnaire. Conditional probabilities of treatment on mild/moderate/severe spasticity (ie, 0-3/4-6/7-10 points in NRS score) were 0.48/0.02/0.003 and 0.27/0.23/0.003 (RR: 1.77/0.09/0.85) in the nabiximols and placebo arm. In patients with mild spasticity, the impact of nabiximols on mental health or vitality subscales resulted in a probability ratio of 1.63. The decomposed mediation effect of spasticity symptom severity between treatment and mental health or vitality outcomes was 99.4% (95% CI: 35.4% to >99.9%) or 93.7% (95% CI: 52.2% to >99.9%). The network propagation was validated for "treatment-mild spasticity" pathway. Bayesian network framework is an innovative and powerful method to investigate dependencies and generate new hypotheses in clinical trial settings.

[1] Cowell RG. Introduction to inference for Bayesian networks. In: Learning in graphical models. NATO ASI series (Series D: Behavioural and social sciences), vol 89. Dordrecht, Springer, 1998.
[2] Friedman N. Learning belief networks in the presence of missing values and hidden variables. In: Proceedings of the Fourteenth International Conference on Machine Learning,, San Francisco: Morgan Kaufmann Publishers; 1997. pp. 125–133.

Poster Sessions

WP17 Causaloptim: a user-friendly r-package for computation of symbolic sharp bounds on counterfactual queries

Jonzon G.*[1], Sachs M.C.[2], Gabriel E.E.[2]
[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet ~ Stockholm ~ Sweden, [2]Section of Biostatistics, Department of Public Health, University of Copenhagen ~ Copenhagen ~ Denmark

Unmeasured confounding will generally render a causal effect unidentifiable unless strong assumptions are imposed.
If we cannot reliably justify such assumptions, we may resort to deriving nontrivial bounds of the effect. In some non-identifiable settings, we can even derive optimally narrow bounds. A methodology for doing so in the simple instrumental variable setting was developed in [1]. This approach has been adapted and applied to certain similar settings, but has not been unified into a general framework that explores its limitations and characteristics. More importantly, the methodology is often viewed as opaque and difficult to apply in practice. We have developed an algorithm for computation of causal bounds in a generalized framework, characterized a set of causal problems to which it applies, and implemented the methodology in an R-package [2] using an intuitive interface and causal notation familiar in clinical epidemiology. We have unified computation of causal bounds for several settings using linear optimization in a single framework and made it accessible for applied research. Our R-package computes symbolic optimally narrow bounds on causal queries given input in the form of a causal diagram (via a graphical user interface) and potential outcomes notation. It outputs the bounds as text strings as well as R-functions, and is applicable in situations where the observed covariates are categorical, including studies of treatment effects with instrumental variables (e.g. Mendelian randomization) or of direct and indirect effects with unmeasured confounding on the effect of the mediator on the outcome. We demonstrate the package using e.g. data from the Lipid Research Clinics Coronary Primary Prevention Trial, where patients were randomized to cholestyramine treatment or placebo and followed up for coronary heart disease. Having access to symbolic expressions of bounds may give valuable insight even in the absence of data. We believe that causal bounds have been underused in applied research in part because of the computational burden and that our accessible R-package opens up such computations to the applied research community.

[1] Alexander Balke and Judea Pearl. "Bounds on Treatment Effects from Studies with Imperfect Compliance". In: Journal of the American Statistical Association 92.439 (1997), pp. 1171–1176.
[2] Gustav Jonzon, Michael C Sachs, and Erin E Gabriel. Accessible Computation of Tight Symbolic Bounds on Causal Effects using an Intuitive Graphical Interface. 2022. doi: 10.48550/ARXIV.2209.03657.

Poster Sessions

WP18 A note on the latent unconfoundedness assumption for hierarchical observational data: a simulation study

Kostouraki A.*[1], Leyrat C.[2], Rachet B.[1], Belot A.[1]
[1]*Inequalities in Cancer Outcomes Network, Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom,* [2]*Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom*

In hierarchical data, when analysing the causal effect of a treatment defined at the individual level, important cluster-level confounders often remain unmeasured, violating the conditional exchangeability assumption. Conditional exchangeability may be relaxed to the so-called "latent unconfoundedness assumption", i.e., treatment groups are exchangeable given both measured individual-level confounders and the specific cluster. Hence, specifying the values for all cluster-level confounders (measured or unmeasured) for a given cluster is equivalent to conditioning on that cluster. Once having obtained unbiased estimates for the cluster-specific average treatment effects (ATE), one could average over clusters to get an unbiased population ATE estimate. Mixed models are applied as a proxy to unobserved cluster-level confounders and to account for correlation of individuals within clusters. However, independence between the random effects and the included covariates must hold to secure consistent estimates from a mixed-effects model. We examine to what extent latent unconfoundedness is plausible when applying either G-computation or Inverse Probability-of-Treatment Weighting (IPTW) estimators combined with mixed-effects. A simulation study explores the situations i) of an unobserved cluster-level covariate and/or ii) of the presence of between-cluster variation in the effect of a confounder. We consider a binary treatment assigned at the individual level and a binary outcome. We analyse the data using a weight and/or outcome models with random effects to account for clustering. Existing estimators are unbiased for the estimation of cluster-specific effects with moderate and constant cluster sizes and when the effect of an individual-level covariate on the treatment probability varies between clusters. In presence of correlation between the omitted cluster-level confounder and an individual-level confounder, existing estimators perform relatively well when targeting the population ATE. Correlation between the random slope and the variance of one individual-level confounder in the outcome mechanism needs further investigation.

[1] Chang T-H, Stuart EA. Propensity score methods for observational studies with clustered data: A review. Statistics in Medicine. 2022;1-15. doi: 10.1002/sim.9437.
[2] Fan Li, Alan M Zaslavsky, and Mary Beth Landrum. "Propensity Score Weighting with Multilevel Data". In: Statistics in Medicine 32.19 (2013), pp. 3373–3387. doi. 10.1002/sim.5786.

WP19 Tyrosine kinase inhibitors discontinuation in chronic myeloid leukemia: a retrospective cohort study

Koumantakis E.*[1], Fava C.[2], Abruzzese E.[3], Annunziata M.[4], Bocchia M.[5], Caocci G.[6], Iurlo A.[7], Cavazzini F.[8], Elena C.[9], Galimberti S.[10], Luciano L.[11], Pietrantuono G.[12], Rapezzi D.[13], Sorà F.[14], Castagnetti F.[15], Breccia M.[16], Scortechini A.R.[17], Intermesoli T.[18], Stagno F.[19], Scappini B.[20], Beltrami G.[21], Ciceri F.[22], De Gobbi M.[2], Leonetti Crescenzi S.[23], Maggi A.[24], Luzi D.[25], Campiotti L.[26], Bonifacio M.[27], Miggiano M.C.[28], Cilloni D.[2], Berchialla P.[2]
[1]*Department of Public Health and Paediatrics, University of Torino ~ Torino ~ Italy,* [2]*Department of Clinical and Biological Sciences, University of Torino ~ Torino ~ Italy,* [3]*Hematology Unit, S. Eugenio Hospital, Tor Vergata University ~ Roma ~ Italy,* [4]*Division of Hematology, Cardarelli Hospital ~ Napoli ~ Italy,* [5]*Azienda Ospedaliera Universitaria, University of Siena ~ Siena ~ Italy,* [6]*Department of Medical Sciences, University of Cagliari ~ Cagliari ~ Italy,* [7]*Haematology Division, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico ~ Milano ~ Italy,* [8]*Department of Medical Sciences -Haematology and Physiopathology of Haemostasis Section ~ Ferrara ~ Italy,* [9]*Hematology Hunit, Fondazione IRCCS Policlinico San Matteo ~ Pavia ~ Italy,* [10]*Hematology Department, University of Pisa ~ Pisa ~ Italy,* [11]*Division of Hematology - Departments of Clinical Medicine and Surgery, University of Napoli Federico II ~ Napoli ~ Italy,* [12]*IRCCS, Centro Di Riferimento Oncologico Della Basilicata ~ Rionero in Vulture ~ Italy,* [13]*S.C. Ematologia, ASO S. Croce e Carle ~ Cuneo ~ Italy,* [14]*Hematology Department, University Cattolica del Sacro Cuore - Policlinico A. Gemelli ~ Roma ~ Italy,* [15]*Institute of Hematology "L. & A. Seràgnoli", St. Orsola University Hospital ~ Bologna ~ Italy,* [16]*Department of Cellular Biotechnologies and Hematology, University La Sapienza ~ Roma ~ Italy,* [17]*Ematologia, Azienda Ospedaliero - Universitaria Ospedali Riuniti ~ Ancona ~ Italy,* [18]*Hematology, Bergamo Hospital ~ Bergamo ~ Italy,* [19]*Division of Hematology and Bone Marrow Transplant, AOU Policlinico—V. Emanuele ~ Catania ~ Italy,* [20]*AOU Careggi ~ Firenze ~ Italy,* [21]*Policlinico San Martino ~ Genova ~ Italy,* [22]*San Raffaele Scientific Institute ~ Milano ~ Italy,* [23]*Ematologia, Ospedale S. Giovanni Addolorata ~ Roma ~ Italy,* [24]*Division of Hematology, Hospital "S.G. Moscati" ~ Taranto ~ Italy,* [25]*Ematologia, A.O. Santa Maria - Terni S.C Oncoematologia ~ Terni ~ Italy,* [26]*Department of Clinical and Experimental Medicine, Università dell'Insubria ~ Varese ~ Italy,* [27]*Divisione di Ematologia, Istituti Ospitalieri di Verona, Policlinico G.B. Rossi ~ Verona ~ Italy,* [28]*Ematologia, Ospedale Vicenza ~ Vicenza ~ Italy*

In the last 15 years different studies analyzed the outcome of patients with a sustained deep molecular response (DMR) who discontinued tyrosine kinase inhibitors (TKIs), and demonstrated that it is safe to discontinue treatment with tyrosine-kinase according to the current recommendations. We aim at evaluating TFR in the setting of clinical practice for bringing out unmet clinical needs and optimizing already consolidated practices and obtaining increasingly personalized treatments based on their disease profile. We designed a retrospective and prospective observational study on patients with Chronic Myeloid Leukemia (CML) who discontinued TKIs in Italy. In the analysis will be considered patients with at least 1 year of follow-up. On the first 435 patients enrolled, we evaluated the impact of the last TKI administered (first or second generation) before discontinuation on DMR's loss. Potential considered confounders and prognostic factors are age, sex, Sokal score, ELTS risk, type of transcript, duration of TKI therapy, time to DMR, DMR duration, line of therapy at discontinuation, MR's depth, and reasons for stopping. Causal machine learning approaches based on targeted minimum loss-based estimation (TMLE) and causal survival forests, based on multiple causal decision trees, will be adopted and results will be compared to the inverse probability of treatment weighting (IPTW) method to estimate an adjusted hazard ratio for DMR's loss and TKI's groups (1,2). A simulation study will be performed to compare the performance of the estimators in terms of bias, coverage probability, type I and type II error. The advantages and disadvantages of the approaches will be discussed. In our population, we did not show a significant impact of type of TKI at discontinuation on DMR's loss. Estimators that allow for better separation between baseline covariates and treatment effect can avoid potential bias for covariates post hoc selection. Machine Learning approaches are robust enough to model miss-specification.

1. Suk Y, Kang H, Kim JS. Random Forests Approach for Causal Inference with Clustered Observational Data. Multivariate Behav Res. 2021 Nov-Dec;56(6):829-852. doi: 10.1080/00273171.2020.1808437. Epub 2020 Aug 28. PMID: 32856937.
2. Gruber S, van der Laan MJ. Targeted minimum loss based estimator that outperforms a given estimator. Int J Biostat. 2012 May 18;8(1):Article 11. doi: 10.1515/1557-4679.1332. PMID: 22628356; PMCID: PMC6052865.

Poster Sessions

WP20 Tell me what you want, what you really really want: estimands in observational pharmacoepidemiologic studies

Luijken K.*[1], Van Eekelen R.[2], Gardarsdottir H.[3], Groenwold R.[4], Van Geloven N.[4]
[1]University Medical Center Utrecht ~ Utrecht ~ Netherlands, [2]Amsterdam University Medical Center ~ Amsterdam ~ Netherlands, [3]University Utrecht AND University of Iceland ~ Reykjavik ~ Iceland, [4]Leiden University Medical Center ~ Leiden ~ Netherlands

Ideally, the objectives of a pharmacoepidemiologic comparative effectiveness or safety study should dictate its design and data analysis. This paper discusses how defining an estimand is instrumental to this process. We applied the ICH-E9 (Statistical Principles for Clinical Trials) R1 addendum on estimands [1] – which originally focused on randomized trials – to three examples of observational pharmacoepidemiologic comparative effectiveness and safety studies. Five key elements specify the estimand: the population, contrasted treatments, endpoint, intercurrent events, and population-level summary measure. Different estimands were defined for case studies representing three types of pharmacological treatments: (1) single-dose treatments using a case study about the effect of influenza vaccination versus no vaccination on mortality risk in an adult population of ≥60 years of age; (2) sustained-treatments using a case study about the effect of dipeptidyl peptidase 4 inhibitor versus glucagon-like peptide-1 agonist on hypoglycemia risk in treatment of uncontrolled diabetes; and (3) as-needed treatments using a case study on the effect of nitroglycerin spray as-needed versus no nitroglycerin on syncope risk in treatment of stabile angina pectoris. The case studies illustrated that a seemingly clear research question can still be open to multiple interpretations. Defining an estimand ensures that the study targets a treatment effect that aligns with the treatment decision the study aims to inform. Estimand definitions further help to inform choices regarding study design and data-analysis and clarify how to interpret study findings.

[1] ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands- sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf

WP21 Handling informative patient monitoring in routinely-collected data used to estimate treatment effects

Pirondini L.*[1], Diaz Ordaz K.[2], Keogh R.[1]
[1]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom, [2]University College London ~ London ~ United Kingdom

Routinely-collected hospital data provide opportunities to gain understanding of treatment effects that would not be feasible in randomised trials and that reflect their impact in realistic clinical practice. However, these data present major challenges of time-dependent confounding, which must be addressed in the analysis to enable causal inferences to be made. A specific challenge presented by hospital data is that measurements of patients' clinical status are made at high frequency, on differing schedules for each patient dependent on their clinical status, so timing and frequency of measurements is informative. However, many existing causal inference methods assume measurements of treatments and confounders are made at regular time intervals. The aim of this work is to evaluate methods for estimating causal effects of longitudinal treatments in the presence of informative monitoring. We consider a typical data structure in which monitoring depends on patient characteristics and can vary by covariate type, across patients, and over time. This is motivated by hospital data on patients with COVID-19 and questions about optimal mechanical ventilation strategies. We compare methods based on (i) marginal structural models fitted by inverse probability of treatment and monitoring weighting (IPW) [1], (ii) G-computation with imputation of missing covariate data [2], and (iii) targeted maximum likelihood estimation. We evaluate these methods using a simulation study, comparing against a more simple approach using last-observation-carried-forward (LOCF) ignoring monitoring. Data are simulated to represent a range of realistic scenarios with time-varying treatment and covariates, in which monitoring depends on past covariate and treatment levels. We also illustrate methods in a real-world example using routinely-collected hospital data. We show that ignoring monitoring can result in bias, the size of which depends on informativeness of the monitoring process. IPW-based methods result in the smallest bias, and, when it can be assumed that treatment is only affected by measured covariates, IPW-based methods do not require imputation of missing covariate data. However, methods based on g-formula with imputation benefit from relatively higher efficiency. Data with informative monitoring are common in observational studies, but there is a lack of readily- implementable methods to handle them. We describe three methods and evaluate their performance.

[1] N Kreif, O Sofrygin, JA Schmittdiel, AS Adams, RW Grant, Z Zhu, MJ van der Laan, R Neugebauer, Exploiting non-systematic covariate monitoring to broaden the scope of evidence about the causal effects of adaptive treatment strategies. Biometrics, 2021; 77: 329–342.
[2] C Leyrat, JR Carpenter, S Bailly, EJ Williamson, Common Methods for Handling Missing Data in Marginal Structural Models: What Works and Why. American Journal of Epidemiology, 2020; 190: 663-672

WP22 Estimating the effect of treatment delay: g-computation versus clone- censor-reweight approach

Prosepe I.*, Le Cessie S., Van Geloven N.
Leiden University Medical Center ~ Leiden ~ Netherlands

When patients seek treatment, doctors may decide to delay treatment initiation, to see if natural recovery occurs. The impact of the chosen treatment strategy (e.g. starting treatment at 3 months or wait at least 3 months before starting treatment) is often unknown. In this work we compare two methods that can, under specific assumptions, quantify the causal effect of different treatment strategies using observational data subject to baseline confounding. Our estimand of interest is the marginal cumulative proportion of recovered patients under different treatment strategies. The first method combines multistate modelling with g-computation to target this estimand. The multistate model is similar to an illness-death model where patients can transition from disease to treatment, from treatment to recovery and from disease to recovery. Each transition is modeled via Cox proportional hazards models, with all relevant confounders included as covariates. The waiting time before entering the treatment state is included as one of the covariates for the transition from treatment to recovery. The cumulative percentage of recovered patients is obtained subsequently employing g-computation. The second method is the clone-censor-reweight approach. In this method, patients are artificially censored at the moment they deviate from the delay strategy of interest and reweighting is applied to account for dependency between the censoring and the outcome process, where weights depend on the set of confounders. Estimation of cumulative proportions of recovered patients then proceeds by the weighted Kaplan-Meier method. The assumptions needed for the two methods are contrasted. We run a simulation to numerically compare the methods regarding accuracy and efficiency. We generate synthetic observational data and then assess the difference between the estimated and true cumulative percentage of recovered patients, had all patients transitioned to treatment according to the same treatment strategy. We apply both methods to a cohort of 1896 couples with unexplained subfertility who seek artificial reproductive therapy. We estimate the cumulative proportion of pregnancies 1.5 years after diagnosis under different treatment strategies. Uncertainty is quantified by bootstrapping. Our comparison provides insight into the performance of both method regarding their ability to estimate the effect of different treatment strategies from observational data.

[1] M A Hernán, BMJ, 360, 2018; k182

WP23 Sensitivity analysis with matched pairs from observational studies

Saha P.*, Md Abdul B., Ahm Mahbub L.
Dhaka University ~ Dhaka ~ Bangladesh

The causal conclusion from the observational studies relies on the assumption about overt or hidden bias in the data. Overt bias can be avoided using methods like matching or stratification, and hidden bias cannot be overcome due to the unavailability of the required data. Sensitivity analysis can be performed to examine the magnitude of the unobserved confounding required to alter the inferential decisions. Rosenbaum's sensitivity analysis framework is one of the pioneering frameworks [1,2]. In this study, the performance of Rosenbaum's sensitivity framework is examined through simulation studies. This framework is also applied to real data; the causal effect of maternal education on the number of antenatal care (ANC) visits in Bangladesh is assessed and the sensitivity of the results is examined through Rosenbaum's method. Rosenbaum's framework seems to work well for large sample sizes, and the bias decreases with higher coefficients of the observed and unobserved covariates related to the treatment. As a real-life example, the causal effect of maternal education on the number of ANC visits in Bangladesh is assessed in this study. Using nationally representative survey data, the number of ANC visits and maternal education are found to be associated, but this association may be biased due to observed and unobserved confounders. To overcome the overt bias in the association, a matched sample of different treatment (maternal education) levels is obtained using the observed covariate. In the matched sample, the causal odds ratio between maternal education and ANC visit was 2.67. Using Rosenbaum's approach, it was found that an unobserved confounder should exist to alter the inferential decisions, which increases the odds of ANC visits threefold. The existence of such a confounder is rare. So, the causal estimate can be said to be insensitive to hidden bias from our analysis. Rosenbaum's approach should be used in large samples and the method is most effective when the sensitivity coefficients have large impact. We found a significant causal effect of maternal education on the number of ANC visits.

[1] Rosenbaum, P. R. and Rubin, D. B. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society: Series B (Methodological), 45(2), 1983, 212–218.
[2] Rosenbaum, P. R. Observational Studies, Springer, 2002.

Poster Sessions

WP24 On causal inference for the relative survival setting

Smith M.*[1], Leyrat C.[1], Belot A.[1], Luque Fernandez M.A.[2], Maringe C.[1], Kostouraki A.[1], Rachet B.[1]
[1]*London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom,* [2]*University of Granada ~ Granada ~ Spain*

Most competing risks settings rely on a strong assumption that the exact cause of death is known. In the relative survival setting, we do not rely on records for the cause of death and instead use information from population life tables, which provide estimates of the expected mortality rates in the general population. We aim to define a framework for estimating causal effects of a treatment on cause-specific quantities of death in the context of the relative survival setting. First, cause-specific hazards (the disease-specific hazard and the other-cause hazard) are estimated by using information from the relative survival setting in the form of population life tables (stratified by sociodemographic characteristics).[1] We derive probabilities for the event-type (disease-specific death or other-cause death) from the relationship between these cause-specific hazards. These probabilities are then used to weight the all-cause (observed) mortality.[2] After applying the weights, the total causal effect is estimated using the g-formula in a conventional competing risk analysis.[3] The performance of our method is assessed using a simulation study. Under identifiability assumptions, our method is unbiased, and has optimal coverage, when there are sufficiently stratified life tables and correctly specified excess hazard models. Quantification of bias and coverage of the total causal effect is discussed in scenarios when either the life table is insufficiently stratified or when the excess hazard model is incorrectly specified. Causal effects of a treatment on cause-specific death can be obtained within a well-defined framework using information from the relative survival setting. We show here that our approach can easily adapt causal methods developed in the cause-specific survival setting to the relative survival setting (i.e., in the absence of reliable information on the cause of death).

[1] Pohar Perme, M., Estève, J. & Rachet, B. Analysing population-based cancer survival – settling the controversies. BMC Cancer 16, 933 (2016). https://doi.org/10.1186/s12885-016-2967-9
[2] Maringe, C., Pohar Perme, M., Stare, J. & Rachet, B. Explained variation of excess hazard models. Statistics in Medicine, 37(14):2284-2300 (2018). https://doi.org/10.1002/sim.7645
[3] Young, J.G., Stensrud, M.J., Tchetgen Tchetgen, E.J. & Hernan, M.A. A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine, 39(8):1199–1236 (2020). doi:10.1002/sim.8471

WP25 Missing data handling and model-agnostic tests for treatment effect heterogeneity in randomised trials

Van Vogt E.*[2], Diaz--Ordaz K.[1]
[1]*University College London ~ London ~ United Kingdom,* [2]*Imperial College London ~ London ~ United Kingdom*

Randomised Clinical trials (RCTs) are the gold standard for evaluating treatment effectiveness. For RCT results to be generalisable to the greater population, they must include a diverse set of patients however, this increased diversity may lead to non-random variation in the treatment effects, so we are interested in going beyond average treatment effects and estimating heterogeneous treatment effects (HTEs). Classical approaches to HTE estimation typically involve parametric estimation of interaction coefficients between treatment and patient covariates. Causal machine learning (CML) approaches instead focus on using non-parametric or semi-parametric models to learn the conditional average treatment effect (CATE), the treatment effect conditional on patient covariates. We re-analyse data from the TRACT trial, which compared liberal vs conservative blood transfusion (30 vs 20ml/kg) on in-hospital mortality in children with uncomplicated severe anaemia in Uganda and Malawi [1]. Our primary outcome was 28-day mortality. We considered both non-parametric (causal forest, Bayesian causal forest) and parametric (hierarchical lasso) models to estimate the CATE. After CATE estimation, we implement a general "omnibus" test to establish the presence of treatment heterogeneity [2]. We also use variable importance measures (VIMP) to study the drivers of these heterogeneity. Missing data handling methods in this field are limited, here we compare complete case analysis with inverse probability weighting. Using the omnibus test in conjunction with CML methods and variable importance measures, we find strong evidence of HTE through temperature variation and undernutrition. This compares with the original RCTs planned subgroup analysis, where there was a significant interaction between presence of fever and the treatment . CML methods can control for type I error without the need for a priori specification of covariates. This flexible, agnostic approach allows for exploration of subgroups with heterogenous effects which might not otherwise be considered. We have demonstrated that IPW is a potential candidate for more sophisticated handling of missing data. Further, the agnostic tests applied in this work don't rely on the individual properties of CML algorithms, allowing for easy comparison between models.

1. Maitland K, Kiguli S, Olupot-Olupot P, Engoru C, Mallewa M, Saramago Goncalves P, et al. Immediate Transfusion in African Children with Uncomplicated Severe Anemia. New England Journal of Medicine. 2019 Aug 1;381(5):407–19.
2. Imai K, Li ML. Statistical Inference for Heterogeneous Treatment Effects Discovered by Generic Machine Learning in Randomized Experiments. arXiv:220314511 [stat] [Internet]. 2022 Mar 28 [cited 2023 Mar 31]; Available from: https://arxiv.org/abs/2203.14511

WP26 Multistate models for the analysis of time-to-event data: the case of myeloproliferative neoplasms

Carobbio A.*, Carioli G., Ghirardi A., Barbui T.
FROM Fondazione per la Ricerca Ospedale di Bergamo ~ Bergamo ~ Italy

Survival analyses in medical research are mainly referred to Kaplan–Meier curves and Cox models. Chronic Myeloproliferative neoplasms (MPNs) are characterized by a long period of disease before death, during which several major outcomes occur, such as thromboses and disease evolution in overt myelofibrosis (MF) and acute leukemia (AL). Among MPNs, epidemiologic studies have demonstrated that essential thrombocythemia (ET) and prefibrotic primary myelofibrosis (pre-PMF), despite similar clinical presentation at diagnosis, have very different outcomes later. However, these studies have focused on one isolated outcome at a time, using Cox-proportional hazard regression models, without considering the entire spectrum of multiple intermediate disease states. This situation calls for a multistate model to get a novel and more in-depth insight of factors that may influence this progressive transitioning, provide accurate information on how these specific clinical changes occur and express an impact on survival. A parametric Markov multistate survival model[1] was applied to 791 ET and 382 pre-PMF patients to analyze data of survival considering as possible intermediate states the occurrence of an incident thrombotic event and/or the evolution to overt MF and/or AL. After 10 years, the state occupation probability of being event-free was 70% and 50% in ET and pre-PMF, respectively. The probability of remaining in the status of thrombosis was much higher in ET and decreases only after 20 years, in favor of the probability of death and evolution in overt MF. On the other hands, pre-PMF showed a more rapid decline of the probability of survival free-from-events, due to an earlier mortality and incidence of hematological evolutions, which instead developed later in the ET. This trend was particularly evident looking at the probability of death expected directly from diagnosis (i.e., regardless of the pathways through hematological evolutions) which was double in pre-PMF than ET, reaching almost 30%, 60% and 80% vs. 15%, 30% and 60% at 5, 10 and 20 years, respectively. Analysis of multiple pathways in time-to-event data, such as MPNs, needs to extend beyond the Kaplan-Meier estimator and Cox models and claims multistate models as the gold standard methodology.

[1] Crowther MJ, Lambert PC. Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. Stat Med. 2017;36:4719–4742.

WP27 Increased cardiac risk after a second malignant neoplasm among childhood cancer survivors, a fccss study

Charrier T.*[1], Allodji R.[2], Latouche A.[3], De Vathaire F.[2]
[1]*Université Paris-Saclay ~ Villejuif ~ France,* [2]*INSERM ~ Villejuif ~ France,* [3]*INSERM ~ Saint Cloud ~ France*

Advances in cancer treatment have significantly improved childhood cancer survival, with five-year survival exceeding 80% in most European countries today. The growing population of childhood cancer survivors (CCS) exceeds 300,000 people in Europe, and is known to suffer from many late effects from childhood cancer treatment. Among the most severe late effects are Second Malignant Neoplasm (SMN) and Cardiac Disease (CD). Previous works have identified important risk factors of CD, but the impact of other late-effects, such as SMN, has been ignored. We have studied the effect of time-dependent SMN status on both the cumulative incidence and the instantaneous risk of CD, while accounting for the competing risk of death. To study the effect of SMN on the cumulative incidence of CD, we used an additive model combined with a landmark strategy. We chose the additive model for ease of interpretation, and the landmark strategy to account for the time-dependent status of SMN. We used multiple landmark times, and two time scales (time after childhood cancer diagnosis, and patient age) to get a full overview of the evolution of SMN effect with time. To study the effect of SMN on the cause-specific hazard of CD, we used a proportional cause specific hazard model, with SMN defined as a time-dependent covariate. In both cases, we accounted for death as a competing risk to properly evaluate the impact of SMN. The motivating example is the French Childhood Cancer Survivors Study, which follows 7,670 five- year childhood cancer survivors. With a median follow-up of 30 years, 378 CDs, including 49 after a SMN, were identified. When adjusting on radiotherapy and chemotherapy doses, age at childhood cancer diagnosis, and sex, we found a 2-fold increase in the cause-specific-hazard of CD after a SMN. We also found that among patients who survived at least 25 years after childhood cancer diagnosis, experiencing a SMN increased cumulative incidence of CD 3.64% (95% CI: 0.38%-6.90%). We quantified the increased risk of cardiac disease after a second malignant neoplasm among childhood cancer survivors at multiple time points, helping better identify survivors at high cardiac risk.

[1] Cortese G, Andersen PK. Competing risks and time-dependent covariates. Biom J. 2010 Feb;52(1):138–58.

WP28 A semi-parametric, non-proportional competing risks model predicting outcomes after kidney transplantation

Coemans M.*[1], Verbeke G.[2], Naesens M.[3]
[1]Leuven Biostatistics and Statistical Bioinformatics Centre (L-Biostat), KU Leuven ~ Leuven ~ Belgium, [2]Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-Biostat), Universiteit Hasselt and KU Leuven ~ Hasselt and Leuven ~ Belgium, [3]Department of Microbiology, Immunology and Transplantation, KU Leuven ~ Leuven ~ Belgium

In kidney transplantation, older donors increase the risk of post-transplant graft failure, while older recipients lower this risk. No clinically useful model exists that quantifies the cumulative incidence of graft failure or recipient death for all possible donor-recipient age combinations. We performed a retrospective cohort study on adult kidney transplantations collected in a large European database between 2000 and 2019 (n=162 780). The relation between donor/recipient age and the post-transplant outcomes, graft failure and recipient death with functioning graft, was studied using a multivariable competing risks survival model. We modeled the cause-specific hazards of both outcomes via competing risk-censored Cox models and combined these to attain absolute risk estimates. Donor and recipient age were treated as continuous predictors, and their effects were accommodated to potential non-linearity and non-proportionality. Five- and ten-year cumulative incidence estimates of graft failure and recipient death were calculated for all donor-recipient age combinations. Recipient age showed a non-proportional effect in both cause-specific hazards models. Donor age was the strongest predictor of the rate of graft failure, while recipient age was the strongest predictor of the rate of recipient death. Recipient age had a protective effect on the graft failure rate that stabilized at approximately 60 years. At the cumulative incidence scale however, the risk of graft failure consistently decreased with a higher recipient age, and consistently increased with a higher donor age. Indeed, due to more deaths with a functioning graft in older recipients, the absolute risk of graft failure did not stabilize after 60 years. The cumulative incidence of recipient death primarily increased due to aging recipients. Old-to-young transplantations were at highest risk of graft failure, while young-to-old and old-to-old transplantations were at lower risk. Young-to-old and old-to-old transplantations were at higher risk of death. Using competing risks methodology, we calculated unbiased and clinically useful cumulative incidence estimates of allograft failure and recipient death with a functioning graft for all donor-recipient age combinations in a large European kidney transplant cohort. The risk of graft failure was highest in old- to-young transplantations, and lowest in young-to-old transplantations.

WP29 Analyzing duration of response in phase ii oncology trials

Cyrille S.*[1], Jiang C.[1], Latouche A.[1], Rotolo F.[2], Paoletti X.[1]
[1]INSERM U900 STAMPM, Institut Curie ~ Paris ~ France, [2]Innate Pharma ~ Paris ~ France

The overall objective of phase II oncology trials is to evaluate the anti-tumor activity of new drugs. Duration of response (DoR) is a clinically important endpoint used to support drug approval applications and to inform drug development decisions. Standard approach to analyze DoR is to estimate the median DoR in patients who achieve a response. However, this measure in the subset of patients who responded does not capture the complete information of the tumor response data. Additionally, comparative analysis of DoR in responders is likely to be biased since responders may not be comparable between the randomized treatment arms with respect to baseline characteristics and is formally discouraged by the European Medicines Agency. As an alternative measure, the probability-of-being-in-response function was defined to evaluate tumor response across all randomized patients [1]. Recently, this approach has been derived in the context of illness-death multistate models: the time spent in response (TSiR) [2]. It combines the response rate (RR) and DoR. We evaluated the statistical properties of TSiR in the context of randomized trials. We simulated a multistate model that illustrates patient's trajectories after randomization with three transitions: randomization to response, randomization to progression and, response to progression. We considered different scenarios with moderate sample sizes as seen in phase II clinical trials and various RR (from 40% to 60%). An exponential distribution was assumed for all transitions. We analyzed the effect of covariates on TSiR. A prognostic factor associated with DoR and RR but balanced between arms was introduced. This study found that the parametric estimator of TSiR was robust with small bias. The non-parametric estimator of TSiR had higher bias than the parametric estimator. These findings suggest that the estimator of TSiR is a relevant measure of anti-tumor activity. It has attractive properties including the incorporation of covariates in order to evaluate the effect of potential factors on response in all randomized patients.

[1] C. B. Begg, M. Larson, Biometrics, 38(1), 1982, 59-66.
[2] C. Jiang, F. Rotolo, A. Leary, A. Latouche, X. Paoletti, Revue d'Épidémiologie et de Santé Publique, 70, 2022, S63-S85.

Poster Sessions

WP30 A multi-state model evaluating the association of oxygen therapy with the course of cystic fibrosis in europe

Simone G.*[1], Annalisa O.*[2], Anna Z.[2], Federico A.[2]
[1]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Healthcare Professions Department ~ Milano ~ Italy, [2]University of Milan, Department of Clinical Sciences and Community Health, Laboratory of Medical Statistics, Biometry and Epidemiology "G. A. Maccacaro" ~ Milano ~ Italy

The most serious complications of cystic fibrosis (CF) relate to respiratory failure, leading to hypoxemia, which is the time when the respiratory system is profoundly compromised by the disease. Oxygen therapy (OT) is then prescribed to restore oxygen levels in the blood. [1] Also, some people with CF (pwCF) can have lung transplantation (LTx) during their life. Association between dependence OT and natural disease progression in pwCF becomes challenging, and it has not been estimated yet. We therefore used the multi-state model to estimate the transition probabilities from being alive without LTx to LTx and to death, and from being alive after LTx to death in pwCF with and without OT. We used 10 years' data from the 35-country European CF Society Patient Registry (ECFSPR). A multi-state regression model was fitted using age as timescale to assess the effects of individual risk factors on transition probabilities. We considered 48,343 pwCF aged 6 to 50 years. OT (HR 5.78, 95%CI: 5.32 - 6.29) and abnormal FEV1 (HR 6.41, 95%CI: 5.28 - 7.79) were strongly associated with the probability of having LTx; chronic infection with Burkholderia cepacia complex (HR 3.19, 95%CI: 2.78 - 3.67), abnormal FEV1 (HR 5.00, 95%CI: 4.11 - 6.08) and the need for OT (HR 4.32, 95%CI: 3.93 - 4.76) showed the greatest association with the probability of dying without LTx. Once pwCF received LTx, OT (HR 1.75, 95%CI: 1.41 - 2.16) and abnormal FEV1 (HR 1.63, 95%CI: 1.18 - 2.25) were the main factors associated with the probability of dying. We also found an association between gross national income and the probability of receiving LTx, which is lower for pwCF living in low- income European countries. Oxygen therapy, as a proxy for disease severity, is associated with poor survival in pwCF, even after LTx. Harmonization of CF care throughout European countries remains of paramount importance.

[1] Elphick HE, Mallory G. Oxygen therapy for cystic fibrosis. Cochrane Database Syst Rev. 2009;(1):CD003884. doi:10.1002/14651858.CD003884.pub3

WP31 Graphical and multistate modelling of home dialysis uptake in patients who need kidney replacement therapy

Solis--Trapala I.*[1], Potts J.[1], Pearse C.[2], Damery S.[3], Fotheringham J.[4], Hill H.[4], Phillips--Darby L.[1], Iestyn W.[3], Davies S.[1]
[1]Keele University ~ Staffordshire ~ United Kingdom, [2]MRC Lifecourse Epidemiology Centre ~ Southampton ~ United Kingdom, [3]University of Birmingham ~ Birmingham ~ United Kingdom, [4]University of Sheffield ~ Sheffield ~ United Kingdom

Introduction: People with kidney failure can have dialysis treatment at home or at an out-patient unit. UK National Guidelines recommend home therapy (HT) treatment; however, despite efforts to increase the use of HT, uptake is poor. The real-world evidence "Inter-CEPt" study [1] used a mixed- methods approach to identify modifiable factors to produce a package of interventions to increase HT uptake. Objectives: 1) Describe direct and indirect paths of associations of patient- and centre-level factors with HT uptake. 2) Model the patient treatment history and mortality. Methods: A chain graph model for HT uptake one year from starting kidney replacement therapy (KRT) was specified using the insights obtained from an ethnographic study which also informed the design of a national survey of renal units across England. The graphical model was estimated using data from the survey linked to patient-level data from the UK Renal Registry (UKRR). Parametric multistate models were developed to estimate instantaneous rates and probabilities of transitions between home therapies, in-centre hemodialysis (ICHD), transplantation, and to death for 93,473 patients starting KRT in England 2005-2019. Results: Renal centres that fostered opportunities for staff to engage in research, had run quality improvement projects on home dialysis, had hosted a home dialysis road show and offered assisted peritoneal dialysis were directly associated with increased odds of therapy uptake. Patients in centres with perceived stress on staff capacity to deliver dialysis had decreased odds of therapy uptake. There was evidence of inequality trends on the uptake of home therapies with Asian, Mixed, Black and Other Ethnic Groups having lower odds of being on home therapy than White patients. Five years after starting KRT, patients in the most socioeconomically deprived group had higher probability of moving from HT to ICHD or dying and lower probability of receiving a transplant than those in the least deprived group. Conclusions: By linking patient-level data from the UKRR with centre-level survey data and developing a chain graph model, we were able to identify candidate factors for subsequent intervention development, and highlighted health inequalities in treatment and outcomes using a multistate model.

[1] Tshimologo, M., Allen, K., Coyle, D., Damery, S., Dikomitis, L., Fotheringham, J., Hill, H., Lambie, M., Phillips-Darby, L., Solis-Trapala, I., Williams, I., Davies, S.J., 2022. Intervening to eliminate the centre-effect variation in home dialysis use: protocol for Inter-CEPt—a sequential mixed-methods study designing an intervention bundle. BMJ Open 12, e060922. https://doi.org/10.1136/bmjopen- 2022-060922

Poster Sessions

WP32 Sglt2i better slow down chronic kidney disease progression in type 2 diabetes patients: a multistate analysis.

Tansawet A.*[1], Looareesuwan P.[2], Thakkinstian A.[2]
[1]*Navamindradhiraj University ~ Bangkok ~ Thailand,* [2]*Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University ~ Bangkok ~ Thailand*

Guidelines have recommended tight glycemic control in type 2 diabetes (T2D) patients to slow down the disease process, which could be achieved by combining multiple medications. However, how second-line medications affect chronic kidney disease (CKD) progression has seldom been studied. Data from T2D patients with CKD stage 3 were retrieved for analysis. Groups of second-line diabetes medications including sulfonylurea (SU), thiazolidinedione (TZD), dipeptidyl peptidase-4 inhibitor (DPP4i), and sodium-glucose cotransporter 2 inhibitor (SGLT2i) were the medications of interest and assessed as a separate independent factor. Medications along with covariates that were significantly retained in each transition-specific parametric survival regression (Weibull distribution) were integrated into a multistate model. This multistate model had two intermediate states (i.e., CKD stage 4 and 5) and one absorbing state (i.e., death). Subsequently, transition probabilities in each state and probability difference were estimated from 1000-replicated bootstrapping for each covariate pattern of interest. Data from 14,834 T2D patients with CKD stage 3 were used in the analysis. Apart from medications, body mass index, hypertension, hemoglobin A1C level, high-density lipoprotein cholesterol level, cardiovascular disease (CVD), peripheral vascular disease (PVD), and diabetic retinopathy (DR) were included in the multistate model. In patients without macro and microvascular complications, 90.7% of patients receiving SGLT2i still stayed in CKD stage 3 at 10-year follow-up compared to 73.5%, 78.2%, and 70.1% of patients receiving SU, TZD, and DPP4i, respectively. At 10 years, SGLT2i yielded a significantly lower probability of progression to CKD stage 5 and death with the difference (95% confidence interval) of -4.5% (-5%, -4.1%), -3.3% (-3.7%, -2.9%), and -4.9% (-5.4%, -4.5%) for CKD stage 5; and -7.4% (-8%, -6.8%), -5.5% (-6.1%, -4.9%), and -11.9% (-12.6%, -11.2%) for death, compared with SU, TZD, and DPP4i, respectively. Significant effects of SGLT2i compared with other medications were also observed in patients with CVD, PVD, and DR. SGLT2i yielded the lowest probability of CKD progression compared with other second-line diabetes medications.

WP33 Sars-cov-2 trends in italy, germany and school opening during the omicron variant: a quasi-experimental study

Bellerba F.*[1], Bardeck N.[2], Böhm M.[2], Raimondi S.[1], Abecasis A.[3], Pirkl M.[2], D'ecclesiis O.[1], Incardona F.[4], Gandini S.[1]
[1]*European Institute of Oncology ~ Milan ~ Italy,* [2]*University Clinics of Cologne ~ Cologne ~ Germany,* [3]*NOVA University of Lisbon ~ Lisbon ~ Portugal,* [4]*EuResist Network ~ Rome ~ Italy*

We investigated the potential impact of school reopening on SARS-CoV-2 transmission in Italy and Germany in autumn 2022. The investigation faced several methodological challenges, including the selection of the most appropriate statistics to investigate the link between the infection's spread and the different school reopening dates. The reproduction number (R) has been widely employed by countries during the pandemic to estimate the transmission dynamics of the virus and for the public health decision making. In literature, there are two types of R: the instantaneous reproduction number (Rt) for real-time estimation and the case reproduction number (Rc) for retrospective analysis. Both metrics heavily rely on the parametrization of the generation time. Under appropriate temporal assumptions, Rt is approximately 1 when the model agnostic growth rate (rt) is 0[1]. We confirmed that the parametrization regarding Omicron variant was appropriate for our timeframe. We used daily symptomatic cases in Italy and weekly confirmed cases in Germany. We pooled data from regions/states that opened on the same day and compared the Rc curves by age groups in relation to school opening dates. We calculated the time from school opening to the day of increase or increase in velocity of Rc. We employed a staggered difference-in-differences analysis[2] to assess whether the school reopening significantly affected the trend of infections while considering the different dates of reopening. We used different metrics (daily and weekly log cases, rt, Rc, log and original, daily and weekly) by country, to verify stability of results. Considering the analysis based on Rc, we found a significant decrease in Rc following school openings in the 6-19 years population (Overall average treatment effect for the treated subpopulation (O-ATT):-0.69 [95%CI:-0.99;-0.39] for Italy; O-ATT:-0.28 [95%CI:-0.32;-0.243] for Germany). No difference compared to the situation before the opening of schools in the adult population (O-ATT:-0.02 [95%CI:-0.05; -0.02] for Italy; O-ATT: -0.06 [95%CI:-0.09;-0.03] for Germany). Multivariable models adjusting for confounders confirmed these results. The increasing trend of the Sars-Cov-2 in autumn 2022 appeared to be driven mainly by the geographical location, seasonal changes and overall population behavior than by school openings.

[1] Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative than reproduction numbers? J R Stat Soc Ser A Stat Soc. 2022 May 26:10.1111/rssa.12867. doi: 10.1111/rssa.12867.
[2] Callaway B, Sant'Anna PHC. Difference-in-Differences with multiple time periods. J Econometrics. 2021;225(2):200-230. ISSN 0304-4076. doi: 10.1016/j.jeconom.2020.12.001.

WP34 Mental health of family members and friends of covid-19 patients: an observational cohort study

Lovik A.*[1], González--Hijón J.[1], Valdimarsdóttir U.[2], Fang F.[1], Covidment C.[1]
[1]*Karolinska Institutet ~ Solna ~ Sweden,* [2]*University of Iceland ~ Reykjavík ~ Iceland*

Little is known regarding the mental health impact of having a family member and/or a close friend with COVID-19 of different severity. The aim of the present study was to assess depressive and anxiety symptoms of family members and friends of people diagnosed with COVID-19. Methods. The study included five prospective cohorts from four countries (Iceland, Norway, Sweden, and the UK) with self-reported data on COVID-19 and symptoms of depression and anxiety during March 2020-March 2022. We calculated the prevalence ratio (PR) of depression and anxiety in relation to having a significant person with COVID-19 and performed a longitudinal analysis in the Swedish cohort to describe the temporal pattern of the results. Results. 162,237 and 168,783 individuals were included in the analysis of depression and anxiety, respectively, of whom 24,718 and 27,003 reported a significant person with COVID-19. Overall, the PR was 1.07 (95% CI: 1.05-1.10) for depression and 1.08 (95% CI: 1.03-1.13) for anxiety in relation to having a significant other with COVID-19. The respective PRs for depression and anxiety were 1.04 (95% CI: 1.01-1.07) and 1.03 (95% CI: 0.98-1.07) if the significant person was never hospitalized, 1.15 (95% CI: 1.08-1.23) and 1.24 (95% CI: 1.14-1.34) if hospitalized, 1.42 (95% CI: 1.27-1.57) and 1.45 (95% CI: 1.31-1.60) if ICU admitted, and 1.34 (95% CI: 1.22-1.46) and 1.36 (95% CI: 1.22-1.51) if the significant person died. Individuals reporting a significant person hospitalized, ICU admitted, or deceased showed higher prevalence of depression and anxiety during the entire 12 months after the diagnosis of the significant person. Having a significant person with COVID-19, especially severe COVID-19, is associated with an increased risk of depression and anxiety. These findings motivate enhanced clinical surveillance of relatives and friends of patients suffering severe COVID-19 or other potential future pandemics.

WP35 Life expectancy changes and covid-19 vaccination campaign impact on all- cause mortality in italy

Nova A.*, Fazia T., Zanella M., Bernardinelli L.
Università ~ Pavia ~ Italy

COVID-19 pandemic triggered an unprecedented rise in mortality that reversed the long-term increasing life expectancy (LE), with implications for public health. Given the different geographical and temporal spread of the epidemic in the Italian regions, a high heterogeneity was observed in LE changes from 2019 to 2021. COVID-19 vaccination campaign, started in 2021, played a key role in improving LE, as the benefits of vaccines are well established in the literature [1]. However, despite the positive results, there remains skepticism around vaccination in terms of potential side effects or lack of efficacy [2]. Therefore, we aimed to investigate, up to 31 December 2022, LE changes in the Italian regions and the impact of the vaccination campaign in terms of number of all-cause deaths averted. Annual period life tables were used to estimate LE by gender. Compared to 2021, an overall national increase in LE was found for both males (+0.46 years, 73% driven by 60-79 age group) and females (+0.13 years, 39% driven by 60-79 age group). Centre and South of Italy's regions generally showed higher LE increases compared to the regions in North. However, LEs has not yet reached 2019 levels. We then assessed the impact of COVID-19 vaccination rates by estimating numbers of averted all- cause deaths simulating a counterfactual scenario in which vaccines were not delivered. Using publicly available data we implemented an over-dispersed Poisson regression model, which included weekly age and sex-specific proportions of individuals vaccinated with two doses and, separately, with booster dose in each region. Variables such as variants' frequency, demographics, climate, restrictions levels and weekly COVID-19 detected infections were also considered as potentially relevant confounders. We estimated that 354,660 all-cause deaths (95% CI: 225,634-503,246) were prevented due to the implementation of the vaccination campaign.
Our findings highlight an increase in LE towards pre-pandemic levels in almost all Italian regions. We further provide support to a positive and safe impact of the COVID-19 vaccination program which allowed to prevent a considerable number of all-cause deaths and consequently to improve LE.

[1] Watson OJ, Barnsley G, Toor J, et al. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. Lancet Infect Dis. 2022 Sep;22(9):1293-1302.
[2] Dubé E, MacDonald NE. COVID-19 vaccine hesitancy. Nat Rev Nephrol. 2022 Jul;18(7):409-410.

WP36 Agent-based models for the covid-19 epidemic: understanding the effects of determinants of disease spread

Roy P.*
for the CovDyn Group (Covid Dynamics). 1) Université de Lyon, F-69000 Lyon, France 2) Université Claude Bernard Lyon 1, F- 69622 Lyon, France 3) CNRS UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, F-69100 Villeurbanne, France 4) Service de Bi

The outbreak of the SARS-CoV-2 virus, enhanced by rapid spreads of variants, has caused a major international health crisis. An agent-based model was designed to simulate the evolution of the epidemic in France taking into account the main determinants of the disease's spread. The first model [1] aimed to predict the evolution of the COVID-19 epidemic over an 18-month period under the combined effects of non-pharmaceutical interventions (NPIs) and vaccination strategies, taking into account the increasing proportion of the Alpha variant at the beginning of 2021. An extension [2] was proposed to predict COVID-19 mortality over a period of 400 days (until the end of 2022) based on several scenarios combining vaccination strategies, NPIs, and degrees of decline in vaccine immunity over time. It seemed mandatory to vaccinate the highest possible proportion of the population within 12, or better, 9 months. The race against the epidemic and the variants of the virus was a question of vaccination strategy. The proposed simulation approach allowed a better understanding of the combined effects of the various determinants of the disease. The model dealt successfully with single measures or complex combinations. It can help choosing them according to future epidemic features, vaccination extensions, and population immune statuses. As the COVID-19 epidemic is not yet under full control, the proposed approach could be useful for planning future vaccination campaigns. Representative data on the prevalence of infection are needed to anticipate rapid changes in disease patterns, so that the results of analyses of these data can be combined with the available results of other analytical studies and clinical trials. By refining and above all simplifying the use of such a model, health authorities could have at their disposal a tool for analysing local or national situations and for assisting in the decision to control epidemics.

[1] Pageaud S, Pothier C, Rigotti C, Eyraud-Loisel A, Bertoglio JP, Bienvenüe A, Leboisne N, Ponthus N, Gauchon R, Gueyffier F, Vanhems P, Iwaz J, Loisel S, Roy P, On Behalf Of The Group CovDyn Covid Dynamics. Vaccines (Basel). 2021 Dec 10;9(12):1462.
[2] Pageaud S, Eyraud-Loisel A, Bertoglio JP, Bienvenüe A, Leboisne N, Pothier C, Rigotti C, Ponthus N, Gauchon R, Gueyffier F, Vanhems P, Iwaz J, Loisel S, Roy P, On Behalf Of The CovDyn Group Covid Dynamics. 2022 Nov 28;10(12):2033.

WP37 A combined criterion for dose finding in phase i clinical trials

Alam M.I.*, Sarwar T.
University of Dhaka ~ Dhaka ~ Bangladesh

A popular design used in phase I clinical trials is the continual reassessment method (CRM) (O'Quigley et al., 1990). The CRM allocates that dose to the patients with the estimated probability of toxicity closest to the target. The maximum tolerated dose (MTD) in a phase I trial can also be found under the framework of D-optimum design. The D-optimum design is unpopular among clinicians since it often allocates doses to patients from extremes of the design region. But it can determine the MTD more precisely than the CRM in many cases (Alam, 2016). This paper aims to check whether any bridge between the CRM and D-optimum design is possible. More specifically, we intend to investigate a combined criterion as a dose-optimization tool. The proposed combined criterion is a linear combination of CRM and D. Since toxicity is the endpoint, we use a two-parameter logistic regression model. The Bayesian technique is used to estimate the model parameters since the maximum likelihood estimates cannot be obtained until enough data are collected. Six plausible dose-response scenarios are investigated in a simulation study. Each of the patients in a trial receives doses following the combined criterion. Starting with the lowest dose, each trial continues until it reaches n cohorts. Two performance metrics, decision efficiency and sampling efficiency, are used to evaluate the designs. The simulation results show that the combined criterion has higher decision efficiency than the CRM in most cases. The sampling efficiency of the combined design is much better than the D-optimum design in all six scenarios. Compared to the CRM, the sampling efficiency of the combined design is also slightly higher. So the combined design allocates doses to cohorts more sensibly.

O'Quigley J., Pepe M., Fisher L. (1990). Continual reassessment method: A practical design for phase I clinical trials in cancer. Biometrics 46(1): 33–48.
Alam M.I. (2016). A comparison between the continual reassessment method and D-optimum design for dose finding in phase I clinical trials. Biometrical Letters 53(2): 69–82

Poster Sessions

WP38 High-order asymptotic intervals for the toxicity probability in phase i clinical trials

Althobety A.*
King AbdulAziz University ~ Jeddah ~ Saudi Arabia

Phase I dose-finding studies aim to develop a safe and efficient dose of a new treatment. The selected dose is the maximum tolerated dose (MTD), which is the dose that has a toxicity rate close to a prespecified level of toxicity. This study concerns the continual reassessment method (CRM) proposed by O'Quigley, Pepe and Fisher (1990). It is a dose-finding design that assumes a mathematical model for the relationship between the dose and toxicity probability. It is designed to treat as many patients as possible close to the MTD and sequentially estimate the next recommended dose and its toxicity rate. To support the experimenter's decision, an approximate confidence interval for the toxicity probability can be obtained to give some precision around the estimated toxicity rate. Due to the limited number of treated patients and an insufficient amount of information, inaccurate results could be observed. Therefore, a high-order approximation would be a solution to provide an accurate result. To date, there have not been many attempts to examine the Cornish-Fisher inversion; also, the accuracy of the saddlepoint approximation (Lugannani and Rice, 1980) has not been explored in this context. This study assesses and compares the use of Cornish-Fisher inversion and saddlepoint approximation in improving the coverage levels of the confidence interval for the toxicity probability.
The dose-finding trial based on Elkind et al. (2008) is redesigned by considering the Bayesian CRM to investigate the behaviour of high-order approximate interval of the toxicity rate. Simulation studies are conducted by sequentially estimating the next recommended dose. Finally, approximate intervals for the toxicity rate at the MTD are obtained using the Cornish-Fisher inversion and saddlepoint approximation. Results show that the coverage levels are slightly improved according to the Cornish- Fisher approximation, whereas the saddlepoint approximation resulted in anticonservative levels. This investigation highlights the usefulness of studying high-order approximations to achieve accurate coverage levels and offers new insights into the approximations' adequacy in practical work. However, in dose-finding trials, comparing the adequacy of the approximations may differ according to the trial's natural and design inputs. Elkind, M. S., Sacco, R. L., MacArthur, R. B., Fink, D. J., Peerschke, E., Andrews, H., and Cheung, K. (2008). The Neuroprotection with Statin Therapy for Acute Recovery Trial (NeuSTART): an adaptive design phase I dose-escalation study of high-dose lovastatin in acute ischemic stroke. International Journal of Stroke, 3(3), 210-218.

Lugannani, R., and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. Advances in Applied Probability, 12(2), 475-490.
O'Quigley, J., Pepe, M., and Fisher, L. (1990). Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics, 46(1), 33–48.

Poster Sessions

WP39 Bayesian decision analysis for clinical trial design with binary outcome: illustration for ebola virus disease

Belhadi D.*[1], Cho J.[2], Malvy D.[3], Mentré F.[1], Lo A.W.[2], Laouénan C.[1]
[1]Université Paris Cité, Inserm, IAME ~ Paris ~ France, [2]MIT Laboratory for Financial Engineering ~ Cambridge, MA ~ United States of America, [3]Department for Infectious and Tropical Diseases, University Hospital Center Pellegrin ~ Bordeaux ~ France

When designing trials for diseases with high mortality rates and few therapies available, the use of a 5% type I error to calculate sample sizes can be questioned. Bayesian Decision Analysis (BDA) for trial design allows to incorporate the burden of disease when designing trials [1-3]. We aim to adapt the BDA to binary outcomes and to calculate optimal sample sizes and type I errors using as illustration Ebola virus disease treatment trials [4]. We consider a fixed two-arm randomized trial with a binary primary outcome and two types of clinical trial loss: post-trial loss, for not approving an effective treatment or approving a non-effective treatment; in-trial loss, for administrating a non-effective treatment during the trial. We set parameters to fit an Ebola outbreak context using mortality as an outcome: target population size N varying from 50 to 5000; mortality rate p_control from 0.30 to 0.70; mortality reduction ratio (RR) of an effective drug in {0.20; 0.35; 0.50}. The optimal type I error α^* and sample size per arm n* are compared to a standard one-sided test with α_ref=0.025, 90% power and sample size n_ref. We adapted the BDA model to be used with binary outcomes. An R function was developed (R software v.4.2.2). To obtain optimal type I errors and sample sizes, parameters needs to be specified including (N, p_control, RR) and parameters for the treatment side effects and disease severity. Regarding Ebola trials, when N is low, α^* are higher than α_ref and n* are lower than n_ref; e.g. with N=500 and p_control=0.45, n* for RR in {0.20; 0.35; 0.50} are respectively {152; 117; 73} compared to n_ref: {620; 193; 88}. Results differ when N and RR are large: α^* are lower than α_ref and n* are higher than n_ref, especially for high mortality rates. This BDA adaptation aims at helping researchers when designing trials with a binary primary outcome, especially for diseases with high mortality rates and few therapies available. Through our illustration, we explored the influence of target population size, mortality rate and expected treatment effect on the optimal type I error and sample size.

[1] Isakov L, Lo AW, Montazerhodjat V. Is the FDA too conservative or too aggressive?: A Bayesian decision analysis of clinical trial design. Journal of econometrics. 2019;211(1):117–36.
[2] Xu Q, Cho J, Ben Chaouch Z, Lo AW. Incorporating patient preferences and burden-of-disease in evaluating ALS drug candidate AMX0035: a Bayesian decision analysis perspective. Amyotroph Lateral Scler Frontotemporal Degener. 2022 Oct 26;1–8.
[3] Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW. Use of Bayesian decision analysis to minimize harm in patient-centered randomized clinical trials in oncology. JAMA oncology. 2017;3(9):e170123–e170123.
[4] Belhadi D, El Baied M, Mulier G, Malvy D, Mentré F, Laouénan C. The number of cases, mortality and treatments of viral hemorrhagic fevers: A systematic review. PLoS Negl Trop Dis. 2022 Oct;16(10):e0010889.

WP40 Optimal allocation of clusters to sequences in stepped wedge trials with binary outcome data

Laura E.*, Michael G., James W.
Newcastle University ~ Newcastle upon Tyne ~ United Kingdom

In stepped wedge design, clusters are allocated to sequences that specify when the cluster will crossover to the intervention. Much of the methodology regarding the optimal design of stepped wedge trials assumes normally distributed outcome data. However, stepped wedge designs are frequently used in practice with binary outcomes. Accordingly, in this work, by adapting existing methods for individually randomized crossover trials, we combine a simulation study with analytical derivations to take a Bayesian approach to determining optimal sequence allocations for stepped wedge trials with binary data. The analysis of the trial is assumed throughout to be via generalised estimating equations. Prior distributions for model parameters present in the variance-covariance matrix are introduced to estimate a weighted average of the treatment effect variance, which is then numerically minimized by leveraging an efficient routine for multi-dimensional integration. The sequence allocations obtained via the proposed approach are then be compared to those returned by existing results on optimal design for normal data, by applying a suitable normal approximation for the binary outcomes. Through this, the viability of retaining the easier-to-implement normal data results is assessed. Results are currently being synthesized, with early indications being that equal allocation of clusters to sequences remains rarely an optimal design. Furthermore, results suggest that the exact correlation structure may have a notable impact on whether a normal approximation yields near-optimal sequences for binary outcome data. Singh, S.P. and Mukhopadhyay, S. (2016) "Bayesian crossover designs for Generalized Linear Models," Computational Statistics & Data Analysis, 104, pp. 35–50.

WP41 Trial estimands: using the composite strategy for intercurrent events when the outcome variable is continuous

Grobler A.*, Lee K.
Murdoch Childrens Research Institute ~ Melbourne ~ Australia

The updated ICH-E9(R1) guideline identified five strategies to handle intercurrent events(ICE) when defining the estimand of interest[1]. The composite strategy is recommended when the occurrence of the ICE provides information about the effect of the treatment and is implemented by including the ICE in the definition of the outcome. This is straight-forward when the outcome is binary, since the ICE can be incorporated into a composite outcome variable. It is less clear how to implement this strategy if the outcome is continuous. We explore how the composite strategy can be implemented in the context of a continuous outcome variable and discuss implications for the definition of the estimand and statistical analysis. Potential strategies when incorporating the ICE into the outcome are: 1) convert the continuous outcome to a binary outcome, or 2) assign the occurrence of the ICE a value on the continuous scale (e.g. worst outcome). Option 1 changes the estimand, with well-known loss of information and power. With option 2 there may be a natural numerical value for the ICE (e.g. ICE of death on the Karnofsky Performance Scale where death has a value of 0), If there is no natural value, the summary measure defined in the estimand should not be influenced by the specific arbitrary value assigned to the ICE. Summary measures possible for Option 2 are difference in means, medians or trimmed means, or the common odds ratio when summarising ordered categorical outcomes. The associated statistical methods could be rank-based methods, quintile regression, or analysing ordered categorical outcomes using the proportional odds model[2]. We highlight advantages, disadvantages and statistical considerations for each of these, and illustrate the difference that the choice of estimand and analysis approach can have on the sample size and resulting treatment estimates from a real clinical trial. Unless the outcome variable is binary, using the composite strategy to handling ICE raises questions regarding the summary measure in the estimand and the statistical approach to estimate it. The implications of these decisions should be considered when using this strategy in the context of a continuous outcome.

[1] International Council for Harmonisation of Technical Requirements for Pharmaceuticals For Human Use. ICH Harmonised Guideline. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). 2019.
[2] O.N. Keene, Pharmaceutical statistics, 18, 2019, 78-84

WP42 Efficient dose insertion in phase 1 trials: an adaptive approach using model- based design

Lai X.*[1], Cheung K.[2], Yap C.[1]
[1]The Institute of Cancer Research ~ London ~ United Kingdom, [2]Columbia University ~ New York ~ United States of America

Phase I clinical trials commonly aim to determine the maximum tolerated dose (MTD) of a new drug. However, this approach can be problematic when there is limited prior information about the tolerability of pre-specified doses, and the true MTD may be far from any of those doses. Previous work has examined techniques to adaptively insert additional doses when none are acceptable for trials using model-assisted designs, but none has considered dose insertion specifically for model-based designs, such as the continual reassessment method (CRM). Our work considers different strategies to adapt the CRM to allow for dose insertion when necessary. The trial utilises a CRM design to recommend a dose with dose-limiting toxicity (DLT) rate that is closest to the target DLT rate. New doses can be inserted by applying any dose insertion criteria or by the trial's investigators team if needed. After identifying the location of dose insertion, the prior DLT rate for the inserted dose level is determined by our proposed simple or model recalibration method. The prior is updated, and the new dose is used to treat the next cohort. The trial continues until maximum number of patients are enrolled or stopped early according to predetermined rules. We performed simulation studies to evaluate the performance of the approaches. The recalibration method is superior in selecting the inserted dose as MTD and in treating more patients at the inserted dose compared to the simple method. Both performed well in scenarios where dose insertions were needed at intermediate intervals, but the insertion was suboptimal when all preset doses had true toxicity rates lower than the target. However, in the scenario where all preset doses were too toxic, the insertion method was able to insert less toxic doses that were correctly identified as the MTD. The proposed dose-insertion approaches do not rely on the determination of the exact dose through interpolation/extrapolation. The methods demonstrated better performance in terms of MTD selection to other existing dose insertion methods through simulation. It allows flexibility in inserting additional doses for a CRM trial, when required, and thus increases the probability of correctly identifying desirable doses.

[1] Chu, Y., Pan, H., & Yuan, Y. (2016). Adaptive dose modification for phase I clinical trials. Statistics in medicine, 35(20), 3497-3508.
[2] Hu, B., Bekele, B. N., & Ji, Y. (2013). Adaptive dose insertion in early phase clinical trials. Clinical Trials, 10(2), 216-224.

WP43 Incorporation of historical data for basket trials in the early phase

Li W.*[1], Jaki T.[2], Mozgunov P.[1]
[1]MRC Biostatistics Unit ~ Cambridge ~ United Kingdom, [2]University of Regensburg ~ Regensburg ~ Germany

Basket trials are designed to study a single targeted therapy in the context of multiple diseases or disease types sharing common molecular alterations. To achieve concurrent borrowing of information across multiple baskets, several Bayesian methods have been proposed in recent years. Meanwhile, information from historical studies can be incorporated into the current study depending on the commensurability in study designs. In this talk, we will propose Bayesian basket trial designs allowing for concurrent and non-concurrent borrowing. Furthermore, we will evaluate the robustness of the proposed approaches to shift in the historical and current studies. We adopt the methodology of Ibrahim and Chen [1] on power priors to discount information from historical basket(s) while allowing for borrowing between current baskets. We consider two cases: with fixed and random power parameters a0. To implement the latter, we use adapted path sampling for the approximation of scaling constant to mitigate the violation of likelihood principle. This novel power prior Bayesian hierarchical model (PP-BHM) approach assumes exchangeability of treatment effect across historical and current baskets while down-weighting the historical data. Furthermore, we develop an alternative method called modified power prior (MPP)[2], by using historical data discounted via either a fixed or random a0 to inform prior information of treatment effects in the current basket trial. A hierarchical model is used for treatment effect borrowing across current baskets. We will demonstrate the performance of the proposed approaches under different numbers of baskets. We will show how the choice of a fixed a0 impacts operating characteristics under homogeneous and heterogeneous scenarios. We also demonstrate the flexibility of using a random a0. Finally, we will illustrate the proposed approach is considerably more sensitive to heterogeneous cases (i.e. a0 rapidly goes down to zero) than in homogenous cases (i.e. a0 increases very slowly). We encourage increased concentration on the use of historical data for early-phase basket trials. It is crucial to ensure all relevant evidence should be collected and appropriately synthesised to support inferences about contemporary clinical trials and hence to facilitate decision-making of clinical trials and increase benefits for patients.

[1] J. G. Ibrahim, M-H, Chen, Power Prior Distributions for Regression Models. Statistical Science. 2000; 15(1): 46-60.
[2] Y. Y. Duan, E. P. Smith, Evaluating water quality using power priors to incorporate historical information. Environmetrics. 2006; 17:95–106.

WP44 Joint modeling of pharmacodynamic biomarker and safety in ph1 clinical trials in oncology: sanofi experience

Rotolo F.*[1], Chaoui W.[1], Lin J.[2], Wang R.[2], Wang W.[2], Zhang Y.[2]
[1]Oncology Biostatistics, Biostatistics and Programming Department, Sanofi R&D ~ Montpellier ~ France, [2]Oncology Biostatistics, Biostatistics and Programming Department, Sanofi R&D ~ Cambridge ~ United States of America

With the advent of targeted therapies and immunotherapies in oncology, a strictly increasing dose- effect relationship is not necessarily observed, notably in terms of antitumor activity[1]. Also, some recent drugs showed very little safety issues, which calls into question the classical requirement that the recommended dose should show a minimal rate of toxicity to be potentially active and effective. Recent works proposed alternative dose selection models, like Bayesian joint models considering toxicity together with either clinical and/or biological activity[2]. Despite clinical activity is more relevant from a patient perspective, biological activity is likely observed earlier and at lower doses. We implemented this model in a real clinical trial at Sanofi, and investigated the robustness of this joint model to the specification of the prior distributions and of the biomarker response threshold. The THOR-707-101 trial (NCT04009681) included patients in four cohorts, with different schedules of SAR444245 and drug combinations. We applied the bi-variate and three-variate versions of the joint model for toxicity and each of: three mechanism-of-action biomarkers, ctDNA as biological activity biomarker, +/- clinical response. We studied via simulations the impact of the biomarker threshold for biological activity. We evaluated in the application the impact on results of different prior distributions. The choice of the biomarker threshold had in general a minor impact on the recommendation of the dose. The specification of the prior distribution had negligible impact across a wide panel of models. The Bayesian MCMC algorithm allowed including in the analyses also patients with missing information on either margin. In a phase-1 trial with very low toxicity, the contribution of biomarkers can be of great value to select the recommended dose. The use of a joint model allows estimating the dose-effect relationship on both toxicity and activity, together with their association. The estimates of posterior means had quite wide uncertainty, but the use of several biomarkers accounting for different aspects of the biology helped to partially mitigating such drawback. Sensitivity analysis on the biomarker threshold can be useful in application, while the choice of the prior distributions seems without any impact

[1] Colin P, Delattre M, Minini P, Micallef S. (2017) An Escalation for Bivariate Binary Endpoints Controlling the Risk of Overtoxicity (EBE-CRO): Managing Efficacy and Toxicity in Early Oncology Clinical Trials. J Biopharm Stat;27(6):1054-1072. doi: 10.1080/10543406.2017.1295248
[2] Zhang Y, Xu Z, Lin J, H Quan (2022). Dose Finding via Efficacy Biomarkers and Toxicity Endpoints in Immuno-Oncology Clinical Trials. JSM 2022.
Acknowledgements:
The data were previously presented at American Society for Cancer Research (AACR) Annual Meeting 2023, April 14 - 19, 2023, Orlando, FL, USA. Medical writing support for the parent abstract (AACR 2023) was provided by Theresa Carneiro of Sanofi. Editorial assistance for this encore abstract (EADO 2023) was provided by Ajay Francis Christopher of Sanofi. Investigators and contributors: S Fu, G S Falchook, M Barve, M McKean, T J Tan, C Lemech, C E Chee, T Meniawy, N Marina, G Abbadessa, R Meng, H Wang, J Deng.

WP45 Active level set estimation in phase i dose-finding clinical trials

Seno K.*, Matsui K., Matsui S.
Department of Biostatistics, Nagoya University Graduate School of Medicine ~ Nagoya ~ Japan

Many statistical dose-finding designs for phase I clinical trials, such as CRM and BOIN, estimate the maximum tolerated dose (MTD) based on parametric models for the dose-toxicity relationship. Recently, Takahashi et al. proposed a nonparametric design based on Bayesian optimization [1] in the framework of active learning. In this study, we propose a new design based on another active learning approach, called active level set estimation (ALSE) [2], which searches for a set of input points that take values above or below a given threshold level representing the target toxicity rate for an unknown function of the dose-toxicity curve using as few observation points as possible. We model the dose-toxicity curve nonparametrically using a Gaussian process prior. Once data are obtained, the next dose is determined based on the uncertainty in the posterior distribution of the dose-toxicity curve. After classifying the candidate doses into the upper and lower sets based on the target toxicity rate, we expect that MTD lies near the boundary between these two sets. In simulation studies, the proposed design performed as well as or better than the 3+3 design, CRM, and BOIN in many toxicity scenarios in terms of the probability of selecting the correct MTD and that of avoiding overdose administration. The new dose-finding method based on ALSE is expected to be effective for estimating MTD in phase I clinical trials. This method can be extended to more complex dose-finding settings, involving those with multiple agents and multiple response variables.

[1] A. Takahashi and T. Suzuki, Contemporary Clinical Trials Communications, 21, 2021, 100753.
[2] A. Gotovos, N. Casati, G. Hitz, and A. Krause, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, AAAI Press, 2013, 1344-1350.

Poster Sessions

WP46 The use of basket trials in gynaecology

Stocking K.*[1], Wason J.[2], Watson A.[3], Wilkinson J.[1], Kirkham J.J.[1], Vail A.[1]
[1]Centre for Biostatistics, The University of Manchester, Manchester Academic Health Science Centre Manchester, United Kingdom ~ Manchester ~ United Kingdom, [2]Biostatistics Research Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK ~ Newcastle upon Tyne ~ United Kingdom, [3]Department of Obstetrics and Gynaecology, Tameside & Glossop Acute Services NHS Trust, Ashton-Under-Lyne, UK ~ Ashton-Under-Lyne ~ United Kingdom

In recent years, there has been a move towards providing personalised treatments for patients. The interest in targeted research has led to the development of innovative trial designs such as basket trials. Basket trial designs typically evaluate a targeted intervention in multiple disease types, with a common biomarker. In gynaecology it is common that patients with the same diagnosis experience different symptoms. Frequently, these patients receive the same treatment. Previously, we identified that clinical trials in Polycystic Ovarian Syndrome and Endometriosis were not efficient, often making outcome-based eligibility restriction, or measuring and analysing one or more clinical outcomes that were not relevant to all participants [1]. The current research landscape in gynaecology is not fit-for- purpose, and creates research waste by failing to serve the population available. The purpose of the current work is to explore the applicability of basket trial designs for patients with heterogeneous symptoms in gynaecology, and to identify methodological barriers and challenges. We present the development of existing basket trial methodologies for use in settings where patients with the same diagnosis experience heterogeneous symptoms. In this scenario, we consider sub-trials to be defined by patients' symptoms. We conducted a scoping review of the use of basket trials and the statistical methodologies underpinning them. We focussed in particular on the selection of outcome measure(s), where there is not necessarily a clear choice, and the considerations needed for treatment effect/response borrowing between sub-trials when participants in different baskets experience different symptoms and therefore have different outcomes of interest [2]. Our work considers multivariate extensions, i.e. modelling multiple outcomes at once. We present a simulation study which examines efficiency gains, when compared with usual practice, alongside an overview of the current research landscape, interim findings and provide information on ongoing research. Basket trial methodologies have been extended for use outside of oncology, however, are currently underutilised. Gynaecological conditions are strong candidates for a basket trial design, due to their different clinical manifestations.

[1] Stocking K, Watson A, Wilkinson J, Kirkham JJ, Vail A. A systematic review of methodological approaches used by Cochrane reviews in gynaecology, and their component trials, to incorporate diversity in bothersome symptoms. ICTMC 2022: 6th International Clinical Trials Methodology Conference. https://doi.org/10.5281/zenodo.7741866
[2] Zheng H, Wason JMS. Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. Biostatistics. 2022;23(1):120-35.

WP47 Why do mental health trials only analyse a single primary outcome? We can do better

Stringer D.*, Carter B., Emsley R.
Department of Biostatistics and Health Informatics, King's College London ~ London ~ United Kingdom

Randomised controlled trials (RCTs) in mental health often analyse and report a large number of outcomes, over multiple timepoints. Outcomes that are correlated, and which may measure similar underlying constructs, are typically analysed separately, for example, subscales of a psychometric questionnaire. This approach is inefficient; multivariate modelling techniques may offer increased power whilst still estimating unbiased marginal treatment effects. The objective is to identify and evaluate methods for multivariate modelling in the trial setting, extending previous work by Vickerstaff et al [1] to include repeated measures. Firstly, we carried out a systematic review to assess the evidence base for the analysis and reporting of multiple outcomes in mental health RCTs, comparing approaches with current CONSORT and other regulatory guidance. The review included all mental health trials published in 2019 and 2020 in 9 leading mental and general health journals. We found most trials analysed a single primary outcome, with other outcomes defined as secondary in separate models, and gave no consideration to multiplicity or correlated measures. The median number of outcomes reported was 8 (IQR 6). Current guidance cautions to limit the number of endpoints and that secondary outcomes should be treated as "exploratory" or "supportive". Secondly, we carried out a literature review of methods to jointly analyse multiple outcomes. Several appropriate methods were identified including latent variable and multilevel modelling [2]. The latent variable approach models the correlation between outcomes as a common latent factor. Multilevel models use random effects to model the multivariate hierarchical structure. Thirdly, we will conduct a Monte Carlo simulation study to explore the proposed methods in a longitudinal setting by assessing bias, efficiency, and power under different scenarios. We expect these methods to be unbiased, and show increased efficiency when measures are correlated or there is substantial missing data. We will show these methods can be fit using structural equation modelling in Stata. We show that methods to analyse trial outcomes multivariately increase power. These findings may offer increased efficiency in the design of mental health trials, and therefore, may be cheaper to run, or have a greater chance of showing a conclusive result.

[1] V. Vickerstaff, G. Ambler, and R. Z. Omar, 'A comparison of methods for analysing multiple outcome measures in randomised controlled trials using a simulation study', Biom. J., vol. 63, no. 3, pp. 599–615, Mar. 2021, doi: 10.1002/bimj.201900040.
[2] H. Goldstein, J. Carpenter, M. G. Kenward, and K. A. Levin, 'Multilevel models with multivariate mixed response types', Stat. Model., vol. 9, no. 3, pp. 173–197, Oct. 2009, doi: 10.1177/1471082X0800900301.

WP48 On controlling the type i error rate in an interrupted group sequential trial with interim analyses

Yarahmadi A.*, Stallard N.
University of Warwick ~ Coventry ~ United Kingdom

Motivated by the experience of COVID-19 trials, we consider group sequential clinical trials in which studies may be terminated at some future date when they become infeasible. If the trial has included interim analyses, then it may be impossible to prove that the decision to terminate the study has been made independent of the observed data, and therefore that the type I error rate has not been inflated. Moreover, in some extreme circumstances, the sample size calculation may also be challenging due to the lack of knowledge about likely case numbers. We show how to ensure type I error rate control using the conditional error approach as described by [3] to adjust the critical value for final analysis following termination of the study. As an alternative method to overcome challenges associated with sample size calculation, we propose using a test based on the sequential probability ratio test ([2], [1]) using a spending function to construct a test that has no maximum sample size. We first demonstrate how to construct the tests described above. Then, using extensive simulation studies we assess the behaviour of the type I error rate, as well as the power in all proposed designs, correspondingly.

[1] Siegmund, D. (1985). Sequential analysis: tests and confidence intervals. Springer Science & Business Media.
[2] Wald, A. (1947). Sequential Analysis. John Wiley and Sons, 1st edition.
[3] Wassmer, G. and Brannath, W. (2016). Group sequential and confirmatory adaptive designs in clinical trials. Springer.

Poster Sessions

WP49 Longitudinal functional data analysis with applications in biomechanics

Andrew J.*
University of Galway ~ Galway ~ Ireland

Functional Data Analysis (FDA) is the analysis of data that are in a form of a function, that is, that is, a real valued function in a compact interval I = [0,T] on a real line. FDA can easily handle high dimensional data by approximating the infinite dimensional space by a finite one, and can easily handle missing data problems. Current advancements in technology have led to fast and accurate procedures for capturing/measuring data, leading to generation of massive data which brings many opportunities for research and data analysis. We explored the data from 18 athletes running 400m under normal conditions (healthy) and again under fatigued conditions with sensors mounted on their left and right shanks, thighs and one on the lumbar spine (five total sensors). Each sensor captured 16 signals 256 times per second: acceleration in X, Y and Z directions, euler pitch, roll, and yaw, angular velocity in X, Y, and Z directions, magnetometer in X, Y, and Z directions, and quaternion in W, X, Y, and Z. The long record of any of the 16 signals can be broken into strides forming a series of strides arising longitudinally, and for each sensor, all the 16 signals together bring multivariate data structure. Win the present study we make use of the Functional Mixed Model (FMM) with functional response – Euler pitch angle from a sensors mounted on the shanks capturing the knee angle (measured in degree), and scalar covariates – fatigue (normal vs fatigue) and leg (right vs left). This model enables us to capture the correlation between strides of the runner, mean function of the runner, and the grand mean function of all the runners. This may help to identify subjects (runners) with similar and different mean functions. In this poster we will present preliminary results showing the difference between normal and fatigue and between the right and left legs. Our next focus will be to utilise the entire space of the data by exploring the additional benefits of analysing all the variables together by extending our model to multivariate FMM. Wang, Jane-Ling, Jeng-Min Chiou, and Hans-Georg Müller. "Functional data analysis." Annual Review of Statistics and its application 3 (2016): 257-295

WP50 Growth mixture modelling to identify sodium patterns. Application to icu data in traumatic brain injury (TBI)

Graziano F.*, Rebora P., Galimberti S.
Bicocca Bioinformatics Biostatistics and Bioimaging Center B4, School of Medicine and Surgery, University of Milano- Bicocca ~ Milano ~ Italy

The growing accessibility to electronic data has sparked interest in modeling the changes in biomarkers over time, and the potential to detect clusters of patients who share particular patterns is attractive. However, this task is complicated by various factors, such as dropouts, correlations between observations, and heterogeneous trajectories of measurements. The Growth Mixture Modelling (GMM) is a highly flexible method accommodating for the aforementioned complexities able to capture latent subgroups of individuals who share a common profile over time. We explore this approach to identify sodium patterns and baseline predictors of patterns group membership in the context of traumatic brain injury (TBI). Patients from the CENTER-TBI cohort admitted to Intensive Care Unit (ICU) who had sodium measure for at least three days within the first week were included in the analysis. Candidate models were investigated by varying the number and the shape (linear or modelled by splines) of trajectories, and by the inclusion of baseline covariates on trauma severity (e.g. age, pupil's reactivity and Glasgow Coma Scale). The best model was chosen by AIC, high entropy, and minimum acceptable class size (e.g., ≥5%). Logistic regression was used to assess the association between identified patterns and poor neurological outcomes at 6 months. A total of 1376 patients, with 7040 sodium measurements, fulfilled the study criteria. The final model included three sodium trajectories and a spline term with three nodes for the growth (smallest AIC, entropy>0.7 and smallest class size of 18%) and age, pupil's reactivity and Glasgow Coma Scale were select as predictors of trajectory group membership. The classes were labelled by the degree of sodium increase: "fast-increase" (class-1, 22%), "moderate-increase" (class-2, 60%) and "stability" (class-3, 18%). In the regression model, a significant association on poor outcome was found only for class-1 compared with class-3 (OR=4.9,95%CI=2.73-6.81). This study confirmed the flexibility of GMM in identifying distinct subgroups within a population based on growth patterns, even accounting for individual differences. In TBI patients admitted to ICU, severity was a predictor of their sodium trajectory. The fast increase trajectory was significantly associated with unfavorable outcome reflecting more severe patients and, probably, more aggressive treatment. Proust-Lima, Amieva and Jacqmin-Gadda (2013). Analysis of multivariate mixed longitudinal data : a flexible latent process approach, British Journal of Mathematical and Statistical Psychology 66(3): 470-87.

WP51 Uncertainty computation at finite distance in nonlinear mixed models: evaluation of a new bayesian method

Guhl M.*[1], Bertrand J.[1], Guedj J.[1], Mentré F.[2], Comets E.[3]
[1]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, F-75018 Paris, France ~ Paris ~ France, [2]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME and AP-HP Hôpital Bichat, Département d'Epidémiologie Biostatistiques et Recherche Clinique ~ Paris ~ France, [3]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME and Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085 ~ Paris ~ France

The standard error (SE) of the maximum likelihood estimate (MLE) of the population parameter vector in nonlinear mixed effect models (NLMEM) is usually estimated as the inverse of the Fisher Information Matrix (FIM). However, at finite distance, the FIM can underestimate the SE of NLMEM [1]. As the limit distributions of the MLE and the maximum a posterior estimator in a Bayesian framework are equivalent, the standard deviation of the posterior distribution, obtained in Stan via the HMC algorithm, has been shown to be a proxy for the SE [1]. Here, we develop a similar method using the Metropolis Hastings (MH) algorithm and implement it in the saemix R package. We assess it with different simulation sets and a real dataset. Our simulation study used a one-compartment pharmacokinetic model with linear absorption and elimination, first with N=150 patients with n=10 samples per patient, then with N=12 and n=3. Evaluation was based on MH acceptation rates, boxplots of SE (the target being the empirical SE obtained with SAEM) and 95% coverage rates computed over 1000 datasets. We compared our method to the FIM (Asympt), the HMC algorithm in Stan (Post) and the Sampling Importance Resampling method (SIR). For N=150 and n=10, all methods performed well. For N=12 and n=3, Asympt, SIR and MH underestimated the SE. Coverage rates were below the prediction interval. MH gave improved results when inflating the variance of the kernel. The SE were overestimated by Post, giving coverage rates over or under the target. Further work is needed to investigate suitable priors. Acceptation rates of MH were between 15% and 40%, but decreased in additional simulations with increased variability or strong correlations. We applied this method to model the evolution of a clinical score in patients hospitalised for Covid19, using real data from the Discovery trial [2]. MH gave SE lower than Asympt, Post and SIR. Acceptation rates were critically low, linked to the more complex variability structure with variances higher than 100% and strong correlations. Further work is needed to calibrate our implemented method on challenging settings.

[1] Loingeville F, Bertrand J, Nguyen T, Sharan S, Feng K, Sun W, Han J, Grosser S, Zhao L, Fang L, Möllenhoff K, Dette H, Mentré F, New model-based bioequivalence statistical approaches for pharmacokinetic studies with sparse sampling, AAPS J 2020.
[2] Ader F, Bouscambert-Duchamp M, Hites M, Peiffer-Smadja N, Poissy J et al., Remdesivir plus standard of care versus standard of care alone for the treatment of patients admitted to hospital with COVID-19 (DisCoVeRy): a phase 3, randomised, controlled, open-label trial, Lancet Infect Dis 2022.

WP52 Time series analysis on historical death data of Milan from the Milan registers (1452-1845)

Luconi E.*[3], Marano G.[3], Nodari R.[1], Comandatore F.[1], Galli M.[2], Boracchi P.[2], Biganzoli E.[2]
[1]Romeo ed Enrica Invernizzi Paediatric Research Centre, Department of Biomedical and Clinical Sciences (DIBIC), University of Milan ~ Milano ~ Italy, [2]Department of Biomedical and Clinical Sciences (DIBIC) & DSRC, University of Milan ~ Milano ~ Italy, [3]Department of Biomedical and Clinical Sciences (DIBIC), University of Milan ~ Milano ~ Italy

Giuseppe Ferrario (1802-1870) was an early medical statistician graduated in medicine and surgery. The second volume of his opera, entitled "Statistica medica di Milano dal secolo 15. fino ai nostri giorni" [1], reports detailed information about Milan. It contains tables of the monthly deaths in Milan (1452-1845). In this long period, events such as famine, epidemics, migrations, and wars happened. Therefore, studying historical time series can help to identify significant events, periods with similar behavior, and cyclicities (seasonality and longer ones). Cycles could be associated with recurrent diseases such as smallpox.
Data earlier than 1581 were excluded because of incompleteness. The time series (monthly deaths) was first divided into sub-periods (as customary for long time series) of about fifty years. The analysis was performed in two steps: In each period, firstly, Kernel smoothing (local and global) was applied to identify possible trends; subsequently, the E-divisive method was used to identify changepoints, and finally, outliers were identified with Chebichev inequality. The R package envoutliers [2] was used because it has been developed on environmental data affected by several external variables. The impact and sometimes the presence of major events such as plague, other epidemics, and famine are unknown, especially in the earlier periods decomposition methods were applied to identify relevant patterns in the segments identified by changepoints. The peak of the 1630 plague is clearly shown. In addition, other identified outliers could be related to events described in the opera of Corradi, such as 1693 smallpox, 1719 dysentery, bilious fever, smallpox, and 1743 flu. The decomposition of the data between 1591 and 1626 allows us to suggest a frequency cycle of 44 months, probably due to smallpox, and assess the presence of seasonality. The considered approach performs well for the study of historical time series with unknown impact of external variables and allow the identification of outliers, changepoints, and relevant patterns

[1] G. Ferrario, Statistica medica di Milano dal secolo XV. fino ai nostri giorni escluso il militare: 2. Vol. Vol. 2, Guglielmini & Redaelli, 1840.
[2] M. Čampulová, J. Michálek, P. Mikuška, D. Bokal, Nonparametric algorithm for identification of outliers in environmental data, Journal of Chemometrics, 32(5), 2018, e2997

WP53 The efficacy of a motivational interview in heart failure patients and their caregivers: a dyadic analysis

Occhino G.*[1], Ausili D.L.L.[1], Vellone E.[2], Pucciarelli G.[2], Rebora P.[1]
[1]University of Milano-Bicocca ~ Milano ~ Italy, [2]University of Rome Tor Vergata ~ Rome ~ Italy

The topic of this work focuses on the use of multilevel models on dyadic data. The project was driven by the MOTIVATE-HF, a randomized clinical trial on the effectiveness of a motivational interview (MI) in improving self-care in patients with heart failure (HF) [1]. A secondary aim of the trial was to evaluate the effect of MI on mutuality (i.e. the perceived positive quality of the care relationship between a patient and a caregiver), that was repeatedly measured during follow-up [2]. Patients and caregivers have been recruited between 2014 and 2018 in three Italian centres. After enrolment each patient-caregiver dyad was randomized (1:1:1) on one of the three arms: 1) MI intervention to patients only; 2) MI intervention to patients and caregivers; 3) standard of care for patients and caregivers. Follow-up assessment was performed at 3, 6, 9 and 12 months after enrolment. 510 HF patient-caregiver dyads were enrolled and randomized. To evaluate the effect of MI on mutuality in HF patient-caregiver dyads, a longitudinal dyadic model was applied with the dyad as the unit of analysis. The dependence within measures was accounted for by including a random intercept and slope in the within-dyads level. The between-dyads level included the treatment arm, its interaction with the visit number and caregiver living with patient as covariates. MI did not show any impact on changes in the patient and caregiver mutuality during the follow-up time. In fact, Arms 1 and 2 did not improve significantly more than Arm 3. Interestingly, the same results were also found when applying two separate longitudinal mixed models on patients and caregivers. The interdependence between mutuality of the dyad members, estimated by the dyadic model resulted 0.58 ($p<0.001$), indicating moderate covariation within the dyads. This project shows the application of multilevel models to longitudinal dyadic data in healthcare research. The advantage of using a dyadic model, in which the care dyad is examined as unit of analysis, instead of two separate models, is the possibility to control for the interdependency of scores among dyad members.

[1] Vellone, E., Paturzo, M., D'Agostino, F., Petruzzo, A., Masci, S., Ausili, D., . . . Riegel, B. Contemp Clin Trials, 55, 2017, 34-38.
[2] Lyons KS, Sayer AG, Archbold PG, Hornbrook MC, Stewart BJ. Res Nurs Health, 30, 2007, 84–98.

WP54 Meta-analysis of patient-reported outcomes: methodological proposal and application to rcts in heart failure

Oriecuia C.*[1], Sala I.[2], Tomasoni D.[3], Bagnardi V.[2], Specchia C.[1]
[1]University of Brescia ~ Brescia ~ Italy, [2]University of Milan-Bicocca ~ Milano ~ Italy, [3]ASST Spedali Civili of Brescia ~ Brescia ~ Italy

Randomized clinical trials (RCTs) are increasingly utilizing patient-reported outcomes (PROs) to ensure a patient-focused approach to treatment development and regulation. However, the lack of standardization in PROs collection and analysis in RCTs creates challenges in synthesizing evidences. This study aims to address this issue by proposing a meta-analytical methodology that enables the comparison of continuous PROs between treatment groups. Although meta-analysis of individual patient data (IPD) is considered the most reliable method for synthesizing results from different trials, obtaining IPD can be a challenging resource-intensive process. We suggest, using a validated algorithm1, to reconstruct pseudo-IPD for a continuous PRO at baseline and pre-specified follow-up timepoints, using available published aggregate data. Then, we propose, using the one-stage approach for IPD meta-analysis based on linear mixed-effects regression models1, to obtain pooled estimates of treatment differences at the pre-planned timepoints while adjusting for baseline score, and to explore the interaction between time and treatment effect. As a motivating example, we will apply the proposed methodology to PROs data from published RCTs testing SGLT2 inhibitors (SGLT2i) versus placebo in heart failure patients. Our primary endpoints will be the difference in mean change of the Kansas City Cardiomyopathy Questionnaire2 overall score between the SGLT2i group and the placebo group at 3 and 6 months from baseline. Preliminary results show with the two-stage approach an estimated difference in mean change of 1.92 (SE=0.28) at 3 months and 1.75 (SE=0.46) at 6 months, whereas with the one-stage approach 1.64 (SE=0.19) at 3 months and 1.77 (SE=0.21) at 6 months, respectively. Both suggest SGLT2i slightly improve patients' health status, but the one-stage approach provides lower standard errors. Further simulation-based analyses will assess whether this method also reduces the bias in estimating the treatment effect and will confirm if this approach could improve statistical power. Overall, the use of one-stage approach based on linear mixed-effects models and pseudo-IPD could be a powerful approach for meta-analyzing continuous PROs in clinical studies. Its flexibility and ability to improve statistical power and potentially to reduce bias make it a valuable tool for meta-analysis.

[1] K. Papadimitropoulou, T. Stijnen, R.D. Riley, O.M. Dekkers, S. Le Cessie, Meta-analysis of continuous outcomes: Using pseudo IPD created from aggregate data to adjust for baseline imbalance and assess treatment-by-baseline modification., Res Synth Methods, 11(6), 2020, 780-794.
[2] J.A. Spertus, P.G. Jones, A.T. Sandhu, S.V. Arnold, Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care: JACC State-of-the-Art Review., J Am Coll Cardiol, 76(20), 2020, 2379-2390.

WP55 Modeling a disease-modifying treatment with a piecewise geodesic mixed- effect model

Poulet P.*[1], Jedynak B.[2], Durrleman S.[1]
[1]Inria (ARAMIS), Paris Brain Institute, Sorbonne university ~ Paris ~ France, [2]Portland University ~ Portland ~ United States of America

Disease progression models have been developed in the last decades to describe the evolution of slow progressive disease such as neurodegenerative ones. However such models are not yet built to take treatment effect into account, while new disease-modifying treatments are starting to emerge. We aim at answering the need to predict how treatment will affect disease progression with a dedicated model built upon a disease course mapping model [1]. Our base model is a non-linear mixed-effect model. More especially it describes the progression of patients' markers as a geodesic in a Riemannian manifold. The choice of the manifold amounts to choosing the template curves. This model has been proven to be suited to describe the natural history of the diseases, but is not flexible enough to measure a change of course due to an intervention. Therefore we propose a parametric way to add breakpoints in the trajectory corresponding to the start of a treatment strategy. The model ends up being piecewise geodesic between breakpoints. We propose a method to jointly estimate the treatment effect and the model parameters using a modified Monte-Carlo Markov Chain Stochastic Approximation Expectation Maximization algorithm. We show that our model is able to recover ground truth on simulated data and we provide experimental evidence for the required number of patients and visits in a predefined scenario. We then apply our method to an Alzheimer's disease clinical trial but did not show a significant disease- modifying effect. We proposed a theoretical model to account for a disease-modifying treatment effect in a disease progression model. We propose a method for its estimation and showed that it recovers ground truth in simulations.

[1] J.-B. Schiratti, S. Allassonnière, O. Colliot, et S. Durrleman, « A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations », J. Mach. Learn. Res., vol. 18, n° 1, p. 4840-4872, janv. 2017.

WP56 Safety signal detection in biochemical measures in randomized clinical trials, a bayesian mixed effect model

Revers A.*, Hoogland J., Hof M., Zwinderman A.
Amsterdam UMC location University of Amsterdam ~ Amsterdam ~ Netherlands

Clinical trials often evaluate the safety of new pharmaceutical treatments through the use of numerous biochemical measures, which reflect the functions of different organs and vital systemic processes. These biochemical variables are repeatedly measured when the trial has a substantial follow-up time. The statistical analysis of these biochemical measures poses several challenges, particularly in dealing with multiple biochemical measures of interest and in determining when to specify a signal as a safety signal of interest. As biochemical measures are typically correlated, since dysregulation of an organ/systemic process may show up in several biochemical measures, it may be useful to jointly analyze them. Jointly analyzing them could lead to a gain in statistical power as well as a reduction of the false discovery rate. To do so, we use a hierarchical Bayesian approach that links typical mixed-effects models for the different (repeated) biochemical measure through their treatment effect parameters. Hence, the estimation of the treatment effect parameters of interest gains in precision by sharing information between groups of biochemical measures. This method for analyzing biochemical measures is an extension of a recently developed method for the statistical analysis of large number of counts of adverse effects occurring in clinical trials [1]. With respect to safety signal detection, several outcomes based on overall treatment effects and patient-specific signals are explored. The Western Electric rules of statistical process control partly inspire these outcome measures. As an illustrative example, data from a trial in chemoradiation patients with resectable pancreatic cancer is used. Overall, the proposed method provides a robust statistical approach for the detection of safety signals in clinical trials with repeatedly measured biochemical safety variables.

[1] Revers, A., Hof, M. H., & Zwinderman, A. H. "BAHAMA: A Bayesian Hierarchical Model for the Detection of MedDRA®-Coded Adverse Events in Randomized Controlled Trials." Drug safety vol. 45,9 (2022): 961-970.

WP57 Joint model for multiple longitudinal responses with informative time measurements

Sousa I.*
University of Minho ~ Braga ~ Portugal

In longitudinal studies individuals are measured repeatedly over a period of time for a response variable of interest. In classical longitudinal models the longitudinal observed process is considered independent of the times when measurements are taken. However, in medical context it is common that patients in worst health condition are more often observed, whereas patients under control do not need to be seen so many times. Therefore, longitudinal models for data with this characteristic should allow for an association between longitudinal and time measurements processes. In this work we consider a response longitudinal variable with Gaussian distribution. We propose a model where the follow-up time process is stochastic. The model is described through the joint distribution of the observed process and the follow-up time process. Estimation of model parameters is through maximum likelihood. We conducted a simulation study of longitudinal data where model parameter estimates are compared, when using the model proposed and ignoring the association between processes. Finally, the model proposed is applied to a real data set when monitoring for biomarkers CEA and CA15.3 on breast cancer progression. In this case the follow-up time process should be considered dependent on the longitudinal outcome process. Results are presented showing that, ignoring the latent process of time measurements brings bias results when the collected time points are associated with the observed process. McKeigue P. (2022), Fitting joint models of longitudinal observations and time to event by sequential Bayesian updating, Statistical Methods in Medical Research, 31 (10), 1934-1941. Asar O., Bolin D.,Diggle P.J., Wallin J. (2020), Linear mixed effects models for non-Gaussian continuous repeated measurement data, Journal of the Royal Statistical Society Series C-Applied, 69 (5), 1015-1065 Szczesniak R.D., Su W., Brokamp C., Keogh R.H., Pestian J.P., Seid M., Diggle P.J., Clancy J.P. (2020), Dynamic predictive probabilities to monitor rapid cystic fibrosis disease progression, Statistics in Medicine, 39 (6), 740-756

Poster Sessions

WP58 Multimodal prediction of echocardiography prescription

Bailly A.*[1], Blanc C.[1], Francis E.[2], Jamal F.[3], Roy P.[1]
[1]Service de Biostatistique-Bioinformatique, Pôle Santé Publique, Hospices Civils de Lyon, Lyon, France ~ Lyon ~ France,
[2]EverTeam Software ~ Paris ~ France, [3]izyCardio - cardioparc ~ Lyon ~ France

Echocardiography prescription during an appointment could be an indicator of a pathology. Predicting this exam allowed to early detect patients with potential pathology in order to accelerate their care. Different information could be obtained from the patient upstream the consultation, either on textual modality or on structured modality. Using both in multimodal models could allowed to create predictive model for the exam. This work aimed to use these data to predict echocardiography. Different ways to use both modalities in a same model exist following two major paradigms [1]: early fusion if the fusion occurs before the prediction or late fusion if the fusion occurs after the prediction. Literature does not highlight a consensus about the superiority of one way on the other. Data were collected from 23.069 patients. Each patient provided a text with the reason for the consultation and few structured data. All observations were associated with the presence of an echocardiography prescription. Each modality was considered independently and conjointly in models. Structured data were analyzed with feed forward neural networks while Transformers based models [2] were used for text modality. Many multimodal approaches were used and compared, with both early and late fusion paradigms. Six early fusion approaches and two late fusion approaches were tried in order to compare different fusion mechanisms. Each considered approach was evaluated with F1-score computed on a 10-fold bootstrap splitting. Unimodal model using only textual data get performance higher than unimodal model using structured data (mean F1-score of 0.68 and 0.61 respectively). However, using both modalities in multimodal models did not provide better performance than the unimodal model using textual data, whatever the approach considered (with mean F1-score around 0.69 for all multimodal approaches). Using multimodal approaches did not allow to obtain better performance than using unimodal model involving only text data. In exam echocardiography prediction, it seems that using only textual data is enough to achieve the best performance. Nevertheless, as echocardiography is a non-invasive exam, praticians prescribe it even if there is no real emergency for the patient which could be reflected by data.

[1] Gargi Joshi, Rahee Walambe, and Ketan Kotecha. A review on explainability in multimodal deep neural nets. IEEE Access, 9:59800–59821, 2021, doi: 10.1109/ACCESS.2021.3070212.
[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017.

Poster Sessions

WP59 Diverse concepts of machine learning robustness in healthcare: a scoping review

Balendran A.*[1], Beji C.[1], Brion Bouvier F.[1], Evgeniou T.[2], Porcher R.[1]
[1]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS) ~ Paris ~ France, [2]Insead ~ Fontainebleau ~ France

While AI-based solutions in healthcare, and more precisely the AI models used in those solutions, have shown comparable performance—and sometimes better performance in some instances—than medical experts, their apparent performance does not take into account various sources of perturbations that can affect the model behavior, also known as robustness. Prior work formalized the robustness of a ML model as the change in performance when compared to an altered version where one of the components has been perturbed [1]. In recent years, Adversarial attack has become a popular way to assess a model's robustness in healthcare [2]. However, ML models can also be subject to many other different types of perturbations, affecting the performance of a model in different ways, especially in healthcare. Our goal is to provide a comprehensive overview of machine learning robustness in healthcare, focusing on diagnosis and prognosis. We conducted a scoping review by searching through PubMed, IEEE Xplore, and Web of Science. These three databases cover a broad range of machine learning literature in healthcare, making them suitable for our review. We supplemented our database search with studies found on Google Scholar, relevant AI/ML conferences/workshops, and manually. Studies were eligible if they assessed the robustness of supervised ML models for diagnosis or a prognosis in healthcare. Studies that propose new methods to improve the robustness of a model were excluded. Preliminary results indicate that robustness also encompasses concepts such as robustness to data labeling, missing data, data from underrepresented populations, etc. We also propose a mapping of the different studies along dimensions such as nature of the perturbation, data modality, type of model involved, quality score for the evaluation, stage of a model life cycle.

As AI/ML research and applications in healthcare are rapidly advancing, it is important to have a unified framework to serve as a resource for researchers, practitioners, patients, and policymakers to understand the challenges of robustness in AI/ML solutions in healthcare for future research and development.

[1] J. M. Zhang, M. Harman, L. Ma and Y. Liu, Machine Learning Testing: Survey, Landscapes and Horizons, in IEEE Transactions on Software Engineering, vol. 48, no. 1, pp. 1-36, 1 Jan. 2022
[2] Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. Science. 2019 Mar 22;363(6433):1287-1289

WP60 A pragmatic regularization strategy for collinearity-tolerant selection of nonlinear relations

Buch G.*, Gieswinkel A., Wild P.S.
Preventive Cardiology and Preventive Medicine, Department of Cardiology, University Medical Center of the Johannes Gutenberg University Mainz ~ 55131 Mainz ~ Germany

Elastic net regression [1] selects variables while accounting for collinearity among predictors, which increases interpretability. However, the technique is limited to identifying only linear relationships. In biology, there are nonlinear and particularly U-shaped relationships of processes relevant to health development and progression. Therefore, collinearity-tolerant selection of nonlinear relationships is required. As a strategy for collinearity-tolerant selection of nonlinear relationships, the functional form of predictors was modeled with linear splines and the same selected using the Group Least Absolute Shrinkage and Selection Operator (G-LASSO) [2] in combination with an additional l_2-norm applied at the group level. Its performance was compared in a simulation study using classical LASSO [3] and Elastic net with different fractional polynomials. In a real-world application analyzing data from a heart failure study (MyoVasc, ClinicalTrials.gov Identifier: NCT04064450), the approaches were applied and the generated models were compared in terms of their prediction performance using a hold-out data set. Linear spline selection with G-LASSO and an additional l_2-norm regularization showed superiority in variable selection performance, independent of whether only linear, only U-shaped, or a mixture of both relationships had to be identified. In particular, the proposed strategy often performed better than classical strategies, even when the latter had only the correct functional form of predictors available. The additional group-level l_2-norm resulted in moderate improvement over a strategy without it, especially in scenarios with low correlated predictors. The best performance was obtained with high alpha values (such as 0.9), i.e., when the influence of the additional l_2-norm regularization was small. In the real-world application, similar performance of the compared methods in predictive performance was observed, but the generated models were of different sizes and overlapped only moderately. A linear spline modeling strategy combined with a regularized l_2-norm group selection operator is a pragmatic approach that has several appealing properties. It can reduce the feature space to a predictive subset, taking into account the high correlation between predictors, and model the functional form in a flexible way. The generated models are easy to interpret since the functional form is estimated with simple splines, making the technique attractive for practical application.

[1] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.
[2] Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.
[3] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

WP61 Predicting attitudes towards covid-19 vaccination for children: a machine learning-based approach

Chiavenna C.*[1], Leone L.P.[1], Lucchini L.[1], Rotesi T.[2], Omer S.B.[3], Cucciniello M.[1], Pin P.[4], Melegaro A.[1]
[1]Bocconi University ~ Milano ~ Italy, [2]University of Lausanne ~ Lausanne ~ Switzerland, [3]Yale University ~ New Haven, Connecticut ~ United States of America, [4]University of Siena ~ Siena ~ Italy

Quantifying parental hesitancy for COVID-19 vaccination and identifying its determinants are key public health questions. We surveyed 6947 adults, in Italy (N=3633) and the UK (N=3314), between June and July 2021. We assessed willingness to accept COVID-19 vaccination for their kids, if parents, or likelihood to recommend vaccination for their friends' kids, if childless, on a 0-100 scale. Based on their answer, respondents were then classified as hesitant [0,30), undecided [30, 70] or keen (70, 100]. Information on a set of putative predictors was also collected. Predictive power was estimated with two machine learning algorithms, random forest and extreme gradient boosting. Our tuned random forests predicted vaccine hesitancy achieving a macro-F1 score of 0.877 for parents and 0.792 for non-parents, however both models performed better on the polarized classes than on the undecided. For parents, vaccine confidence, vaccination history, trust in government and beliefs about who should be responsible of vaccination decision ranked among the top 5 predictors. Results were similar for non-parents, except for a more sizeable impact of the child's age and of confidence in vaccine providers. Campaigns aiming to reduce vaccine hesitancy will need information on behavioural factors to target relevant population subgroups that would not be identifiable through socio-demographic profiling.

1. MacDonald NE; SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine. 2015;33(34):4161-4164. doi:10.1016/j.vaccine.2015.04.036
2. Ruiz JB, Bell RA. Parental COVID-19 Vaccine Hesitancy in the United States. Public Health Reports. 2022;137(6):1162-1169. doi:10.1177/00333549221114346
3. Viswanath, K., Bekalu, M., Dhawan, D. et al. Individual and social determinants of COVID-19 vaccine uptake. BMC Public Health 21, 818 (2021). https://doi.org/10.1186/s12889-021-10862-1

WP62 Breast lesion malignancy prediction using machine learning and deep learning approaches on ultrasound images

Gaeta A.*[1], Nicosia L.[2], Pesapane F.[2], Bozzini A.C.[2], Vignati S.[1], Bellerba F.[1], Ballerini D.[3], Origgi D.[4], Sangalli C.[5], De Marco P.[4], Castiglione Minischetti G.[4], Cassano E.[2], Gandini S.[1], Raimondi S.[1], Cesarini M.[6]

[1]Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, ~ Milan ~ Italy, [2]Breast Imaging Division, Radiology Department, IEO, European Institute of Oncology IRCCS ~ Milan ~ Italy, [3]Breast Radiology IRCCS Istituto dei Tumori ~ Milan ~ Italy, [4]Medical Physics Unit, IEO, European Institute of Oncology IRCCS ~ Milan ~ Italy, [5]Data Management, European Institute of Oncology IRCCS ~ Milan ~ Italy, [6]University of Milan-Bicocca ~ Milan ~ Italy

One of the most common types of cancer among women is breast cancer. For women aged >50, mammography of the breast is the current recommended screening examination for lesion detection. Nonetheless, ultrasound (US) is widely used in young women and for thick breasts. Malignancy detection is a challenge for US examination, since less than one in ten biopsies required after breast US are cancerous. The aim of this work is to investigate whether Machine Learning and Deep Learning approaches can improve the detection of a malignant lesion over the existing methods. The 420 pre-biopsy DICOM images were acquired using either a Samsung RSV80A or RSV85 Healthcare device. Only images with a lesion were included. Manual segmentations of the lesions were performed and handcrafted image radiomic texture features were extracted using the LIFEx software [1]. Deep Features were extracted using a pre-trained neural network VGG19 on ImageNet [2]. The gold standard for the breast cancer malignancy was the histological result. An AutoML library and the logistic Lasso regression were were used to build the models, that were than tested on the test set after being trained on the training set (proportion 20:80). The models were also evaluated with the addition of patients' age and BI-RADS. S-detect performance were evaluated for a subgroup of patients (54% of the sample). Between the models constructed with the information given by the images, the best one was the model that involved the Deep Feature (Sensitivity (SE); 95%Confidence Intervals (CI)= 73.81; 57.96-86.14; Specificity (SP); 95%CI=71.43; 55.42-84.28). While the model that used jointly deep feature and handcrafted feature with age reached a SE of 85.71 (71.46-94.57) and SP of 61.90 (45.64-76.43). The inclusion of age and BI-RADS ® increased the performance, reaching SE (95%CI)= 90.48 (77.38-97.34) and a SP (95%CI)=90.48 (77.38-97.34). Hand Crafted Features, extracted using mathematical formulas and requiring manual contouring (which consumes specialised labour) performs no better than models with features extracted using the VGG19 pre-trained network. However, the highest predictivity is obtained with the inclusion of the physicians' experience (BI-RADS), despite the radiomic effort.

[1] C. Nioche, F. Orlhac, S. Boughdad, S. ReuzÅLe, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. Frouin, and I. Buvat, "Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity," Cancer research, vol. 78, no. 16, pp. 4786–4789, 2018.
[2] M. Byra, "Discriminant analysis of neural style representations for breast lesion classification in ultrasound," biocybernetics and biomedical engineering, vol. 38, no. 3, pp. 684–690, 2018.

Poster Sessions

WP63 Evaluation of eligibility criteria impact on study outcome and patient count

Jreich R.*[1], Zhang H.[2], Van De Velde H.[3], Meng Z.[4], Wang F.[2]

[1]R&D Data and Data Sciences-Clinical Modeling & Evidence integration, Sanofi, France ~ Chilly-Mazarin ~ France, [2]Department of Population Health Sciences, Weill Cornell Medical College, Cornell University, United States ~ New York ~ United States of America, [3]R&D Oncology dev, Sanofi, Belgium ~ Machelen ~ Belgium, [4]R&D Data and Data Sciences- Clinical Modeling & Evidence integration, Sanofi, United States ~ Boston ~ United States of America

Restrictive and sometimes not fully justified eligibility criteria (EC) are among the factors that could slow down patient enrollment speed. This issue was highlighted by several institutes such as the US National Cancer institute, the American Society of Clinal Oncology (ASCO) (Kim, 2011) (Duggal, 2021). We validated an existing AI algorithm, Trial Pathfinder (Liu, 2021) that allowed to systematically quantify EC's impact on treatment efficacy analysis targeting Relapsed and Refractory Multiple Myeloma (RRMM) domain disease. We have also implemented a bootstrapped version of this algorithm to better understand uncertainties around estimates. For a better assessment of Real-World Data (RWD) quality we considered 2 different RWD data sources Flatiron and OPTUM. The pathfinder was applied independently on 10 historical RRMM trials. The decision rule of criterion removal/relaxation was based on both Shapley values and percentage of patient count excluded by the criterion using RWD. The final judgment was made jointly across all RRMM clinical trialists. The results showed some potential removal/relaxation at level of baseline concomitant conditions such as HIV infection status, cardiac or pulmonary conditions and some laboratory tests such as neutrophil count, creatinine clearance and platelet count. After identification of the subset of inclusion/exclusion criteria that could be relaxed without harming study efficacy it is essential to validate that relaxation of this subset of EC's does not impact study safety outcome.

Kim, Edward S., Thomas S. Uldrick, Caroline Schenkel, Suanna S. Bruinooge, R. Donald Harvey, Allison Magnuson, Alexander Spira et al. "Continuing to broaden eligibility criteria to make clinical trials more representative and inclusive: ASCO–friends of cancer research joint research statement." Clinical Cancer Research 27, no. 9 (2021): 2394-2399.
Duggal, Mili, Leonard Sacks, and Kaveeta P. Vasisht. "Eligibility criteria and clinical trials: An FDA perspective." Contemporary Clinical Trials 109 (2021): 106515.
Liu, Ruishan, Shemra Rizzo, Samuel Whipple, Navdeep Pal, Arturo Lopez Pineda, Michael Lu, Brandon Arnieri et al. "Evaluating eligibility criteria of oncology trials using real-world data and AI." Nature 592, no. 7855 (2021): 629-633.

WP64 Prediction of polypharmacy in half a million adults iranian population: comparing machine learning algorithms

Kamyari N.*[1], Seyedtabib M.[2]

[1]Abadan University of Medical Sciences ~ Abadan ~ Iran, Islamic Republic of, [2]Ahvaz Jundishapur University of Medical Sciences ~ Ahvaz ~ Iran, Islamic Republic of

Polypharmacy (PP) is increasingly common in Iran, and contributes to the substantial burden of drug- related morbidity, increasing the potential for drug interactions and potentially inappropriate medications. Machine learning algorithms (ML) can be employed as an alternative solution for the prediction of PP. Therefore, our study aimed to compare several ML algorithms to predict the PP using the health insurance claims data and choose the best-performing algorithm as a predictive tool for decision making. This population-based cross-sectional study was performed between April 2021 and March 2022. After feature selection, information about 550 thousand patients was obtained from National Center for Health Insurance Research (NCHIR). Afterwards, several ML algorithms were trained to predict PP. Finally, to assess the models' performance, the metrics derived from the confusion matrix were calculated. It was found that ML provides a reasonable level of accuracy in predicting polypharmacy. Therefore, the prediction models based on ML, especially the RF algorithm, performed better than other methods for predicting PP in Iranian people in terms of the performance criteria.

[1] Kamyari N, Soltanian AR, Mahjub H, Moghimbeigi A, Shahali Z. Mapping Drug Prescription, Polypharmacy, and Pharmaceutical Spending in Older Adults in Iran: A Multilevel Analysis Based on Claims Data. Med J Islam Repub Iran. 2021;35(1):1–12.
[2] Kamyari N, Soltanian AR, Mahjub H, Moghimbeigi A, Seyedtabib M. Zero-augmented beta-prime model for multilevel semi-continuous data: a Bayesian inference. BMC Med Res Methodol [Internet]. 2022;22(1):283. Available from: https://doi.org/10.1186/s12874-022-01736-0.

WP65 Signal detection in medical devices in spontaneous reports using natural language processing

Murray C.[1], Stanford T.[2], Kelly T.*[2], Mitchell L.[1], Lim R.[2], Bala I.[1], Ali A.[3], Pratt N.[2], Gillam M.[3]
[1]Department of Mathematics, The University of Adelaide ~ Adelaide ~ Australia, [2]Clinical and Health Sciences, University of South Australia ~ Adelaide ~ Australia, [3]Allied Health and Human Performance, University of South Australia ~ Adelaide ~ Australia

Adverse events from implantable medical devices are commonly reported to regulatory bodies in the form of unstructured free text in spontaneous reports. Detecting safety signals from the reports for post-market surveillance can be challenging. The objectives of this study were to (1) profile the unstructured text into mixtures of clinically relevant topics and (2) use the topics to detect potential safety signals from use of urogynaecological mesh. The Database of Adverse Event Notifications (DAEN) maintained by the Australian Therapeutic Goods Administration (TGA) was searched for reports on urogynaecological and hernia meshfrom 2012- 2017. Topic modelling, a Natural Language Processing technique, was used to cluster common groupings of words into topics. Reports containing the most frequent clinical topic were considered events in retrospective disproportionality analysis of urogynaecological mesh, with hernia mesh as a comparator. The maximised sequential probability ratio test (maxSPRT) [1] accounted for multiple testing through alpha spending, while the Bayesian Confidence Propagation Neural Network (BCPNN) was used with and without adjusting for multiple testing. Testing was performed at monthly intervals, if new data were accumulated in the interval, over the study period, commencing in 2012. Results: Words associated with 'pain' comprised the most frequent clinical topic. A signal was detected from maxSPRT in December 2014 after accumulating 29 events in urogynaecological mesh and 6 in hernia mesh. BCPNN without multiple testing adjustment detected safety signals in August 2014 and with exponential spending multiple adjustment, in December 2014. Urogynaecological mesh was withdrawn from the Australian market in 2018, while our retrospective analysis detected signals between August – December 2014. We have demonstrated the potential of using topic modelling in spontaneous reports for signal detection in post-market surveillance.

Funding: This project was funded by the Australian National Health and Medical Research Council grant number GNT2002589.

[1] Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, and Platt R (2011). A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance, Sequential Analysis 30: 58–78.

WP66 Deep learning-based prediction of major arrhythmic events in dilated cardiomyopathy

Corrado L.*[1], Mattia C.[2], Laura D.M.[2], Martina P.M.[2], Sabino I.[2], Dario G.[1], Francesco T.[2]
[1]Unit of Biostatistics, Epidemiology, and Public Health - Dep. Cardiac, Thoracic, Vascular Sciences, and Public Health - University of Padova ~ Padova ~ Italy, [2]Dep. Cardiac, Thoracic, Vascular Sciences, and Public Health - University of Padova ~ Padova ~ Italy

Major arrhythmic events (MAEs)include sudden cardiac death, cardiac arrest due to ventricular fibrillation, sustained ventricular tachycardia lasting ≥30s or causing hemodynamic collapse in <30s, and appropriate implantable cardiac defibrillator intervention.Predicting MAEs in dilated cardiomyopathy (DCM) remains a significant clinical challenge.[1] Emerging technologies, such as computational models and artificial intelligence (AI), show promise in enhancing our ability to predict MAEs in DCM. Although, AI approach has not been tested in this field.[2] In this proof-of-concept study, we introduce a deep learning (DL)-based model, DARP-D, which utilizes multidimensional cardiac magnetic resonance data (cine videos, hypervideos, LGE images, and hyperimages) and clinical covariates to predict and monitor individual patient risk curves for MAEs over time. We trained and validated the DARP-D model on 70% of a sample of 154 DCM patients, and tested it on the remaining 30%. The model achieved a 95% confidence interval (CI) for Harrell's C concordance indexes ranging from 0.12 to 0.68 on the test set. Our study demonstrates the feasibility and novelty of the DL approach for arrhythmic risk prediction in DCM. The DARP-D model effectively analyzes cardiac motion, tissue characteristics, and baseline covariates to predict individual patient risk curves for major arrhythmic events.

[1] Iliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing Risk Stratification for Sudden Death in Dilated Cardiomyopathy. The Past, Present, and Future. Circulation (2017) 136:215-231. doi: 10.1161/CIRCULATIONAHA.116.027134
[2] Corianò M, Tona F. Strategies for Sudden Cardiac Death Prevention. Biomedicines (2022) 10(3):639. doi: 10.3390/biomedicines10030639.

Poster Sessions

WP67 Demag predicts the effects of variants in actionable genes with structural and evolutionary features

Luppino F.*[1], Adzhubei I.A.[2], Cassa C.A.[3], Toth--Petroczy A.[1]
[1]Max Planck Institute of Molecular Cell Biology and Genetics ~ Dresden ~ Germany, [2]Department of Biomedical Informatics, Harvard Medical School ~ Boston ~ United States of America, [3]Brigham and Women's Hospital Division of Genetics, Harvard Medical School ~ Boston ~ United States of America

Despite the increasing use of genomic sequencing in clinical practice, the interpretation of rare genetic variants remains challenging even in well-studied disease genes, resulting in many patients with Variants of Uncertain Significance (VUSs). Computational Variant Effect Predictors (VEPs) provide valuable evidence in variant assessment, but they are prone to misclassifying benign variants, contributing to false positives. The 'partners score' provides a general framework for modeling epistatic interactions, integrating both clinical and functional information. Here, we develop Deciphering Mutations in Actionable Genes (DeMAG), a supervised classifier for missense variants trained using extensive diagnostic data available in 59 actionable disease genes (American College of Medical Genetics and Genomics Secondary Findings v2.0, ACMG SF v2.0). DeMAG improves performance over existing VEPs by reaching balanced specificity (82%) and sensitivity (94%) on clinical data, and includes a novel epistatic feature, the 'partners score', which leverages evolutionary and structural partnerships of residues. We provide our tool and predictions for all missense variants in 316 clinically actionable disease genes (demag.org) to facilitate the interpretation of variants and improve clinical decision-making. https://doi.org/10.1101/2022.06.15.496230

WP68 Detection of recurrent cancer from emr using natural language processing: a systematic review

Ponthongmak W.*[1], Sangariyavanich E.[1], Tansawet A.[1], Theera--Ampornpunt N.[1], Mckay G.[2], Attia J.[3], Numthavaj P.[1], Thakkinstian A.[1]
[1]Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand ~ Bangkok ~ Thailand, [2]Centre for Public Health, Queen's University Belfast ~ Belfast ~ United Kingdom, [3]Centre for Clinical Epidemiology and Biostatistics, School of Medicine and Public Health, University of Newcastle ~ Newcastle ~ Australia

Background: Cancer recurrence is a public health concern, and its documentation is mainly recorded in clinical narratives. Extracting recurrence status from electronic medical records (EMRs) is a time-consuming and labor-intensive task that is prone to errors. Natural language processing (NLP) tecniques has shown potential in identifying recurrent cancer diagnoses from EMRs. Therefore, this study aims to systematically review the NLP techniques and algorithms used for this purpose and describe their performance.
Objective: The objective of this study is to systematically review the NLP techniques and algorithms in identifying recurrent cancer diagnoses from EMRs.
Method: PubMed, Scopus, ACM Digital Library, and IEEE databases were searched to identify studies that used NLP techniques and algorithms to identify recurrent cancer diagnoses from EMRs. Two independent reviewers extracted data from the included studies, such as text representation, algorithm, type of clinical notes, and model performance. The risk of bias (ROB) was assessed using the PROBAST tool.
Results: From 412 studies, 17 studies were included, of which 15 developed models without externally validated model. 3 word representations, including statistical, context-free, and contextual word representations, were applied in 12, 6, and 3 studies, with corresponding median F1 scores of 0.43, 0.87, and 0.72, respectively. The algorithms applied consisted of rule-based, machine learning, and deep learning approaches with median F1 scores of 0.71, 0.43, and 0.76, respectively.
This systematic review found that deep learning with word representation using PubMedBERT, which trained on medical domains, have better performance in identifying recurrent cancer diagnoses from EMRs. However, the limited input text length of 512 tokens is a challenge, as it is computationally expensive for the training process with large data samples. Therefore, other competitive large language models should be considered to address these weaknesses. This study highlights the potential of NLP techniques and algorithms in identifying recurrent cancer from EMRs and emphasizes the need for further research in this area.

WP69 Outlier detection for tree-structured model in regression problem

Seiwa N.*, Shimokawa A.
Tokyo University of Science ~ Tokyo ~ Japan

In many research fields including the medical research, outlier detection is an important subject. Outliers are generally considered as the points that have different characteristics from normal points, and there are two different ways to detect these points. The first one is considering only the values of points. This method is treated as an area of unsupervised learning in the field of machine learning. The methods using ensemble learning have been particularly studied in recent years, and isolation forest is a representative method (Liu and Zhou, 2008). The second approach is detecting outliers that do not fit the model. Although the residual-based methods for outlier detection are generally used, the use of those methods are limited to very simple models such as linear regression model. In this research, we propose a new method to detect outliers that affect the prediction accuracy of the tree- structured model. In the isolation forest, the number of partitions required to isolate a data point is measured by the path length from the root node to the leaf node in each tree. If the number is small, the data point is considered easy to isolate and is detected as outlier. In normal decision tree construction, in generally, outliers that have a large impact on response prediction also tend to be separated in shallow layers of the tree. Therefore, we propose the method to evaluate outliers using the path length until the data points are isolated. From simulation studies, it was shown that the proposed method has superior performance to detect outliers in terms of prediction accuracy. We proposed a new method to detect outliers for tree-structured model. As a further development, we think that the method can be extended to detect outliers not only in continuous responses but also in discrete and survival time responses. In addition to the simulation results, we will show the results of the proposed method on actual data in the conference.

F. T. Liu and Z.-H. Zhou. Isolation forest, In Proceedings of the 8th IEEE International Conference on Data Mining, 413-422, 2008.

WP70 Weakly-supervised classification of clinical documents: a case study on italian discharge letters

Torri V.*[1], Barbieri E.[2], Cantarutti A.[3], Giaquinto C.[2], Ieva F.[1]
[1]MOX - Modelling and Scientific Computing Lab, Department of Mathematics, Politecnico di Milano ~ Milan ~ Italy, [2]Division of Pediatric Infectious Diseases, Department for Woman and Child Health, University of Padua ~ Padua ~ Italy, [3]Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca ~ Milan ~ Italy

Discharge letters that are produced during hospitalizations contain valuable information on patients' conditions that is not stored in a structured format but only as free text. The disease for which a patient is admitted to/discharged from the hospital is the most relevant information enclosed in a discharge letter, and a model able to classify discharge letters with respect to certain diseases can be particularly useful for both cohort selection and epidemiological analysis. This is a typical Natural Language Processing (NLP) problem, known as document classification. Training NLP models requires an annotated dataset, but unfortunately it is not currently available in Italian. The aim of this work is to show how it is possible to develop such a model for the Italian context by applying a weakly- supervised approach. A novel NLP-based pipeline for Italian sentences, based on a fine-tuned version of the transformer- based model BERT[1], is used to cluster diagnoses contained in discharge letters as text strings and not coded according to standard ICD9-CM definitions. The last Italian version of BERT, Umberto, has been fine-tuned with Italian texts from the medical domain and different dimensionality reduction techniques have been investigated to feed its embeddings to a clustering algorithm. A second-level clustering on a keywords-based representation is applied with the aim of merging groups of diagnoses referred to the same disease. The final resulting clusters are subsequently used as disease labels by the BERT-based model, empowered with an additional layer, to associate each discharge letter to a given pathology. Results on the identification of viral infections on a dataset of 7000 Italian discharge letters from various hospitals in the Veneto Region show improvements with respect to state-of-the-art techniques. Evaluation is performed over a test set whose labels have been manually validated. We tackled for the first time the problem of disease classification on Italian discharge letters, achieving good performances even in absence of an annotated dataset, proving that this is not necessarily a limiting factor for the development of NLP models in scarce-resource domains and languages.

[1] J. Devlin, M. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pages 4171–4186

Poster Sessions

WP71 Machine learning regression models for predicting hospital stay for general medicine-specific patients

Addisu Jember Z.[1], Pierpaolo P.[1], Paolo T.[2], Rossella M.*[3], Lorenzo C.[1]
[1]Department of Electrical, Electronic, and Information Engineering Guglielmo Marconi, University of Bologna, 40126 Bologna ~ Bologna ~ Italy, [2]Enterprise Information Systems for Integrated Care and Research Data Management, IRCCS Azienda Ospedaliero- Universitaria di Bologna, 40138 Bologna ~ Bologna ~ Italy, [3]Department of Statistical Sciences, University of Bologna, 40126 Bologna ~ Bologna ~ Italy

The General medicine (GM) department has the highest patient volume and heterogeneities in the hospital. Examining hospitalization data closely is crucial because patients come with a range of conditions or traits. Length of stay (LoS) is influenced by a patient's medical background, demographic factors, arrival mode at the hospital, triage evaluation, and type of illness or problem [1]. LoS is a variable that can vary widely, making it difficult to be estimated accurately, but doing so is highly beneficial also for the evaluation of hospital resource organization. The objective was to estimate and compare machine learning (ML) regression models that predict the actual number of LoS days as opposed to a usual classification method, by utilizing demographic and clinical information derived from the observable characteristics at admission. We included patients in the GM department who were admitted through the ED at Sant'Orsola Malpighi University Hospital in Bologna, Italy. The data were collected from January 1st, 2022 to October 26th, 2022. Nine ML regression models were used to predict LoS by analyzing historical data and patient information. The model's performance was assessed based on root mean squared prediction error (RMSPE), and mean absolute prediction error (MAPE). Moreover, we used K-means clustering to divide patients' criticalities into four clusters prior to ML analysis. Our study of 3757 eligible patients with an average LoS of 13 days, with a standard deviation of 11.8, we used a training cohort of 2630 patients (70%) and a test cohort of 1127 patients (30%). The XGBR model had the lowest predicted error for both RMSPE (11.00 days) and MAE (7.52 days) and identified sex, arrival mode (by own means/walk-in or by ambulance), triage category (light blue, orange, green, or red), age group (70 and older, 50 to 69, or 30 to 49), and reported specific problems such as hematochezia/rectorrage/melena, pain at the side, pre-syncope, pallor/anemia, generalized asthenia, and request for prescription/performance. The ML models developed in this study reported good predictive performance, with the XGBR model exhibiting the lowest prediction error. This model can help physicians administer appropriate clinical interventions for GM patients.

1. Shea, S., Sideli, R.V., Dumouchel, W., Pulver, G., Arons, R.R., Clayton, P.D.: Computer-generated informational messages directed to physicians: effect on length of hospital stay. J. Am. Med. Inform. Assoc. 2(1), 58–64 (1995).

WP72 Missforest v2 – missing data imputation for prediction settings

Albu E.*, Gao S., Van Calster B., Wynants L.
Department of development and Regeneration, KU Leuven , Belgium ~ Leuven ~ Belgium

When applying clinical prediction models in practice, data such as electronic health records may have missing data at prediction time. Few practical solutions exist to handle missing values in real time. We adapted the popular missForest imputation algorithm for prediction settings, and compared it to other imputation methods on clinical datasets. missForest is an iterative imputation algorithm based on random forests (RF). We adapted the missForest R package for faster computation time and for saving imputation models applicable at prediction time (missForest v2). The convergence criterion has been unified for continuous and categorical variables. missForest v2 is compared to mean/mode imputation, k-nearest neighbours, bagging and two iterative RF algorithms (miceRanger and IterativeImputer) on three clinical datasets with missing values and binary outcomes (n: 15000 to 20000, missingness rate for the variable with most missing values: 20% to 49%). For each imputation method, following steps are repeated 100 times: train/test split; imputation of train set and learning of imputation models; development of prediction models on train set using RF and logistic regression with cubic splines; imputation of test set using the "learned" imputation models; making predictions on the imputed test set; performance evaluation. Preliminary results show that the imputation methods make little difference in overall predictive performance, with mean/mode imputation being comparable with other imputation methods. Median AUROCs for worst vs. best imputation method are 0.628 – 0.634 (dataset1), 0.898 – 0.90 (dataset2), and 0.801 – 0.804 (dataset3) for RF, and 0.624 – 0.637, 0.889 – 0.890 and 0.798 – 0.808 for logistic regression. Median calibration slope for worst vs. best imputation method are: 0.962 – 0.989, 1.054 – 1.032, 1.068 – 1.044 for RF and 0.832 – 0.925, 0.989 – 0.994 and 0.966 – 0.971 for logistic regression. The preliminary results show no clear advantage of using sophisticated imputation methods. The results in applied prediction settings can though vary in function of the amount of missingness, how well variables can be predicted based on the other variables and the importance of variables in the prediction model. missForest v2 provides similar results compared to other imputation algorithms but comes with additional usability features.

[1] Stekhoven, D. J., & Bühlmann, P. (2012). MissForest – non-parametric missing value imputation for mixed-type data. Bioinformatics, 28(1), 112-118.

WP73 Evaluating bias in causal mediation effects with non-adherence and missing data: better simulate than never

Chis Ster A.*, Landau S., Emsley R.
King's College London ~ London ~ United Kingdom

Many clinical trials suffer from participant non-adherence. A standard intention-to-treat (ITT) analysis will estimate the causal effect of treatment offer without bias, though ignores the impact of non- adherence. To account for non-adherence in an analysis, one can compute a complier-average causal effect (CACE) analysis to provide an unbiased estimate for the average causal effect of treatment receipt in the subgroup of participants who comply with their randomisation. Clinical trials in mental health often evaluate complex interventions such as psychotherapy. Evaluating how complex interventions lead to changes in the outcome (the mechanism) is key for the development of more effective interventions. A mediation analysis aims to decompose a total treatment effect into a mediated effect, one that operates via changing the mediator, and a direct effect. However, current methods for mediation analysis in clinical trials decompose the ITT effect, and the corresponding effects ignore the impact of participant non-adherence. Previous work has decomposed the CACE into a direct effect, the Complier Average Natural Direct Effect (CANDE), and a mediated effect, the Complier Average Causal Mediated Effect (CACME) using Structural Equation Models (SEMs). An accompanying Stata command has been developed for practical implementation of this method. However, the reliability and interpretability of the estimates are affected by missing data. The aim of this work is to evaluate the performance of linear SEMs for estimating the CACE, CACME, and CANDE when there are missing data. A Monte Carlo simulation study is conducted to evaluate the bias in CACE, CACME, and CANDE when there are missing data in the mediator and/or outcome. We construct three scenarios where the missing data are MCAR, three MAR scenarios, and five MNAR scenarios, to cover a range of realistic scenarios. We vary 8 parameters, including the trial size, the proportion of non-adherence, and the proportion of missing data. Our findings show that linear SEMs provide unbiased estimates of CACE, CACME, and CANDE for all MNAR and MAR scenarios, but not for MNAR scenarios. Trialists should evaluate the missing mechanisms in their study before adopting linear SEMs to estimate the CACE, CANDE, and CACME. Park Soojin & Kürüm Esra, 2020. "A Two-Stage Joint Modeling Method for Causal Mediation Analysis in the Presence of Treatment Noncompliance," Journal of Causal Inference, De Gruyter, vol. 8(1), pages 131-149, January.

WP74 Comparison of imputation methods in the presence of time-varying and non- linear covariates effects

Imad E.B.*[1], Roch G.[2]
[1]Mohamed VI Center for Research & Innovation, Rabat, Morocco. Mohammed VI University of Health Sciences (UM6SS) ~ Casablanca ~ Morocco, [2]Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, ISSPAM, Hop Timone, BioSTIC, Biostatistique et Technologies de l'Information et de la Communication ~ Marseille ~ France

Incomplete datasets are common in clinical and epidemiological applications. Imputation of missing data is often a necessary step that precedes the modeling task. Despite the fact that multiple imputation remains the choice, one cannot use multiple imputation methods before using a machine learning (ML) algorithm, since the latter only takes a single data set as input. Furthermore, in Cox regression analysis, it is necessary to study whether the associations of covariates with the hazard vary over time or have a non-linear effect, for continuous one. Indeed, ignoring this assumption can conduct to misleading conclusions [1]. Furthermore, multiple imputation uses a deterministic imputation model that may differ in real-world applications from the substantive model, thus leading to asymptotically biased parameter estimates [2]. The objective of our study is to investigate whether ML- based imputation methods perform better than classical ones in the presence of time-varying and non- linear covariate effects. Extensive simulations were carried out to compare the performance of some imputation methods: MICE, MICE-PMM, MICE-CART, MICE-RF, MISL, Hotdeck, Single-PMM, Mean/Mode, CART, FAMD, missForest, missRanger, missCforest. Missing values were artificially introduced under "MAR", "MNAR" and "MCAR" mechanisms. Sample size, proportion of missing data and censoring rate have been made varied in order to investigate different situations. Performance was assessed by means of bias, coverage rate, RMSE, normalized imputation error, AUC, IBS and the estimated curve of non- linear and time-vayring effects presented visually over the follow-up. Both MICE and some ML methods produced unbiased estimates in proportional hazard setting. But ML-based methods were more efficient in terms of imputation error and with less biased estimates in the presence of time- vaying and non-linear effects. Also, among the methods compared, parameter estimates were less biased for tree ensemble methods. In the presence of time-varying and non-linear covariates effects, ML-based methods are more suitable for imputation than classical methods. In fact, ML-based methods have an important potential in capturing complex associations between covariates and the survival times, especially in an unknown context about the underlying structure of the data in hand.

[1] Keogh RH, Morris TP. Multiple imputation in Cox regression when there are time-varying effects of covariates. Statistics in medicine 2018;37(25):3661-78.
[2] Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. American journal of epidemiology 2014;179(6):764-74.

WP75 Sensitivity bounds in multilevel modeling of longitudinal data with data missing not at random

Genbäck M.*, Josefsson M.
Umeå University ~ Umeå ~ Sweden

Longitudinal cognitive aging studies almost always suffer from missing data because some participants are lost to follow-up. If the dropout mechanism is related to unmeasured cognitive performance levels the missingness mechanism is thought to be missing not at random (MNAR). In these situations, standard approaches, such as multilevel or mixed effects models, are not valid. In the current study, we develop a sensitivity analysis approach when missing outcome data is thought to be MNAR. In particular, the proposed method estimate uncertainty bounds for regression parameters in a multilevel model. Here, the association between the dropout mechanism and the outcome is modeled through a general sensitivity parameter matrix which determines the strength (and direction) of the departure from a missing at random assumption. We extend current approaches [1] for sensitivity analyses by estimating bounds for different missingness assumptions and sensitivity parameters. We will apply the proposed method to data from a large survey on health, retirement and aging, where interest is to study the effect of a lifestyle factor on cognitive decline in older participants. The approach allows for flexible assumptions on the missing data, where the strength of the assumptions are directly related to the width of the sensitivity bounds. Genbäck, M., Stanghellini, E., & de Luna, X. (2015). Uncertainty intervals for regression parameters with non-ignorable missingness in the outcome. Statistical Papers, 56, 829-847.
Grilli, L., & Rampichini, C. (2010). Selection bias in linear mixed models. Metron, 68, 309-329.

WP76 Practical approach for missing data sensitivity analyses in joint modelling of cognition and dementia risk

Gorbach T.*[1], Carpenter J.[2], Frost C.[2], Josefsson M.[1], Macdougall A.[2], Nicholas J.[2], Nyberg L.[1]
[1]Umeå University ~ Umeå ~ Sweden, [2]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom

Joint modelling [1] of longitudinal cognitive measures and time-to-dementia onset is a natural tool for understanding the relationship between the trajectory of cognitive decline and dementia. However, both the longitudinal cognitive measure model and the time-to-dementia model typically assume data are missing (censored) at random. Unfortunately, in longitudinal studies of ageing, it appears likely that dropout from the longitudinal measurement schedule is non-ignorable. We propose a practical imputation-based approach for exploring the sensitivity of inferences to such non-ignorable dropout and apply it to the analysis of cognition and risk of dementia using data from the Betula longitudinal study [2]. For sensitivity analysis we: (a) impute missing longitudinal cognitive measurements using a pattern- mixture approach applied to the mixed effect submodel while accounting for possibly accelerated cognitive decline after dropout, using contextual knowledge to inform the choice of the sensitivity parameter values; (b) fit the joint to each imputed data set and (c) combine the results for inference using Rubin's rules. Application to the Betula data shows key joint model inferences are robust to contextually plausible non-ignorable missing longitudinal cognitive measures. We have developed a practical, yet general, approach for sensitivity analyses for missing not at random data in joint modelling of longitudinal data and time-to-event data, and applied it to the analysis of data from the Betula longitudinal study of ageing and dementia.

[1] Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R. Boca Raton, FL, USA: CRC press. 2012.
[2] Nilsson LG, Bäckman L, Erngrund K, et al. The Betula Prospective Cohort Study: Memory, Health, and Aging. Aging, Neuropsychology, and Cognition 1997; 4(1): 1–32.

Poster Sessions

WP77 Bayesian latent class analysis correcting for verification and reference standard bias in tb prevalence

Keter A.K.*[1], Vanobberghen F.[2], Lynen L.[1], Van Heerden A.[3], Fehr J.[4], Olivier S.[4], Wong E.B.[4], Glass T.R.[2], Reither K.[2], Jacobs B.K.M.[1], Goetghebeur E.[5]
[1]Institute of Tropical Medicine ~ Antwerp ~ Belgium, [2]University of Basel ~ Basel ~ Switzerland, [3]Human Sciences Research Council ~ Pietermaritzburg ~ South Africa, [4]Africa Health Research Institute ~ Durban ~ South Africa, [5]Ghent University ~ Gent ~ Belgium

Estimation of pulmonary tuberculosis (PTB) prevalence suffers from two bias sources: imperfect reference standard testing on a selected sub-cohort. In our study, a random subset of participants screening positive with at least one TB symptom or abnormal chest X-ray is eligible for TB verification using Xpert MTB/RIF Ultra (Ultra) and culture. Screening has low sensitivity and a composite reference standard of Ultra and culture is also imperfect. Furthermore, a subset of eligible subjects has their TB status eventually verified. When the verified subjects are not comparable to the unverified, estimation suffers 'verification bias'. We estimate PTB prevalence correcting for the reference standard and verification biases in a community-based multimorbidity screening study (CBMSS) in South Africa. We conducted a secondary analysis of 9914 subjects aged ≥15 years from the CBMSS. Verification was planned if the participants were eligible. Besides 3500 ineligible participants, 22% of the 6369 eligible were not verified. We compared naïve Bayesian analysis to Bayesian latent class analysis (BLCA) for i) complete case analysis, ii) assuming the unverified were negative for Ultra and culture, iii) analysis of multiply-imputed datasets. Using BLCA we simultaneously imputed the missing Ultra and culture results in the analysis model under iv) the missing at random (MAR) v) missing not at random (MNAR) assumptions. We relaxed model assumptions to allow conditional dependence between certain sets of tests based on similar biological mechanisms. Unknown model parameters were assigned Gaussian priors. Through simulation (overall true prevalence=2.0%) we evaluated the ability of (iv) and (v) to simultaneously alleviate the biases. Naïve Bayesian analysis produced, complete case analysis: 1.8% (95% CrI:1.5,2.2), assuming unverified were negative for ultra and culture: 0.9% (95% CrI:0.8,1.1), multiply-imputed datasets:1.5% (95% CrI:1.2,1.9). Respectively, BLCA produced 1.4% (95% CrI:0.9,1.4),0.8% (95% CrI:0.6,1.4) and 1.0% (95% CrI:0.7,1.5). Analysis with simultaneous imputation produced, MAR:1.0% (95% CrI:0.7,1.4), MNAR: 1.0% (95% CrI:0.7,1.5). The simulation yielded, MAR: 2.0% (95% CrI:1.4,3.9), MNAR: 2.2% (95% CrI:1.7,2.8). Ignoring uncertainty in imperfect diagnostic tests and information in the selected sample of verified individuals may yield bias that can be alleviated pragmatically. Imputing missing values as negative for Ultra and culture may not be problematic.

[1] Kendall et al. (2021) *[2] Lau et al. (2022)* *[3] Frascella et al. (2021)* *[4] Pillay et al. (2020)*
[5] van Buuren and Groothuis-Oudshoorn (2011) *[6] Keter et al. (2023)*

WP78 The handling of missing data with multiple imputation in observational studies that address causal questions

Mainzer R.M.*[1], Moreno--Betancur M.[1], Nguyen C.D.[1], Simpson J.A.[2], Carlin J.B.[1], Lee K.J.[1]
[1]Murdoch Children's Research Institute ~ Parkville ~ Australia, [2]University of Melbourne ~ Parkville ~ Australia

Observational studies in health research often aim to answer causal questions. Missing data are common in such studies and can occur in the exposure, outcome and/or variables used to control for confounding. The standard classification of data as "missing completely at random", "missing at random" (MAR) or "missing not at random", does not allow for a clear assessment of missingness assumptions when missing values arise in more than one variable. This presents challenges for selecting an analytic approach and determining when a sensitivity analysis under plausible alternative missing data assumptions is required. Our objective was to review how MI is currently used in observational studies that address causal questions, with a focus on if and how (i) missingness assumptions are stated and assessed, (ii) missingness assumptions are used to justify the choice of a complete case analysis or MI for handling missing data, and (iii) sensitivity analyses under alternative plausible assumptions about the missingness mechanism are conducted. We are conducting a scoping review of observational studies, published between January 2019 and December 2021 in five top-ranked epidemiology journals, that aimed to answer causal questions and used MI.[1] We identified 343 studies using a full text search for the term "multiple imputation" on the journal websites; 221 full-texts were assessed for eligibility and 136 studies have been included in the review, which is anticipated to be completed by June 2023. We are extracting information on study characteristics, amount of missing data, missingness assumptions, analysis methods and MI implementation. Our preliminary data reveal that, although missing data are commonly observed among exposure, outcome and confounder variables, missing data assumptions are poorly reported, the selection of missing data approaches are not well-justified, and sensitivity analyses to missing data assumptions are seldom conducted. Further effort is needed to ensure the handling and reporting of missing data in observational studies are appropriate.

[1] Mainzer, R., Moreno-Betancur, M., Nguyen, C., Simpson, J., Carlin, J. and Lee, K., 2023. Handling of missing data with multiple imputation in observational studies that address causal questions: protocol for a scoping review. BMJ Open, 13(2), p.e065576.

WP79 Min-max-median/iqr approach. Comparison with min-max, logistic regression and xgboost.

Aznar--Gimeno R.*[1], Esteban L.M.[2], Sanz G.[3], Del--Hoyo--Alonso R.[1]
[1]Instituto Tecnológico de Aragón ~ Zaragoza ~ Spain, [2]Escuela Universitaria Politécnica de la Almunia, Universidad de Zaragoza ~ La Almunia de Doña Godina, Zaragoza ~ Spain, [3]Universidad de Zaragoza ~ Zaragoza ~ Spain

While combining multiple variables linearly can yield satisfactory diagnostic performance, some algorithms have a drawback of becoming computationally intensive when the number of variables is large. To address this issue, Liu et al. introduced a distribution-free method called the min-max approach. We have developed the Min-Max-Median/IQR algorithm using Youden index optimization [1], which is more computationally intensive but still manageable and provides additional information. On the other hand, machine learning algorithms have been increasingly used in various fields of application, such as the XGBoost algorithm. However, the performance of an algorithm depends on the data and the problem, and it is always necessary to carry out an exhaustive comparative study to analyze the performance of the algorithms and establish certain guidelines. The aim of our study was to compare the performance of our proposed approach with the min-max approach, logistic regression, and XGBoost. Our proposed approach is an extension of the min-max algorithm using a stepwise approach that we have developed [2] following the suggestions of Pepe et al. The idea behind is to reduce the dimension of the problem, considering the min, max, and median/IQR information of the original variables, turning the problem into a three-variable linear combination optimisation problem. We compared our approach with the min-max approach, logistic regression, and XGBoost on a wide range of simulated scenarios and on two real data datasets (diagnosis of Duchenne muscular dystrophy and Maternal mortality risk). Specifically, scenarios simulating different biomarker distributions, discrimination capabilities, and correlation between them were analysed, considering different number of biomarkers and sample sizes. The results showed that machine learning approaches outperformed our approach, particularly in scenarios with biomarkers with different predictive abilities and in biomarker scenarios with different marginal distributions. However, our approach outperformed the rest in scenarios with biomarkers with the same predictive ability and different correlations between groups. Our study presents a comprehensive comparison of various approaches, presenting our proposed approach (Min-Max-Median/IQR) as an alternative to machine learning models such as logistic regression and XGBoost, in certain scenarios where it has demonstrated superior performance, i.e. in scenarios with same predictive ability and different correlations between groups.

[1] Aznar-Gimeno, R., Esteban, L. M., Sanz, G., del-Hoyo-Alonso, R., & Savirón-Cornudella, R. (2021). Incorporating a New Summary Statistic into the Min–Max Approach: A Min–Max–Median, Min– Max–IQR Combination of Biomarkers for Maximising the Youden Index. Mathematics, 9(19), 2497.
[2] Aznar-Gimeno, R., Esteban, L. M., del-Hoyo-Alonso, R., Borque-Fernando, Á., & Sanz, G. (2022). A Stepwise Algorithm for Linearly Combining Biomarkers under Youden Index Maximization. hematics, 10(8), 1221.

WP80 Mortality prediction: one size may not fit all. the italian experience on the validity of the pim 3 score

Comoretto R.*[1], Chiaruttini M.V.[2], Amigoni A.[3], Wolfler A.[4], Moscatelli A.[4], Izzo F.[5], Dusio M.P.[6], Caramelli F.[7], Stancanelli G.[8], Ricci Z.[9], Chidini G.[10], Gitto E.[11], Zito Marinosci G.[12], Gallina R.[13], Picconi E.[14], Rossetti E.[15], Vitale P.[16], Sagredini R.[17], Biban P.[18], Gregori D.[2]
[1]Department of Public Health and Pediatrics, University of Torino ~ Torino ~ Italy, [2]Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padova ~ Padova ~ Italy, [3]Pediatric Intensive Care Unit, Department of Women's and Children's Health, University Hospital of Padova ~ Padova ~ Italy, [4]Neonatal and Pediatric Intensive Care Unit, IRCCS G. Gaslini ~ Genova ~ Italy, [5]Department of Pediatric Anesthesia and Intensive Care, Children's Hospital "Vittore Buzzi" ~ Milano ~ Italy, [6]Unit of Pediatric Intensive Care, Children's Ospedale C. Arrigo ~ Alessandria ~ Italy, [7]Department of Pediatric Anesthesia and Intensive Care Unit, University Hospital Policlinico S.Orsola-IRCCS ~ Bologna ~ Italy, [8]Pediatric Intensive Care Unit, Garibaldi Nomisma Hospital ~ Catania ~ Italy, [9]Pediatric Intensive Care Unit, Meyer Children's University Hospital ~ Firenze ~ Italy, [10]Pediatric Intensive Care Unit, IRCCS De Marchi ~ Milano ~ Italy, [11]Pediatric Intensive Care Unit, University Hospital G. Martino ~ Messina ~ Italy, [12]Unit of Pediatric Intensive care, Ospedale Santobono ~ Napoli ~ Taiwan, [13]Pediatric and Neonatal Intensive Care Unit, Maggiore Della Carità University Hospital ~ Novara ~ Italy, [14]Unit of Pediatric Intensive Care, A. Gemelli Hospital, Catholic University ~ Roma ~ Italy, [15]Department of Paediatric Cardiac Anesthesia and Intensive Care, Children's Hospital Bambino Gesu' ~ Roma ~ Italy, [16]Pediatric Intensive Care Unit, Pediatric Hospital Regina Margherita ~ Torino ~ Italy, [17]Unit of Pediatric Intensive Care, Burlo Garofolo Hospital ~ Trieste ~ Italy, [18]Department of Neonatal and Paediatric Intensive Care, University Hospital ~ Verona ~ Italy

In paediatric intensive care units (PICUs), the prediction of death probability at admission is a milestone in the management of the patient. The paediatric index of mortality (PIM3), developed in 2013 by Straney on Australian and United Kingdom PICUs registries, over the years has become the reference model. However, it is necessary to validate the score in the Italian population. Methods. Data have been used from the largest Italian Network of Paediatric Intensive Care Units (TIPNet) registry, which counts more than 30 Italian PICUs,. The receiving operating characteristics (ROC) of the scoring test and the corresponding Area Under Curve (AUC) with 95%CI have been calculated and the assessment of the model calibration has been performed using the Brier Score and the Hosmer-Lemeshow (HL) test. Then, to calibrate the model, four methods (Platt scaling method, isotonic regression, polynomial regression, and beta calibration) were adopted, and their performance were compared. Results. Among 6,451 patients for which the PIM3 score was calculated, 283 (4.4%) died during PICU hospitalization and 6,170 (95.6%) were discharged alive. Overall, the estimated PIM3 score shows a higher frequency for extreme probabilities close to 0. The ROC curve has an AUC equal to 88.5 (95%CI: 86.4-90.5) and the best threshold used to maximise sensitivity (81.3) and specificity (80.0) is 0.051. The general HL test is 26.77, with a p-value < 0.001, suggesting a poor calibration of the current model. Among the four calibration methods, the best calibration ability has been demonstrated by beta calibration (HL test = 6.13, p = 0.11). Overall, the PIM 3 score has a good discrimination power, but the discriminating threshold is quite low, with no clinical relevance. Furthermore, the current model was not calibrated for the prediction of PICU mortality in the Italian population. The imbalanced outcome in the development dataset (observed rate of death = 3.7%), the different case-mix of Italian population compared to the original one and the presence of missing data that do not allow to handle with a biggest sample could be possible explanations for its poor calibration.

– Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S. & Bertolini, G. Assessing the Calibration of Dichotomous Outcome Models with the Calibration Belt. The Stata Journal 17, 1003–1014 (2017)
– Kull, M., Silva Filho, T. M. & Flach, P. Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. in Electronic Journal of Statistics vol. 54 (Journal of Machine Learning Research, 2017)

WP81 Reporting quality of protocols for studies developing and validating a prediction model: a systematic review

Dhiman P.*, Ma J., Kirtley S., Waldron C., Mouka E., Whittle R., Collins G.
University of Oxford ~ Oxford ~ United Kingdom

Protocol development is a crucial step in mapping out any research project and provides key details on the rationale, objectives, design, methodology, statistical considerations and organisation of a study[1]. While not mandatory, protocols for prediction model research are key to improving the quality of models, ensuring study design and analysis features are considered before conducting the study (e.g., prespecifying the required sample size (with a sample size calculation), and approaches for dealing with missing data). However, the prevalence protocol development for prediction model research and their subsequent reporting quality is unknown. This study aims to review the reporting quality of protocols of studies developing or validating a prediction model for any clinical specialty. We searched MEDLINE and Embase for prediction model study protocols published between 01/12/2020 and 31/12/2022. We reviewed the reporting quality of the protocols against the Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD) statement[2] (excluding items relating to results). Thirty published protocols were included (22 development only, six development with validation and two validation only); most were oncology related (20%); for prognostic (15%) and binary outcomes (63%). A sample size calculation or justification was reported for 89% (n=25) of protocols developing a model, of which nine used recommended sample size formulae to estimate the required sample size to precisely estimate overall risk and minimise overfitting. Twenty-one protocols reported methods to handle missing data, of which 14 used recommended imputation approaches. 22 protocols reported internal validation methods (79%), of which 13 reported recommended resampling approaches (e.g, bootstrapping). Twenty-one protocols detailed a clustered data structure, but only 8 protocols reported methods to handle clustering. 80% of studies did not report study registration but most reported ethics approval (80%). Prediction model protocols indicate that they are more likely to follow recommended guidelines for conduct (e.g, reporting, sample size). We therefore encourage that all prediction model studies develop a protocol before carrying out any analysis. Reporting guidance is needed to ensure all relevant aspects are considered when developing a protocol for prediction model research.

[1] Peat G, Riley RD, Croft P, et al. PLOS Medicine, 11, 2014,e1001671.
[2] Collins GS, Reitsma JB, Altman DG, Ann Intern Med, 162, 2015, 55–63.

WP82 A systematic review of sample size calculation for machine learning prediction model studies in oncology

Tsegaye B.[1], Snell K.[2], Archer L.[2], Kirtley S.[1], Sperrin M.[3], Van Calster B.[4], Riley R.[2], Collins G.[1], Dhiman P.*[1]
[1]University of Oxford ~ Oxford ~ United Kingdom, [2]University of Birmingham ~ Birmingham ~ United Kingdom, [3]University of Manchester ~ Manchester ~ United Kingdom, [4]KU Leuven ~ Leuven ~ Belgium

Researchers have developed many clinical prediction models to produce an estimate of the probability (or risk) of cancer and other cancer related health events for individual patients. With increased patient data availability (e.g., electronic health records) and use of machine learning methods to capture the complex patterns; the number of prediction models in oncology continues to rise. However, studies have found that many of these models are at high risk of bias and are often developed and validated using inadequate sample sizes[1]. The aim of this study is to review how sample size is considered and how much data is used to develop and validate machine learning prediction models for a binary outcome in oncology. We searched MEDLINE for prediction model studies, developed using machine learning methods, published between 01/12/2022 and 31/12/2022. We reviewed the sample size calculations and investigated the sample sizes used to develop and validate the prediction models. We calculated the minimum sample size needed to estimate the overall risk and minimise overfitting for each developed model (Riley et al sample size criteria[2]) and summarised the difference between the calculated and used sample size. Of the 41 included studies, sample size justification was rarely provided. Sample sizes were often reduced by excluding missing data, and randomly splitting into training, tuning and test datasets. Few studies met the minimum required sample size to estimate the overall risk and even fewer met the minimum required sample size to both estimate the overall risk and minimise overfitting. Sample size calculation and justification is rarely reported in studies developing and validating a prediction model for a binary outcome using machine learning. Studies often do not use enough data to meet minimum sample size requirements for their prediction model scenario and exacerbate the issue of using insufficient data by randomly splitting their data and excluding missing data. We strongly encourage researchers to fully and transparently perform and report their sample size calculations, so they meet minimum sample size and reporting requirements for their studies.

[1] Navarro CLA, Damen JAA, Takada T, BMJ, 375, 2021,n2281.
[2[Riley RD, Ensor J, Snell KIE, BMJ,368, 2020, m441.

Poster Sessions

WP83 Baseline prediction model of time-to-flare in rheumatoid arthritis after dmard cessation: the bio-flare study

Hiu S.*[1], Rayner F.[2], Melville A.[3], Bigirumurame T.[1], Anderson A.[2], Dyke B.[4], Kerrigan S.[3], Mcgucken A.[3], Prichard J.[1], Shojaei Shahrokhabadi M.[1], Hilkens C.[2], Buckley C.[4], Mcinnes I.[3], Ng W.[2], Goodyear C.[3], Teare M.D.[1], Filer A.[4], Siebert S.[3], Raza K.[4], Pratt A.[2], Baker K.F.[2], Isaacs J.[2]

[1]Population Health Sciences Institute, Newcastle University ~ Newcastle upon Tyne ~ United Kingdom, [2]Translational and Clinical Research Institute, Newcastle University ~ Newcastle upon Tyne ~ United Kingdom, [3]School of Infection and Immunity, University of Glasgow ~ Glasgow ~ United Kingdom, [4]Institute for Inflammation and Ageing, University of Birmingham ~ Birmingham ~ United Kingdom

Rheumatoid arthritis (RA) is a chronic immune-mediated inflammatory disease. Increases in disease activity manifest as pain and swelling in joints ("flares"). Disease-modifying anti-rheumatic drugs (DMARDs) can treat RA but have side effects, thus drug-free remission is a goal. Several clinical characteristics are known risk factors of flare but reliable prediction has not been achieved. This study explores a prediction model of flare following DMARD cessation using routinely collected baseline clinical characteristics. BIO-FLARE was a multi-centre, prospective, experimental medicine study. N=121 RA patients who were receiving conventional synthetic DMARDs (csDMARD) and achieved remission were recruited (09/2018–12/2020). The non-randomised, open label, single-arm intervention was the complete cessation of csDMARDs. Study visits were at baseline (before cessation) and follow-up at weeks 2, 5, 8, 12, and 24. Flare was defined at follow-up as DAS28-CRP≥3.2 or DAS28-CRP≥2.4 on two separate occasions within two-weeks, allowing ad hoc visits. The outcome was days-to-flare. All modelling was performed with M=10 MICE datasets and B=200 bootstrap replicates per imputation. Sixteen candidate predictors were considered a priori during discussions before running any analysis. Variable selection was performed with regularised Cox PH models with an elastic net penalty; 10-fold cross validation was used to select optimal mixing and tuning parameters. A variable was selected if it had ≥60% bootstrap inclusion frequency across all MICE-bootstrap models. Non-linear functional forms were explored using univariate fractional polynomials with the RA2 closed test procedure. We reduced overfitting by estimating heuristic shrinkage factors using bootstrapping. We internally validated our final model by assessing its optimism-corrected C index and calibration slope, and calibration plot at week 24. The baseline prediction model comprised of four variables: sex (female/male), methotrexate use (yes/no), rheumatoid factor (IU/ml), and anti-citrullinated peptide antibody (IU/ml), and had acceptable discrimination (C_optimism-corrected=0.71) and good calibration (Calibration Slope_optimism-corrected=1.00). The baseline prediction model has potential to aid individualised clinical decision-making to gauge risk of flare following csDMARD cessation, and benchmarking for comparison against future models integrating biological data to improve predictive performance. The present model should be restricted to research purposes due to absence of external validation.

[1] Rayner F, Anderson AE, Baker KF, Buckley CD, Dyke B, Fenton S, et al. BIOlogical Factors that Limit sustAined Remission in rhEumatoid arthritis (the BIO-FLARE study): protocol for a non- randomised longitudinal cohort study. BMC rheumatology. 2021;5.
[2] Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: Springer Cham; 2019.

WP84 A systematic review of prediction models for transition to psychosis in individuals meeting arms criteria

Hunt A.*, Bonnett L.
University of Liverpool ~ Liverpool ~ United Kingdom

Psychotic disorders affect 1% of the UK and the At Risk Mental State (ARMS) criteria identifies individuals at high risk for psychotic disorders. [1] Although multivariable models exist to predict an individual's risk of transition to psychosis, there is a need for improvement, as currently only about 18% of individuals meeting ARMS criteria transition within 12 months.[2] The project aimed to review evidence about existing and validating prediction models for the transition of psychosis within the ARMS criteria and identify variables included. The review critically appraises the risk of bias of the identified models to determine whether reliable prediction models exist. The systematic review was conducted to assess the risk of bias in other, similar, prediction models. The following bibliographic databases were searched: PsycINFO, Medline, EMBASE and CINAHL from 1994 to 2022. Titles, abstracts and subsequently full texts were screened by two independent reviewers using a predefined inclusion criteria. Study quality was assessed using the Prediction model Risk Of Bias ASsessment Tool (PROBAST). The systematic review identified 69 unique prediction models related to a risk of transition to psychosis. However, the quality assessment highlighted only 4 unique prediction models presenting an overall low risk of bias, with many other studies being insufficiently powered for the number of candidate predictors, or lacking enough information to draw a conclusion regarding risk of bias. Predictors found to be consistently important with prediction transition were age, gender, global functioning score, genetic vulnerability and unusual thought content. The systematic review provided critical insight into the choice of prognostic factors for an improved novel prediction model and reinforced the use of the TRIPOD reporting guidelines to ensure clinical practice is informed by the best possible evidence.

[1] Onwumere, J., D. Shiers, and C. Chew-Graham, Understanding the needs of carers of people with psychosis in primary care. 2016, British Journal of General Practice. p. 400-401.
[2] Fusar-Poli, P., et al., Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. Archives of general psychiatry, 2012. 69(3): p. 220-229.

WP85 Clinical utility of predcit a prognostic tool in breast cancer: decision curve analysis

Kannan S.*[1], Hui P.[1], Kothari B.[1], Hawaldar R.[1], Vanmali V.[2], Nair N.[1], Gupta S.[1]
[1]ACTREC, Tata Memorial Centre ~ Navi Mumbai ~ India, [2]Tata Memorial Hospital ~ Mumbai ~ India

Clinicians who want to calculate a patient's risk (or likelihood) of suffering a future event must use prediction models for survival outcomes. A model's statistical prediction power is quantified by measures of discrimination and calibration. Unfortunately, they fall short in assessing if the model may truly enhance clinical decision making. To address this, we carried out an independent external validation study and used decision curve analysis to evaluate the clinical usefulness. Women treated for operable breast cancer between 2008 and 2016 at Tata Memorial Centre in Mumbai, India, with non-metastatic, T1-T2 invasive cancer, HER2neu receptor negative, and 5 years of survival data were chosen for the external validation research. The online PREDICT program version 2.2.1 provided the anticipated 5-year Overall Survival (OS) rate for each patient. The Chi- squared goodness of fit test and the area under the receiver operating characteristic curve were used to evaluate calibration and discriminatory accuracy (AUC). The PREDICT tool's clinical usefulness was assessed using decision curve analysis (DCA), which quantified the net benefits at various threshold probabilities2. The prediction model was tested on the validation dataset for the entire cohort as well as specific subgroups such as the elderly (>=65 years), ER/PR positive, systemic medication, and hormone therapy. STATA version 15 was used to analyse the data for decision curve analysis. For the analysis, 2783 suitable individuals were chosen from a total of 11670 patients registered at the hospital. The discriminatory accuracy for 5-year survival was acceptable (AUC-0.647). The PREDICT tool could not deliver a larger net benefit (NB) across a wider tolerable range of threshold probabilities for forecasting OS, according to decision curve analysis. Unlike standard statistical approaches, which solely assess a prediction model's accuracy, decision curve analysis can tell us whether adopting a model to enhance clinical decision-making will improve patient outcomes. The 5-year OS was underestimated by the PREDICT tool. Clinical utility was insufficient for both the entire cohort and the subgroups.

1. https://breast.predict.nhs.uk
2. Vickers, A. J. et al. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med. Inform. Decis. Mak. 8, 53 (2008).

WP86 Building centile charts of ecg parameters in children from 0 to 16 years old: an italian cross-sectional study

Khaleghi Hashemian D.*[1], Cortinovis I.[1], Bongiorno A.[2], Seganti A.[3], Mannarino S.[4], Badilini F.[5], Ferraroni M.[6], Sanzo A.[7]
[1]Laboratory of Medical Statistics and Biometry "Giulio A. Maccacaro", Department of Clinical Sciences and Community Health, Università degli Studi di Milano ~ Milan ~ Italy, [2]Department of Molecular Medicine, Division of Cardiology, University of Pavia ~ Pavia ~ Italy, [3]Cardiology Department, Fondazione IRCCS Policlinico San Matteo ~ Pavia ~ Italy, [4]Pediatric Cardiology Unit, Pediatric Department, Buzzi Children's Hospital ~ Milan ~ Italy, [5]AMPS-LLC ~ New York ~ United States of America, [6]Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Ca' Granda Ospedale Maggiore Policlinico ~ Milan ~ Italy, [7]Coronary Care Unit and Laboratory of Clinical and Experimental Cardiology, Fondazione IRCCS Policlinico San Matteo ~ Pavia ~ Italy

Literature regarding pediatric electrocardiogram is poor or outdated, if we focus on growth and development of heart. The most recent European work is P.R. Rijnbeek (2001) [1]. Our study aims to redefine some of the parameters and normal values in ECG's findings, building new growth charts (centile charts) for use in clinical assessment. We collected 54839 ECG records in a cross-sectional study (47% females, 53% males) for children from 0 days to 16 years old with normal weight and height. We excluded ECG records of children with known cardiovascular anomalies, lost during follow-up (at least three years) or deceased. We modelled three parameters (i.e., sinus rate, PR interval, and global QRS duration) to build reference centiles for ECG. First, we needed to be sure the total variability during time was the same and due to skewed distribution, we used a transformation to model sinus rate and PR interval. We fitted a linear regression model for each ECG parameter with time at measurement as independent variable, separately for sex and age (≤6 and >6 months). Thanks to the model, we built reference centiles using absolute residuals [2]. To check the accuracy of our prediction model we verified that the data were adequately divided in respect to frequency distribution above and below the centiles. Final step was to create graph of centile charts visualizing scatter plot of the data and 2nd, 50th and 98th percentile lines. The values of our population's ECG parameters seem to be representative of physiological growth of the heart, so they can be used to create growth charts. Our method has proven to be effective in building percentile charts in the field of pediatric cardiology. Further work is needed to verify if effectiveness is maintained for all ECG parameters available in our dataset. Using percentile charts is a tool that can be practical for pediatrician to consult and useful in clinical assessment. A priori hypothesis of equal distribution of clinical parameters within age classes (≤6 and >6 months) still need to be verified.

[1] P. R. Rijnbeek, M. Witsenburg, E. Schrama, J. Hess, J. A. Kors, "New normal limits for the paediatric electrocardiogram", European heart journal, vol. 22, no. 8, pp. 702–711, 2001.
[2] D. Altman, "Construction of age-related reference centiles using absolute residuals", Statistics in medicine, vol. 12, no. 10, pp. 917-924, 1993.

WP87 Uncertainty quantification in the predictions of a ebola outbreak using bayesian data assimilation

Krishnamurthy A.*[1], White T.B.[2]
[1]Mount Royal University ~ Calgary ~ Canada, [2]University of Exeter ~ Exeter ~ United Kingdom

Mathematical modelling of infectious diseases is an interdisciplinary area of increasing interest. We present a spatial variant of the common SEIR (Susceptible-Exposed-Infectious-Recovered) model of epidemiology to capture the transmission dynamics of the spread of Ebola in separate outbreaks in two Central African countries, the 2018-2020 outbreak in the Democratic Republic of Congo (DRC) and the 2022-2023 outbreak in Uganda. The challenge is that population mobility is low in this area of Africa, and distances are long, so spatial epidemics tend to burn out, or they expand only slowly. Predicting the transmission dynamics of Ebola is challenging and comes with a lot of uncertainty. The goal of this research is to quantify this uncertainty and provide insight that would support public health officials towards informed, data-driven decision making. The ensemble optimal statistical interpolation data assimilation method has been shown to produce optimal Bayesian statistical tracking of emerging epidemics[1]. Our simulations show good correspondences between the stochastic model and the available sparse empirical data. A comparison between weekly incidence data set and our compartment model coupled with Bayesian data assimilation highlights the role of a realization conditioned on all prior data and newly arrived data. In general, the compartmental model with data assimilation gives a better fit than the model without data assimilation for the same time period. We present spatio-temporal disease maps for the infectious variable for the progress of Ebola in the North-Kivu and Ituri provinces of DRC during 2018-2020. This case study was conducted using real-world data from the WHO and practical simulation exercises using free and open-source software. We train the model using the DRC data and test it on the 2022- 2023 Uganda data to project the prevalence and deaths. Our analyses sheds light more broadly on how Ebola spreads in a large geographical area with places where no empirical data is recorded or observed. The analysis presented herein can be applied to a large class of compartmental epidemic models.

[1] L. Cobb, A. Krishnamurthy, J. Mandel, and J. Beezley, Bayesian Tracking of Emerging Epidemics using Data Assimilation Methods, Spatial and Spatio-Temporal Epidemiology, vol. 10, 2014, 39-48.

Poster Sessions

WP88 Do researchers consider the time-dependent interplay between time to assessment and severity in acute stroke?

Lu Z.*, Gittins M., Kishore A., Smith C., Vail A.
The University of Manchester ~ Manchester ~ United Kingdom

Characterised by sudden onset of clinical symptoms, stroke progresses rapidly in the first few hours following ictus. Stroke severity on admission, standardly measured using the National Institutes of Health Stroke Scale (NIHSS), is routinely recorded, and typically used to predict further events. Services seek to provide stroke care as early as possible. In practice, there is a complex inter-dependence between severity of developing symptoms and the time taken to be assessed. Symptom severity may evolve as the stroke progresses and as spontaneous recovery takes place. Our impression was that 'time to assessment' is rarely incorporated in practice. However, in perinatal research it is routine to include the analogous temporal variable (gestational age) when using birthweight to model outcomes. Our primary objective was to review whether and, if so, how stroke researchers account for the dynamic inter-relationship between 'time to assessment' and 'admission NIHSS'. We sought to compare this with approaches used by neonatal researchers. Review of papers describing prognostic analyses in the two contexts above. A simple sample size calculation comparing anticipated proportions incorporating the temporal variable (80% vs 20%) required very few studies to give 90% power. We therefore chose to review just articles from leading specialty journals in 2019. We chose the journals for informal 'quality' purposes and the year as the most recent in which collection of clinical data could not have been affected by the pandemic. For identified papers, we categorised the approaches according to the nature of the outcome (binary, time to event). Our primary outcome was whether or not the temporal variable was included. If included, we noted whether by a centile or a regression approach. We further categorised by parametric/non-parametric and continuous/categorical methods. Final results are in progress. Very few studies were identified in the stroke literature that addressed this issue. Conversely, it was routine in the neonatal literature. There is clear scope to improve statistical methods used in stroke research. Further research will be undertaken to consider the factors that determine the extent to which this is worthwhile.

Clausson, B. et al. (2001). Perinatal outcome in SGA births defined by customised versus population- based birthweight standards. BJOG : an international journal of obstetrics and gynaecology, 108(8), pp.830–834. Salas, A.A. et al. (2016). Gestational age and birthweight for risk assessment of neurodevelopmental impairment or death in extremely preterm infants. Archives of disease in childhood. Fetal and neonatal edition, 101(6), pp.F494–F501.
Vieira, M.C. et al. (2019). Determination of birth-weight centile thresholds associated with adverse perinatal outcomes using population, customised, and Intergrowth charts: A Swedish population-based cohort study. PLoS medicine, 16(9), pp.e1002902–e1002902.
Ausbeck, E.B. et al. (2020). 1000: Gestational age versus birthweight to predict outcomes in neonates with extreme prematurity. American journal of obstetrics and gynecology, 222(1), pp.S621–S622.
Markus, H.S. and Bevan, S. (2014). Mechanisms and treatment of ischaemic stroke--insights from genetic associations. Nature reviews. Neurology, 10(12), pp.723–730.
Counsell, C. and Dennis, M. (2001). Systematic Review of Prognostic Models in Patients with Acute Stroke. Cerebrovascular diseases (Basel, Switzerland), 12(3), pp.159–170.
Mistry, E.A. et al. (2021). Predicting 90-Day Outcome After Thrombectomy: Baseline-Adjusted 24-Hour NIHSS Is More Powerful Than NIHSS Score Change. Stroke (1970), 52(8), pp.2547–2553.
Riley, R.D. et al. (2019). Prognosis research in healthcare: concepts, methods, and impact. First edition. R. D. Riley et al., eds. Oxford, United Kingdom: Oxford University Press.
SCHLEGEL, D. et al. (2003). Utility of the NIH Stroke Scale as a predictor of hospital disposition. Stroke (1970), 34(1), pp.134–137.

WP89 Individual treatment effects of sodium-glucose cotransporter-2 inhibitors on chronic kidney disease risk

Lukkunaprasit T.*[1], Siriyotha S.[2], Looareesuwan P.[2], Thakkinstian A.[2]
[1]College of Pharmacy, Rangsit University ~ Pathum Thani ~ Thailand, [2]Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University ~ Bangkok ~ Thailand

Previous evidence suggested average benefits of sodium-glucose cotransporter-2 inhibitors (SGLT2i) in lowering risk of chronic kidney disease (CKD) in patients with type 2 diabetes (T2D)[1] at population level. However, it is still unknown which patients will gain less or more benefits. Therefore, this study aimed to estimate individual treatment effects (ITE) of SGLT2i on lowering CKD risk by applying a counterfactual prediction model. The analysis was performed using data from T2D cohort at Ramathibodi Hospital from 2010 to 2019. A multivariate logistic regression model was constructed to predict factual and counterfactual risks of CKD if patients did and did not receive SGLT2i treatment. The ITEs of SGLT2i were estimated by subtracting the factual risk of CKD from the counterfactual risk of CKD in individual T2D patients[2]. Of 24,777 patients in T2D cohort, there were 316 and 24,461 in SGLT2i and non-SGLT2i groups, respectively. Important predictors included in the counterfactual model were SGLT2i, age, sex, time since first receiving metformin, hypertension, peripheral arterial disease, diabetic retinopathy, and HDL-C. The estimated ITE varied from -0.141 to -0.005 with a median of -0.028. This could be interpreted that about 50% of patients would have ≥ 2.8% lowering CKD risk, and even about 40.2% of them would have ≥ 3% lowering CKD risk if they received SGLT2i compared to if they did not. Patients who gained the greatest benefit from SGLT2i were males, aged ≥ 65 years, had hypertension, peripheral arterial disease, diabetic retinopathy, low HDL-C, and received metformin ≥ 6 months. Our findings provided information of individualized patients who may gain more benefits from SGLT2i use. Our prediction model could be useful for clinical decision making and personalized medicines given characteristics of patients with T2D.

[1] S. Siriyotha, T. Lukkunaprasit, P. Looareesuwan, H. Nimitphong, G.J. McKay, J. Attia, A. Thakkinstian, Effects of second-line antihyperglycemic drugs on the risk of chronic kidney disease: applying a target trial approach to a hospital-based cohort of Thai patients with type 2 diabetes, Cardiovasc Diabetol, 21, 2022, 248.
[2] M. Falcaro, R.B. Newson, P. Sasieni, Stata tip 146: Using margins after a Poisson regression model to estimate the number of events prevented by an intervention. The Stata Journal, 22, 2022, 460–464.

WP90 Predicting number of events using joint model in clinical trials with a time- to-event endpoint

Machida R.*[1], Sakamaki K.[2], Sozu T.[3]
[1]Biostatistics Division, Center for Research Administration and Support, National Cancer Center ~ Tokyo ~ Japan, [2]Center for Data Science, Yokohama City University ~ Yokohama ~ Japan, [3]Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science ~ Tokyo ~ Japan

In event-driven clinical trials with a time-to-event endpoint, such as overall survival and progression- free survival, the primary analysis is conducted when the required number of events is observed. Thus, the study duration has uncertainty, and its evaluation is essential to manage costs, personnel, and labor of the planned study. The number of events (NE) at a certain time point has to be predicted considering the enrollment scheme and distribution of time-to-event outcome and censoring to predict the study duration accurately. In such cases, the precision of predicting the NE would be increased if a longitudinally measured covariate associated with the time-to-event outcome is available after starting the study and incorporated into the prediction method of the NE. We propose a method for predicting the NE during a study using the joint model for the time-to-event outcome and the covariate and evaluate its operating characteristics. We consider predicting the NE in prostate cancer clinical trials, where the prostate-specific antigen level of each participant is longitudinally measured [1]. Participants at a certain time point are classified into three conditions: (a) participant with event, (b) participant without event, and (c) participant not enrolled (potential participant), if lost to follow-up participant is not assumed. The proposed method consists of the following four steps. Step 1: estimating the joint model parameters using interim data of participants (a) and (b). Step 2: generating and imputing date of enrollment and baseline covariates of potential participant (c). Step 3: calculating the probability of occurrence of an event at a future time point in each participant (b) and (c), using the accumulated data until the current time point [2]. Step 4: calculating the expected total number of events from the sums of the number of participants (a) and probabilities of occurrence of events calculated in Step 3. We demonstrate that the proposed method increases the precision of predicting the NE compared with methods using no covariates. The proposed method using the joint model would be useful in predicting the study duration accurately if a longitudinally measured covariate is available.

[1] A. Finelli, T. M. Beer, S, Chowdhury et al., Comparison of Joint and Landmark Modeling for Predicting Cancer Progression in Men With Castration-Resistant Prostate Cancer. A Secondary Post Hoc Analysis of the PREVAIL Randomized Clinical Trial, JAMA Network Open, 4, 2021, e2112426- e2112426.
[2] D. Rizopoulos, Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data, Biometrics, 67, 2011, 819-829.

WP91 Model predictions of the effects of rapid art on hiv and aids

Omata K.*[1], Imahashi M.[2], Noda T.[3], Taniguchi T.[4]
[1]National Center for Global Health and Medicine ~ Tokyo ~ Japan, [2]Nagoya Medical Center ~ Nagoya ~ Japan, [3]Nara Medical University ~ Kashihara ~ Japan, [4]Chiba University ~ Chiba ~ Japan

The antiretroviral therapy (ART) is expected to contribute not only to the treatment of people living with HIV, but also to the reduction of new infections. Currently, the average time between diagnosis and initiation of treatment for HIV infection is estimated to be approximately 40 days in Japan. The purpose of this study is to estimate the future effects of Rapid ART, which shortens this period. Assuming HIV infection in MSM (men who have sex with men) communities, estimates were made using a compartmental model which employs five compartments: susceptible, infectious, diagnosed, AIDS, and treated. The input values were adopted from the number of new HIV infections and AIDS cases reported by the AIDS Prevention Information Network in Japan, and the analysis of clinical data at Nagoya Medical Center and at the National Center for Global Health and Medicine in Japan. By changing the average time from diagnosis to initiation of treatment for HIV infection from 42 days to 0 days (i.e., immediate start of treatment), it was estimated that if this shortening was continued for a certain number of years, the number of new infections per year would decrease by about 45%; it was estimated that the total number of infected people would decrease by about 80%, the number of AIDS cases would decrease by 45%, and uninfected MSM would increase by about 1% due to the reduced risk of infection.
Even a few tens of days in reducing the average time between diagnosis and start of treatment could have a significant effect on reducing HIV infection if this was continued. This effect of Rapid ART is related to the second and third 95 of the UNAIDS 95–95–95 testing and treatment targets [1], which, together with the implementation of the first 95, can be an important HIV control measure.

[1] UNAIDS. Global AIDS Strategy 2021-2026 (2021). https://www.unaids.org/en/Global-AIDS- Strategy-2021-2026.

WP92 Quantifying the added predictive value of prs in cvd risk prediction tools in individuals with morbidity

Petitjean C.*, Ip S., Lambert S., Inouye M., Wood A.
British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge ~ Cambridge ~ United Kingdom

Recently it has become possible to predict individual's genetic predisposition to diseases using polygenic risk scores (PRS), where studies show that PRS may improve cardiovascular disease (CVD) risk prediction model performance [1]. The use of PRS in clinical practice is likely to be first implemented in vulnerable patients, such as patients with pre-existing morbidity. The aim of this study is to compare the added predictive value of PRS in a CVD risk prediction model containing the conventional risk factors, in individuals with and without type 2 diabetes, depression or a history of cancer; conditions which are known to have elevated CVD risk.
Using data from the UK Biobank, we derived Cox regression models considering conventional CVD risk factors with and without PRS in the general population. We assessed the models' calibration and discrimination in different subgroups. On addition of the PRS in the general population, the model's C- index increased from 0.732 (95%CI=0.732-0.732) to 0.740 (0.739-0.740) demonstrating an increase of 0.008 (0.008-0.008). The increase was 50% lower in individuals with diabetes (C-index without PRS= 0.645 [0.642-0.647]; C-index with PRS= 0.648 [0.645-0.651], difference=0.004 [0.003-0.004]) or with depression (C-index without PRS= 0.727 [0.724-0.729], C-index with PRS= 0.730 [0.728-0.733], difference=0.004 [0.004-0.004]), and 5% lower in individuals with a cancer history (C-index without PRS= 0.693 [0.691-0.695], C-index with PRS= 0.701 [0.699-0.702], difference=0.008 [0.007-0.008]). To assess the prognostic contribution of each risk factor, agnostic to the sequence of its' addition to the model, we calculated Shapley values of the C-index. The Shapley values demonstrated that the PRS contributed 12% to the model's discrimination in the general population, and 11%, 8% and 14% in individuals with type 2 diabetes, depression or a history of cancer, respectively. The added predictive value of PRS in a conventional CVD risk prediction tool in the individuals with diabetes or depression is less than in the general population. However, the PRS's relative contribution to the model discrimination is similar in all subgroups.

[1] Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, et al. (2021) Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLoS Med 18(1): e1003498. https://doi.org/10.1371/journal.pmed.1003498

Poster Sessions

WP93 How correlations between markers influence the net benefit increase of a predicitve model

Sabroso--Lasa S.*[1], Jurado F.J.[1], Malats N.[1], Esteban L.M.[2], Alcalá--Nalvaiz J.T.[3]
[1]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO) ~ Madrid ~ Spain, [2]Department of Applied Mathematics, Escuela Universitaria Politécnica La Almunia, University of Zaragoza ~ La Almunia, Zaragoza ~ Spain, [3]Department of Statistical Methods, University of Zaragoza. ~ Zaragoza ~ Spain

Although the factors affecting the selection of biomarkers to be combined are still under debate, a large number of studies are analyzing how correlations between markers influence the increase the diagnostic performance of biomarker scores. However, the vast majority of these analyses are usually assessed using accuracy metrics that do not consider clinical consequences. The objective of our research is to analyze the effect of the sign and magnitude of the correlations between markers in the improvement of the Net Benefit (NB) of a predictive model proposed by Vickers[1]. For this purpose, different scenarios through simulations and a practical example have been developed. Simulations have been carried out both in situations where multivariate normality is assumed and in cases where skewed data are used for which log-normal distribution simulated data have been selected. Furthermore, the NB was assessed by combining 1934 distinct blood lipid metabolites determined by liquid chromatography in 44 pancreatic cancer (PC) cases and 38 hospital controls from the PanGenMic Study, conducted as real-data practical example. In all cases, we used heatmaps to graphically observe the relationship of the increase in the NB with the sign and magnitude of correlations. Results showed that, depending on the discrimination ability of each marker on its own, large positive or negative correlations are important to achieve higher NBs. For markers with a similar discrimination ability, only negative correlation aids to increase the NB of the combination of markers. However, when markers with different predictive power are combined, high positive correlations also improve the clinical utility of the model, but with less effect. Correlations between biomarkers play a fundamental role in the NB of the models and, therefore, in their clinical utility. Thus, considering the correlation between markers when constructing clinical models should always be a step to be considered.

[1] A. J. Vickers, E. B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, 26(6), 2006, 567-74.

WP94 Development of a risk calculator and web application for kidney transplantation with r/shiny

Schwab S.*[1], Sidler D.[2], Haidar F.[3], Kuhn C.[4], Schaub S.[5], Koller M.[5], Mellac K.[5], Thaqi A.[6], Stürzinger U.[5], Tischhauser B.[5], Binet I.[4], Golshayan D.[7], Müller T.[8], Elmer A.[1], Franscini N.[1], Krügel N.[1], Fehr T.[9], Immer F.[1]
[1]Swisstransplant ~ Bern ~ Switzerland, [2]Inselspital, Bern University Hospital ~ Bern ~ Switzerland, [3]University Hospital of Geneva ~ Geneva ~ Switzerland, [4]Kantonsspital St. Gallen ~ St. Gallen ~ Switzerland, [5]University Hospital Basel ~ Basel ~ Switzerland, [6]Federal Office for Public Health ~ Bern ~ Switzerland, [7]Lausanne University Hospital ~ Lausanne ~ Switzerland, [8]University Hospital Zurich ~ Zurich ~ Switzerland, [9]Cantonal Hospital Graubünden ~ Chur ~ Switzerland

In Switzerland, the kidney waiting list contains the largest number of patients and has the longest average waiting time compared to other organs. Measures have been taken to increase the donor pool, among them the transplantation of marginal donor kidneys (lower quality organs with still acceptable medical risks). However, in Switzerland, no widely accepted prognostic instrument for transplantation outcomes is routinely used in clinical practice so far. Therefore, we are currently developing a risk calculator based on our prediction model, KIDMO (kidney prediction model). We will focus on two aspects of this large project: the implementation of the risk calculator in R/Shiny and the involvement of expert clinicians and patients in the design of the tool. The clinical prediction model is developed based on data from a national multi-center cohort study (Swiss Transplant Cohort Study; STCS) and the Swiss Organ Allocation System (SOAS) with over 2,700 kidney recipients transplanted between 2008 and 2021. The primary outcome was kidney graft survival (with death of the recipient as competing risk using the Fine & Grey model); a study protocol is available [1]. The majority of published prediction models in the medical literature have no clinical utility. Therefore, the KIDMO prediction model has been aligned with best practices such as the involvement of expert clinicians and patients in the design of the tool, as well as the publication of a study protocol that prespecified the candidate predictors and the planned statistical analyses. S. Schwab, D. Sidler, F. Haidar, C. Kuhn S. Schaub, M. Koller, K. Mellac, et al. Diagnostic and Prognostic Research, 7(1), 2023, 6. https://doi.org/10.1186/s41512-022-00139-5

WP95 Impact of sample size and events per predictors on performance & stability of psychosis risk prediction models

Gutvillig M.[1], Oliver D.[2], Fusar--Poli P.[1], Stahl D.*[1]
[1]King's College London ~ London ~ United Kingdom, [2]Oxford University ~ Oxford ~ United Kingdom

Introduction: Clinical prediction models are frequently developed on small samples leading to overly optimistic performance and unstable and imprecise estimates. The appropriate sample size for clinical prediction models remains unresolved with sample size simulation studies often limited using artificially simulated data that does not reflect the complexities of applied research. Recent advancements in sample size estimation for the development of clinical prediction models have concentrated on ensuring precise predictions and limiting overfitting. However, these developments focus on using the generalized linear model (2), and may not be appropriate for statistical learning extensions such as regularized methods. Objective: To systematically assess the influence of total sample size and events per candidate predictor parameter (EPP) on a psychosis prediction model (1), a resampling study was conducted using a previously developed and validated risk calculator as a case study (predictors: age, gender, age by gender, ethnicity, International Classification of Diseases (ICD)-10 index diagnosis, 14 symptom and substance use variables). Unlike the original model using a split sample approach, the risk calculator's validation procedure was updated to internal-external validation. A dataset consisting of the electronic health records of 142 923 patients with non-organic non-psychotic mental disorders was used to first examine the internal-external validity of the risk calculator implemented with Leave-One-Study-Site-Out Cross-Validation. Time to a psychosis diagnosis was modelled using Least Absolute Shrinkage and Selection Operator (LASSO) regularised Cox regression. Next, the validated performance (discrimination and calibration), optimism, and variable selection of the model alongside the preciseness of estimates were examined across 28 simulation scenarios (seven EPPs under four total sample sizes). Results allow for providing guidelines about minimum sample sizes for similar clinical datasets and highlight the interconnectedness of total sample size and EPP. Comparison with simulations using standard cox regression will be discussed.

(1) Fusar-Poli P, Rutigliano G, Stahl D, Davies C, Bonoldi I, Reilly T, McGuire P. Development and Validation of a Clinically Based Risk Calculator for the Transdiagnostic Prediction of Psychosis. JAMA Psychiatry. 2017 May 1;74(5):493-500. doi: 10.1001/jamapsychiatry.2017.0284. Erratum in: JAMA Psychiatry. 2018 Jul 1;75(7):759. PMID: 28355424; PMCID: PMC5470394.
(2) Riley, RD, Snell, KIE, Ensor, J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Statistics in Medicine. 2019; 38: 1276– 1296. https://doi.org/10.1002/sim.7992

WP96 Evaluation of systematic reviews of prognostic models for covid-19: an overview of systematic reviews.

Talimtzi P.*[2], Ntolkeras A.[3], Kostopoulos G.[1], Bougioukas K.[2], Pagkalidou E.[2], Ouranidis A.[4], Pataka A.[5], Dardavesis T.[2], Haidich A.[2]
[1]Department of Endocrinology, 424 General Military Hospital ~ Thessaloniki ~ Greece, [2]Department of Hygiene, Social- Preventive Medicine and Medical Statistics, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki ~ Thessaloniki ~ Greece, [3]School of Biology, Aristotle University of Thessaloniki ~ Thessaloniki ~ Greece, [4]Department of Pharmaceutical Technology, School of Pharmacy, Aristotle University of Thessaloniki, 54124 Thessaloniki ~ Thessaloniki ~ Greece, [5]Department of Respiratory Deficiency, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki ~ Thessaloniki ~ Greece

During the COVID-19 pandemic, there was an abundance of prognostic models for the diagnosis and prognosis of this new coronavirus. Many studies have tried to review and critically appraise these models [1, 2]. The purpose of this study is to evaluate the reporting completeness and transparency of systematic reviews (SRs) of prognostic modelsfor COVID-19. MEDLINE and Epistemonikos (epistemonikos.org) were searched for systematic reviews of prognostic models for COVID-19 until December 2022. The ROBIS tool was used to assess the methodological quality of the selected SRs. Ten SRs were retrieved containing from 4 to 310 primary studies and from 6 to 606 prognostic models; none of the SRs synthesized the results in a meta-analysis. Only three of the SRs had their protocols registered and publicly available in a repository for protocols and one of them had its data publicly available in the website. The majority of SRs (70%) had an overall high risk of bias resulting more often from concerns in the synthesis and reporting of findings. The overall corrected covered area (CCA) was 5.9% which shows a small amount of overlapping. Systematic reviews of prognostic models for COVID-19 should follow certain reporting guidelines to enhance transparency so that clinicians can select the appropriate prognostic model for each individual patient. Pre-established protocol, detailed information on both methodology and process followed, as well as clear reporting of findings are essential aspects to which attention should be paid.

[1] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. BMJ. 2020;369:1–22.
[2] Miller JL, Tada M, Goto M, Mohr N, Lee S. Prediction Models for Severe Manifestations and Mortality due to COVID-19: A Rapid Systematic Review. medRxiv. 2021 Jan;2021.01.28.21250718

WP97 Evaluating calibration at moderate-strong level using patient subgroups identified with clustering analysis

Wang J.*[1], Jiu L.[1], Tapia--Galisteo J.[2], Somolinos--Simón F.[2], García--Sáez G.[2], Hernando M.E.[2], Goettsch W.[1]
1Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University ~ Utrecht ~ Netherlands, 2Universidad Politécnica de Madrid ~ Madrid ~ Spain

Prediction models with poor calibration can be misleading and may result in incorrect and potentially harmful decisions. Calibration was deemed as the Achilles heel of prediction models, and the importance of model calibration has received more attention in recent years. Van Calster et al. [1] defined a hierarchy of four increasingly strict levels of calibration: mean, weak, moderate, and strong calibration. They argued that although strong calibration is desirable for individualized decision making, it is unrealistic in practice. Thus moderate calibration is a better attainable goal and can still guarantee that decisions made based on the model are clinically nonharmful. In a recent concept paper by the GRADE working group, the authors suggested to evaluate calibration for subgroups defined either with different risk categories (e.g., lower or higher risk) or based on characteristics not included in the model. In this conceptual paper, we propose a new calibration level, namely moderate-strong level, as an extension to the hierarchy proposed by Van Calster et al. The moderate-strong calibration is evaluated for subgroups identified with clustering analysis of all predictors included in the model. It can provide better assurance than moderate calibration and is still realistic for evaluation, and it is more meaningful than the subgroups proposed by GRADE [2]. The proposed methods are presented with an illustrative case study using T1D Exchange data to validate a prognostic model for type 1 diabetes patients, to demonstrate how subgroups can be identified with clustering analysis and how calibration at moderate-strong level can be evaluated with different numbers of subgroups. This illustrative example shows the potential application of clustering analysis in creating credible and meaningful subgroups, without collecting extra data than model validation, to approximate the strong calibration.

[1] B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M. J. Pencina, & E. W. Steyerberg (2016). A calibration hierarchy for risk models was defined: from utopia to empirical data. Journal of clinical epidemiology, 74, 167-176.
[2] F. Foroutan, G. Guyatt, M. Trivella, et. al (2022). GRADE concept paper 2: Concepts for judging certainty on the calibration of prognostic models in a body of validation studies. Journal of Clinical Epidemiology, 143, 202-211.

WP98 Assessing the reporting of image and ai based cvd diagnostic models: evaluation and implementation of claim

Wang J.*[1], Wang X.[2], Wang W.[2], Zhu M.[2], Guo H.[3], Ding J.[3], Sun J.[4], He K.[2]
[1]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University ~ Utrecht ~ Netherlands, [2]Medical Big Data Research Center, Chinese PLA General Hospital ~ Beijing ~ China, [3]Department of Pulmonary and Critical Care Medicine, Chinese PLA General Hospital ~ Beijing ~ China, [4]Medical School of Chinese PLA ~ Beijing ~ China

The CLAIM checklist [1] was developed as a guide for reporting diagnostic models based on medical imaging and artificial intelligence (AI). In this study, we aim to evaluate the comprehensiveness of the CLAIM, convert it into an adherence assessment form, and systematically review the adherence of image-based AI diagnostic models in cardiovascular diseases (CVD) to the CLAIM. The items in CLAIM were compared with the items in TRIPOD and recommendations for reporting machine learning models [2]. An adherence assessment form was developed by dividing each items into several sub-items with clear criterion and logical operators (conjunction or disjunction). We conducted a systematic literature search of PubMed and Embase using terms related to AI, image, and cardiovascular diseases to identify image-based AI diagnostic models in CVD and assessed the overall adherence per article and per CLAIM item and sub-item. The CLAIM checklist covered almost all items in TRIPOD and the recommendations for reporting machine learning models. In the proposed CLAIM adherence assessment form, 30 out of 42 items were divided into two (17/30), three (8/30), four (4/30), or six (1/30) sub-items. A total of 99 eligible articles were identified for the systematic review. Overall, articles adhered to a median of 64.13% (P25–P75 57.07–69.57%) of CLAIM items. Eighteen CLAIM sub-items were described in 85% or more of the 99 models, nine sub-items in less than 20% of the models. No associations were found between the reporting completeness and the impact factor of a journal (89 studies), sample size (86 studies), or the time needed for acceptance for publication (92 studies) by regression analysis. CLAIM is a comprehensive and reliable reporting guideline and can be used as an assessment tool of the adherence in image-based AI diagnostic models. The reporting quality of such models in the CVD disease field was sub-optimal.

[1] . Mongan, L. Moy, & C. E. Kahn Jr, (2020). Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiology: Artificial Intelligence, 2(2), e200029.
[2] L. M. Stevens, B. J. Mortazavi, R. C. Deo, L. Curtis, & D. P. Kao, (2020). Recommendations for reporting machine learning analyses in clinical research. Circulation: Cardiovascular Quality and Outcomes, 13(10), e006556.

Poster Sessions

WP99 The chi square test and the need to apply corrections for big data

Avram C.*[2], Avram L.[1], Marusteri M.[2]
[1]"Dimitrie Cantemir" University of Târgu-Mureș ~ Târgu-Mureș ~ Romania, [2]George Emil Palade University of Medicine, Pharmacy, Science and Technology of Târgu-Mureș ~ Târgu-Mureș ~ Romania

The chi-square test which is considered one of the fundamental tests of statistics. This test is applied in the case of categorical variables to compare the samples. The test can only be applied to real data, not to percentages, averages or other representations. In the case of the data from the contingency table in which we have a frequency lower than 10, then it is necessary to apply the Yates correction, if the frequencies are lower than 5, then we apply Ficher's Exact Test. Ficher's Exact Test is a test that is calculated completely differently from the chi-square test. The objective is to identify the limits (frequency ratio) at which corrections should be applied for this test. The current data collected contains a lot of records, that is, we have Big data. We analyzed various models for real data that were used in the articles. Applying the chi square text we have the possibility of having lower frequencies (tens) but also high frequencies (thousands). We want to identify which is the ratio to which the corrections must be applied, we have data with the maximum frequency of the order of hundreds and the minimum of the order of tens, but also situations in which the maximum frequency in the quota table is of the order of thousands and the minimum of the order of hundreds. The meta-analysis of this study will be performed on medical data. Results: The calculation of the chi-square test is easy and the interpretation of this test allows the identification of associations between the variables. Depending on the ratio of frequencies in the contingency tables, it is necessary to apply corrections or just that these corrections are currently calculated for small values. What happens if these ratios are kept but we have values in the hundreds? Are we still applying corrections? The Chi-square test is the most applied test in statistics. The need to identify when it is necessary to apply corrections for this test is very important to correctly calculate the test but also to obtain a correct result (p) that verifies the correlations between the variables. Pearson, K. Philosophical Magazine. Series 5. 50 (302), 1900, 157–175. Doi:10.1080/14786440009463897.

WP100 Evaluation of the effectiveness of pcsk9-i in a target trial emulation framework based on ehr

Barbati G.*[1], Gregorio C.[1], Scagnetto A.[2], Indennidate C.[2], Cappelletto C.[2], Di Lenarda A.[2]
[1]Biostatistics Unit, Department of Medical Sciences, University of Trieste ~ Trieste ~ Italy, [2]Cardiovascular Center, Territorial Specialistic Department, University Hospital and Health Services of Trieste ~ Trieste ~ Italy

Low-density lipoprotein (LDL) cholesterol is a modifiable risk factor for cardiovascular (CV) disease. Antibodies that inhibit proprotein convertase subtilisin–kexin type 9 (PCSK9) have emerged as a new class of drugs that effectively lower LDL levels, thus preventing CV events. We designed an observational study following the target trial emulation (TTE) approach [1] to evaluate the effectiveness of PCSK9-i using Electronic Health Records (EHR). A cohort of subjects eligible to PCSK9i was identified from July 2017 (when the drug was available) to December 2020. Administrative censoring date was set at 31-12-2021. Outcomes of interest were all- cause death and the first hospitalization for ACS (Acute Coronary Syndrome), IS (Ischemic Stroke), PAD (Peripheral Arterial Diseases) and others CV-related causes. A group of subjects initiated the PCSK9-i therapy during the enrolment period ("Treated"); index date for them was reset to the start of the therapy, updating all the covariates at that time. Propensity score was used to estimate inverse probability of treatment weights in terms of demographic and clinical covariates. On the weighted dataset, a semi-markov multi-state model was estimated using a spline-based flexible parametric specification for the hazard [2]. Eligible group comprised 1815 subjects; Treated were 161. Treated were slightly younger, prevalently males, with more severe CV conditions and a higher rate of statin treatment w.r.t. to the Eligible. Eligible had higher prevalence of non CV comorbidities. On the weighted dataset, a protective effect of PCSK9-i on death (HR=0.14, 95% CI 0.07-0.27) and a relative risk reduction for the transition towards the first hospitalization (HR=0.78, 95% CI 0.62-0.98), common to all the causes under study, was observed. This is the first observational study that investigated effectiveness of PCSK9i after their introduction. RCTs previously conducted evaluated a composite primary end-point, therefore an estimation of the effect on the hospitalization with respect to the competing transition to death was lacking. We obtained evidence of effectiveness complementary to RTCs by combining the target trial causal inference framework within a community-based real-world EHR database.

[1] Hernán, M.A.; Robins, J.M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available: Table 1. Am. J. Epidemiol. 2016, 183, 758–764.
[2] Jackson, C. (2016). flexsurv: A Platform for Parametric Survival Modeling in R. Journal of Statistical Software, 70(8), 1–33. https://doi.org/10.18637/jss.v070.i08.

WP101 Spatial statistical multilevel modelling of optic nerve deformation in multiple disease groups of glaucoma

Coan L.*[1], Czanner S.[1], Williams B.[2], Willoughby C.[3], Kavitha S.[4], Maccormick I.[5], Vallabh N.[6], Czanner G.[1]
[1]*Liverpool John Moores University ~ Liverpool ~ United Kingdom,* [2]*Lancaster University ~ Lancaster ~ United Kingdom,* [3]*Ulster University ~ Belfast ~ United Kingdom,* [4]*Aravind Eye Hospital ~ Pondicherry ~ India,* [5]*University of Edinburgh ~ Edinburgh ~ United Kingdom,* [6]*University of Liverpool ~ Liverpool ~ United Kingdom*

Glaucoma is the leading cause of irreversible blindness worldwide. A key parameter for diagnosis and monitoring of the disease is the cup-to-disc ratio (CDR) which is obtained from the ratio between the boundaries of the optic disc and cup in a retinal fundus image. Previous works have focused on the one meridian of the vertical-cup-to-disc ratio, dichotomous grouping (healthy vs glaucoma), and do not account for data from two eyes [1]. We propose a framework for retinal fundus images considering the spatial profile of the optic nerve head (ONH) by using a 24CDR imaging profile vector, analysing such profiles with respect to four disease groups (Healthy, Disc Glaucoma, Visual Field Glaucoma, and Disc & Visual Field Glaucoma) and two eyes per patient. We used data from a clinical intervention trial (OHTS), involving 1165 patients and 1893 eyes. The retinal fundus image was automatically segmented into the optic cup and disc using a segmentation algorithm [2]. Then, the CDR was calculated at 15-degree intervals creating the 24CDR profile. MANOVA was applied to study the modality groups with respect to the 24CDR profiles. Difference between the disease groups was found on the 24CDRs imaging profiles ($p < 0.0001$). Post hoc analysis showed a significant difference between healthy vs disc glaucoma and healthy vs disc and visual field glaucoma; no difference between healthy vs visual field glaucoma was found. However, MANOVA assumes independence and constant variance. Multilevel modelling was then applied, adjusting for correlations between two eyes, spatial correlations in profiles, and for heteroscedasticity, while eyes are nested within patients. We detected an overall shape effect between healthy vs all other three groups including visual field glaucoma, (interaction term disease with direction, $p < 0.001$). There was a significant difference in the shape of ONH profiles with respect to the disease group, also we confirmed the significance of incorporating a nested structure of an eye within patient random effect ($p < 0.001$). Spatial multilevel modelling identifies where disease groups differ in the profile, which is novel. Currently, we are extending our framework toward automated discrimination of the four disease groups for unseen eyes.

[1] Coan LJ, Williams BM, Krishna Adithya V, Upadhyaya S, Alkafri A, Czanner S, Venkatesh R, Willoughby CE, Kavitha S, Czanner G. Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review,Survey of ophthalmology, 68, 2023, 17-41.
[2] Krishna Adithya, V.; Williams, B.M.; Czanner, S.; Kavitha, S.; Friedman, D.S.; Willoughby, C.E.; Venkatesh, R.; Czanner, G. Effunet-spagen: An efficient and spatial generative approach to glaucoma ion, Journal of Imaging, 7, 2021, 1-17.

WP102 Quantifying the risk of treatment-related adverse events in oncology: data, issues and models

Coz E.*, Fauvernier M., Maucort--Boulch D.
Université de Lyon, Hospices Civils de Lyon, Pôle Santé Publique, Service de Biostatistique et Bioinformatique, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Équipe Biostatistique-Santé ~ Lyon ~ France

In randomized controlled trials or in observational studies, comparing the occurrence of adverse events according to patients' characteristics (e.g. treatment) is a recurrent question in the analysis of drug safety data. However, due to the complexity of those data, the comparison of toxicity in patients is challenging and is often inappropriately conducted [1]. This study reports on statistical models used in the literature to quantify and compare the risks of treatment-related adverse events. We searched the literature for statistical models to quantify the risk of adverse events regarding some covariates of interest. Four dimensions were identified as relevant to describe treatment-related toxicities: timing, recurrence or multiplicity, severity and duration [2]. We found models that dealt with each of those dimensions, including several recent propositions that considered the most neglected dimensions: severity and duration. Many models also tackled the challenges of the potentially complex therapy pathways of patients that may highly affect the risk evaluation: censoring, competing events, treatment interruption. Across this review, we embraced a large range of absolute or relative risk indicators (e.g. hazards, hazard ratios, probability). Reporting both the absolute and its associate relative risk indicators has been recommended in guidelines like CONSORT's. In addition, more than one risk indicator may be relevant to describe the occurrence of adverse events accurately (e.g. probability and hazards in a competing risks setting) and was considered by some authors. We identified several models and risk indicators to quantify and compare toxicity in patients, and a growing interest for the issue over the last few years. Defining the outcome, which implies to select adverse events of interest among thousands of them, remains a delicate task despite some recent efforts to suggest crucial events to focus on.

[1] R. Phillips, O. Sauzet, V. Cornelius, Statistical methods for the analysis of adverse event data in randomised controlled trials: a scoping review and taxonomy, BMC Med. Res. Methodol. 20 (2020) 288.
[2] B. Cabarrou, C. Gomez-Roca, M. Viala, A. Rabeau, R. Paulon, D. Loirat, N. Munsch, J.-P. Delord, T. Filleron, Modernizing adverse events analysis in oncology clinical trials using alternative approaches: rationale and design of the MOTIVATE trial, Invest. New Drugs. 38 (2020) 1879–1887.

WP103 Balance of time-dependent covariates in real-world data via the fréchet distance

Diaz M.*
Homer Stryker M.D. School of Medicine, Western Michigan University ~ Kalamazoo ~ United States of America

Assessment of covariate balance is a key step in examining the adequacy of inverse probability weighting models to perform comparisons between groups from real-world data. We generally perform it on baseline confounders, but not necessarily on longitudinal ones. We could use pointwise standardized differences, standardized differences of slopes, or weights from the model for such purpose. However, pointwise differences could be cumbersome in the case of densely sampled longitudinal markers and/or measured at different points, or could fail in detecting relevant differences if coincidentally at the evaluation time points the group curves coincide or are very close. Slopes are suitable for linear models but not for more complex curves. Weights do not identify the specific covariate(s) responsible for imbalances. Here I propose the Fréchet distance[1] to assess balance of time-dependent covariates. A set of theoretical curves for which the standardized difference of their slopes was within 10% -an accepted balance level- as well as curves whose comparison was greater than such threshold identified the Fréchet distance equivalent to the 10% mark. Then I applied this threshold and the Fréchet distance to a set of real curves representing the monthly trajectory of hemoglobin A1c among diabetic patients for a period of up to nine years. The distance and the threshold clustered these curves. This clustering displayed particular patterns in the curves autoregressive and moving average components. The Fréchet distance differentiated between curves within as well as beyond the 10% difference threshold. This assessment of covariate balance provides the following advantages: can handle curves of different lengths, shape, and multidimensional nature, arbitrary time points, and considers the curves directionality. Additionally, Fréchet distance allows clustering individual-level curves to determine their heterogeneity within groups. Future work includes examining the utility of this measure with multidimensional curves as well as the systematic comparison of its performance with other approaches.

[1] Eiter T, Mannila H. Computing discrete Fréchet distance. Tech Report Vienna University of Technology, May 1994.

WP104 Big data and statistical inferences problems: is effect size measure an alternative to p-value?

Galotta A.*, Capra N., Marenzi G., Rurali E., Bonomi A.
IRCCS Centro Cardiologico Monzino ~ Milano ~ Italy
This work was partly financed by the Italian Ministry of Health and the Lombardia Region (Grant NET- 2016-02364191; EASY-NET)

The exclusive use of p-value to evaluate statistical inference could be risky and misleading, and this problem is increasingly common in the era of big data. The American Statistical Association (ASA) has published several statements regarding the proper use, interpretation and suggestions on alternatives to p-value. P-value is a direct function of sample size (N): large sample sizes lead to smaller p-values, and the limit of significance level can be reached against the null hypothesis. Effect size can be used combined with p-value to complete the information needed for proper inference. In fact, the effect size is independent of sample size. The aim is to apply different effect size measures as alternatives to p- value on an observational dataset (real world data). In this study, we used health administrative data from the Lombardy region; the records referred to 15954 patients aged 90 years and older with a hospitalization for AMI during 2003-2018. Patients were grouped according to whether or not they had been treated with PCI during the baseline hospitalization. Through the use of SAS Macros [1-2], we calculated the effect size measures for group comparisons: Cohen's D for continuous variables and Phi coefficient for categorical variables. For example, the difference between the two groups under study can be described as, "The mean age was significantly lower in the group undergoing PCI (91.9±2.0) than in the group without PCI (92.7±2.6) ($p<0.0001$, Cohen's d = 0.32705)"; meaning that the difference of the variable is very significant when looking at the p-value and the effect size indicates a small-to-medium effect. For categorical variables, even very significant differences (<0.0001) between the two groups compared had a very small effect size (Phi Coefficient<0.10, negligible association). In conclusion, the exclusive and arbitrary use of pvalue is increasingly criticized. Effect sizes, in combination with pvalue, broaden the concept of inference, making it easier for authors to think statistically. By including these estimates as well, the interpretation of statistical results in terms of statistical significance and effect size is completed.

[1] R Kadel, K. Kip. (2012). A SAS Macro to Compute Effect Size (Cohen's) and its Confidence Interval from Raw Survey Data.
[2] P. Rasmussen, et al. (2013). Paper 491-2013 DICHOTOMIZED_D: A SAS® Macro for Computing Effect Sizes for Artificially Dichotomized Variables. SAS Global Forum 2013

Poster Sessions

WP105 Projections of cardiovascular disease deaths using a microsimulation model in the absence of cohort data

Kalpourtzi N., Gountas I., Touloumi G.*
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens ~ Athens ~ Greece

Cardiovascular disease (CVD) is the leading cause of mortality in Greece. CVD future trends projections are useful to evaluate potential prevention strategies/measures, but the lack of cohort data in the country poses many statistical challenges. We aimed to project CVD deaths in Greece from 2015 to 2035 using a microsimulation model in the absence of cohort data. We developed a microsimulation model that simulates the life course of an open cohort of adults (18- 99 years) in Greece. In 2015, population was simulated based on the EMENO study data, a representative health examination survey of Greek adults with data on CVD risk factors including age, sex, smoking, blood pressure, lipids, diabetes, diet, BMI. To estimate transition probabilities of categorical variables, we adapted a mathematical optimization technique proposed within operational research, for estimating net transition probabilities [1]. For continuous variables, we calculated their percentile ranks based on age, sex, and BMI in the initial population and we assumed that individuals maintained their percentile rank for the entire simulation period[2]. With each year of increasing age, individuals' values changed to match their initial percentile ranks after stratifying by age, sex and BMI. For the trajectory of smoking, using data from the EMENO study on the age of smoking initiation, we conducted a time-to-event analysis censoring individuals who had not yet started smoking at their current age. Using a Cox proportional hazards model, we calculated the probabilities of smoking initiation. A similar procedure was followed to estimate smoking cessation and relapse probabilities. To estimate CVD mortality risk, we used the Framingham score for CVD deaths. Our results indicated that, under status-quo (i.e., without any intervention), the burden of CVD deaths in Greece is projected to worsen over the next two decades; the projected number of CVD deaths increased by 24.3%, between 2015 and 2035. Our findings underscore the need for targeted interventions to mitigate the rising burden of CVD in Greece. Further methodological evaluation is needed to improve the accuracy and validity of future projections in the absence of cohort data, as a combination of several approaches under various assumptions is required.

[1] Kassteele Jv, Hoogenveen RT, Engelfriet PM, Baal PH, Boshuizen HC. Estimating net transition probabilities from cross-sectional data with application to risk factors in chronic disease modeling. Stat Med. 2012 Mar 15;31(6):533-43. doi: 10.1002/sim.4423. Epub 2011 Dec 5. PMID: 22139860.
[2] Kypridemos, C., Allen, K., Hickey, G. L., Guzman-Castillo, M., Bandosz, P., Buchan, I., ... & O'Flaherty, M. (2016). Cardiovascular screening to reduce the burden from cardiovascular disease: microsimulation study to quantify policy options. bmj, 353.

WP106 A new statistical index for evaluating variability in physical state index during pediatric anesthesia

Muhammad Khan N.*[1], Maria Bonardi C.[2], Amigoni A.[2], Gregori D.[1]
[1]*University of Padova ~ Padova ~ Italy,* [2]*University Hospital of Padova ~ Padova ~ Italy*

Monitoring physical state index (PSI) [1] of pediatric patients after administrating anesthesia is important because it can affect health outcomes. It is also necessary to assess the variation of PSI for allowing the proper medical management. The anesthesia period is clinically divided into well-defined five phases; however, recent studies have unexpectedly detected large variations of PSI have even within same phase during pediatric anesthesia. Therefore, a robust method is required to evaluate the phases and take necessary actions immediately. In this research, we developed a statistical tool, the Variability Ratio Index (VARI), based on the weighted measure of the deviation of PSI from its stationary process to evaluate these phases. We incorporated VARI to single-center retrospective study conducted on pediatric patients who underwent to cardiac surgery in extra corporeal circulation with continuous monitor of PSI during general anesthesia. The study included data on 20 children observed at 124699 time points. We observed large variation of PSI within each phase and VARI evaluated all phases with satisfactory results, specifically distinguished freezing and awaking phases from others. The distribution of PSI was non normal with many outliers. There we performed robustness checking of VARI before applying it with the data. We performed 10,000 iterative convergences to identify the distribution of VARI and found that it follows beta distribution. We performed parametric bootstrapping in Bayesian paradigm to check the robustness and to estimate the parameters of the beta distribution with confidence interval. We also did the Monte Carlo simulation to check its robustness. VARI showed robust behavior in both parametric bootstrapping and Monte Carlo simulation. Finally, we developed R package "varifinder" for its use. It is important to closely observe the variation of PSI and evaluate the phases of pediatric anesthesia to promote a healthier nation. The robustness property of VARI ensures its general applicability within clinical data.

[1] Schneider, Gerhard, Susanne Heglmeier, Jürgen Schneider, Gunter Tempel, and Eberhard F. Kochs, Intensive care medicine, 30, 2004, 213-216.

WP107 Quality of life in patients after 1-year percutaneous coronary intervention: real-world data analysis

Siriyotha S.*[1], Limpijankit T.[2], Sansanayudh N.[3], Pattanaprateep O.[1], Thakkinstian A.[1]
[1]*Department of Clinical Epidemiology and Biostatistics,Faculty of Medicine Ramathibodi Hospital, Mahidol University ~ Bangkok ~ Thailand,* [2]*Department of Medicine (Cardiology), Faculty of Medicine Ramathibodi Hospital, Mahidol University ~ Bangkok ~ Thailand,* [3]*Cardiology, Phramongkutklao Hospital ~ Bangkok ~ Thailand*

Clinical efficacy of percutaneous coronary intervention (PCI) for treatment of occlusion in coronary artery disease (CAD) is well documented, but improvement of health related quality of life (HRQoL) after procedure is unclear. This study was conducted using data from the nationwide PCI registry including 39 hospitals across Thailand during May 2018 to August 2019. Baseline data including underlying diseases, CAD presentation, medication use, and procedural details were retrieved. All patients were assessed HRQoL using Thai EQ5D-5L at baseline, after discharge from performing PCI, 6- and 12-month follow up. Factors associated with HRQoL were assessed using a mixed-effect linear regression. A total of 22,741 patients were included with a mean age of 64.2±11.7 years and predominantly male (69%). Common comorbidities were hypertension (67.4%), dyslipidemia (65.4%), diabetes mellitus (44.2%), chronic kidney disease (32.5%), and current smoking (23.2%). Most patients presented with acute coronary syndrome (57.9%) and had triple vessel and/or left main disease (44.9%). After adjusting covariates, mean HRQoL was 73.3 at baseline, which increased to 85.8 at discharge from hospital, and persisted to 86.7 and 86.1at 6- and 12-month follow up improving about 12 units after PCI procedure. Factors associated with improving HRQoL were age, male gender, overweight, CAD presentation, diabetes mellitus, dyslipidemia, cerebrovascular and chronic kidney disease, health insurance, cardiogenic shock, radial access, and procedural success. Our finding suggested that PCI procedure could improve HRQoL. Factors mostly associated with its improvements are cardiogenic shock at presentation, age, chronic kidney disease, and CAD presentation with ST-segment elevation myocardial infarction.

1. Nowbar, Alexandra N, Gitto M, Howard, James P, Francis, Darrel P, Al-Lamee R. Mortality From Ischemic Heart Disease. Analysis of Data From the World Health Organization and Coronary Artery Disease Risk Factors From NCD Risk Factor Collaboration. Circ Cardiovasc Qual Outcomes. 2019;12(June):1–11.
2. Severino P, D'Amato A, Pucci M, Infusino F, Adamo F, Birtolo LI, et al. Ischemic Heart Disease Pathophysiology Paradigms Overview: From Plaque Activation to Microvascular Dysfunction. Int J Mol Sci. 2020 Oct 30;21(21):8118.
3. Herbert T, Rizzolo D. The role of percutaneous coronary intervention in managing patients with stable ischemic heart disease. J Am Acad Physician Assist. 2020 Jun;33(6):18–22.
4. Verheugt FWA. The role of the percutaneous coronary intervention in acute coronary syndrome. 2010.

WP108 "Doing the doctors work" - usefulness of a simple prediction model applied to health register data in sweden

Wagner P.*, Gunnar H.
Lund University ~ Lund ~ Sweden

Cerebral palsy (CP) is a neurological condition of different subtypes that affect muscle coordination. It is diagnosed in early childhood by a pediatric neurologists (PN) based on evaluation of medical history, physical examination, as well as other tests. When a child is diagnosed, the doctor develops a treatment plan to manage symptoms and improve quality of life. Unfortunately, there is a current shortage of PNs in Sweden. We apply a simple prediction model to data from the Swedish national CP follow-up program, CPUP, with the purpose of diagnosing children with CP. In CPUP, children are examined annually. Measurement include; range of motion (ROM) of upper and lower limbs, muscle spasticity, pain, gross motor classification, CP diagnosis etc. (www.cpup.se). The program comprises > 95% of all children in Sweden with CP. In total, we extracted approximately 800 longitudinally measured variables on 4800 children from the register database. Measurement intervals varied between individuals. Expert domain knowledge was used to identify variables for which developmental trajectories were important for determining the correct CP subtype. For these, individual smoothed trends were estimated using mixed effect models coupled with empirical bays estimators. Trends were projected on a grid to generate variables approximating regularly spaced longitudinal measurements. For remaining variables, the last non-missing measurement was used. We fit a simple L1 penalized multinomial logistic regression model to these data for generating predictions. Cross-validation was used to select the optimal penalty parameter. Results: The initial 1500 variables were reduced to an optimal predictor set of 145. Accuracy was 0.78 and sensitivity and specificity for individual CP subtypes ranged from (0.90,0.87), (0.79, 0.85), (0.70, 0.99) to (0.67,0.95) and (0.22, 1.00) for non-classifiable CP. Cohen's kappa between model predictions and physician diagnoses was 0.68. This simple regression model may predict CP subtypes almost as well as PNs. Kappas between senior PN experts are reported to be 0.74 – 0.78 (1) in a controlled study setting. In a total population, is in the present study, this figure may be expected to drop. Studies are under way to compare predictions to PN diagnoses in order to validate this claim.

(1) Gainsborough M, Surman G, Maestri G, Colver A, Cans C. Validity and reliability of the guidelines of the surveillance of cerebral palsy in Europe for the classification of cerebral palsy. Dev Med Child Neurol. 2008 Nov;50(11):828-31. doi: 10.1111/j.1469-8749.2008.03141.x. PMID: 19058397.

Poster Sessions

WP109 Have air pollution levels decreased more in high income areas – a 19-year follow-up in gothenburg, sweden

Andersson E.*, Azzouz M., Ögren M., Molnár P., Stockfelt L.
University of Gothenburg ~ Gothenburg ~ Sweden

Ambient air pollution (for example from traffic exhaust, tire wear, industrial emissions, residential heating) has been linked to e.g. cardiovascular diseases (BMJ 2014, WHO 2016), also at the comparatively low levels in Sweden (Ljungman 2019). Fine particles (with aerodynamic diameter ≤ 2.5 µm) are considered to be responsible for the majority of the health effects. A meta-analysis showed the risk of myocardial infarction to increase by 18 % per 10 µg/m3 of PM2.5 (Front Med 2021). As a proxy for personal long-term exposure, air pollution levels at the home are often used. Several reports show inequity in that people with lower socioeconomic status have higher environmental exposures. The aim was to study the individual PM2.5 exposure trajectories (years 2000-2018) among people living in areas with different socioeconomic status in Gothenburg, Sweden. The Swedish CArdioPulmonary bioImage Study, SCAPIS, is a population-based cohort where ~30 000 men and women (50-64 years old) were recruited 2013-2018 (www.scapis.org). Approximately 6000 were recruited in the Gothenburg area. All participants were investigated using novel biomarkers and imaging. Individual yearly addresses were used to model exposure to air pollution 2000-2018, and to assess the area-level socioeconomic status, based on DeSO data from Statistics Sweden (Demografiska Statistikområden, www.scb.se). Associations between air pollution exposure and area-level socioeconomic status were assessed using both two-stage analysis with derived variables and mixed effect models. Approximately 2500 individuals had lived in the same DeSO area during the entire period 2000-2018. Preliminary results, using two-stage analysis of the association between individual PM2.5 trajectories and area income, showed that exposure in the highest income areas had decreased more, 5.6% per year compared to a decrease of 5.0% in middle income areas. The mixed effect model confirmed this, and indicated that the initial levels (in 2000) were ~10 % lower in the high income areas. Two approaches for analysis of longitudinal data were used, where the two-stage approach is simple to use, and a mixed model can e.g apply different models for the correlation between observations. Preliminary results were similar, indicating lower air pollution levels (PM2.5) in areas with high socioeconomic status.

Bergström et al (2015) The Swedish CArdioPulmonary BioImage Study: objectives and design. J Internal medicine, 278, 645-659 doi: 10.1111/joim.12384.
Cesaroni et al (2014) Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. BMJ, 348:f7412 doi: 10.1136/bmj.f7412.
Diggle et al (2002) Analysis of longitudinal data. Oxford Statistical Science, Series 25, Second edition, Oxford University Press.
Ljungman et al (2019) Long-Term Exposure to Particulate Air Pollution, Black Carbon, and Their Source Components in Relation to Ischemic Heart Disease and Stroke. Environmental Health Perspectives, 127(10) 107012 doi: 10.1289/EHP4757.
WHO (2016) Ambient air pollution: A global assessment of exposure and burden of disease. World Health Organization, www.who.int.
Zou et al (2021) Long-term exposure to ambient air pollution and myocardial infarction: a systematic review and meta-analysis. Front. Med. 8, 616355 doi: 10.3389/fmed.2021.616355.

WP110 Comparison of clinic versus home devices for spirometry assessment

Diaz--Coto S.*, Peacock J., Egner J., Karagas M.
Geisel School of Medicine at Dartmouth, Dartmouth College ~ Hanover ~ United States of America

Spirometry is a pulmonary function test used for monitoring lung health, particularly when the individuals have an underlying respiratory condition such as asthma [1]. Lung function measures are used clinically in individuals to guide treatment decisions and also are used in populations to model risk factors for disease. Spirometry provides a range of measures from one blow that describe the flow rate of exhalation and the volume of the lungs. The greatest challenge is that spirometry requires sustained respiratory effort and training protocols are required for the test administrator to ensure that maximum effort is used. Spirometers used in healthcare clinic settings are quite large machines that need regular calibration and are not easily portable. On the other hand, there are a range of devices that can be used away from clinic settings. In the New Hampshire Birth Cohort Study, we have compared measures made by both a clinic non-portable spirometer and a portable machine in the same children to explore the agreement between the pairs of measurements. The forced vital capacity (FVC), the forced expiratory volume in the first second (FEV1) and the ratio FEV1/FVC were selected for studying the reliability of the spirometry derived measures. We explored through simulations under a wide variety of scenarios [2] how even good agreement in FVC and FEV1 does not imply same results for FEV1/FVC. Good agreements in some of the measures taken by two spirometers on the same children do not necessarily imply a good agreement in the transformations of these variables. This may lead to a miss consideration of the reliability of those transformations and even the interchangeability of the devices. We explored conditions under which both reliability and interchangeability can be supported.

[1] Kana Ram Jat. Primary Care Respiratory Journal. 2013 Jun; 22(2): 221–229.
[2] Berchtold, A. (2016). Test–retest: Agreement or reliability? Methodological Innovations, 9. https://doi.org/10.1177/2059799116672875

WP111 Systematic mediation analysis of genetic loci associated with kidney function in a population-based study

Ghasemi--Semeskandeh D.*[1], Emmert D.[1], König E.[1], Foco L.[1], Gögele M.[1], Barin L.[1], Fujii R.[1], Fuchsberger C.[1], J M Peters D.[2], P Pramstaller P.[1], Pattaro C.[1]
[1]Institute for Biomedicine, Eurac Research ~ Bolzano ~ Italy, [2]Department of Human Genetics, Leiden University Medical Center ~ Leiden ~ Netherlands

Genome-wide association studies (GWAS) identified hundreds of genetic loci associated with kidney function markers. However, most of the underlying biological mechanisms have not been elucidated. Systematic mediation analysis across multiple traits may identify potential pathways. In the Cooperative Health Research in South Tyrol (CHRIS) study (n=10,146) [1], genetic associations with serum creatinine-based estimated glomerular filtration rate (eGFR), the main kidney function marker, were submitted to mediation analysis across 70 quantitative biochemical, anthropometric and blood pressure traits, following a multi-step approach. First, to identify genetic variants relevant to eGFR in CHRIS, we tested for replication 147 genetic loci previously associated with kidney function in a large GWAS (n=567,460; 1-sided α=0.00034) [2], using sex- and age-adjusted linear mixed models on ln(eGFR), accounting for relatedness. Second, we assessed the differences between the estimated association coefficient before (b0) and after (b1) adjustment for each of the 70 traits, in turn. Traits were classified as potential mediators if d=b0−b1 felt outside [10%, 90%] percentile of its expected null distribution. Third, we tested potential mediators for association with the same variant in CHRIS. Results were compared against trait-genetic variant associations from large consortia. Eleven loci, totaling 163 variants, were associated with ln(eGFR) in CHRIS. Replicated variants had 1- to-5.4 times larger effects than in the original, larger GWAS, at similar minor allele frequencies. We identified 4 to 12 traits as potential mediators at each variant. When further testing the association between variants and potential mediators, only 3 traits were eventually classified as mediators. Specifically, serum magnesium, serum urate, and the activated partial thromboplastin time (aPTT) resulted being partial mediators of the eGFR associations at SHROOM3, IGF1R, and SLC34A1 loci, respectively. At SLC34A1, aPTT adjustment corresponded to a 21% larger effect of the variant on ln(eGFR); at SHROOM3, serum magnesium adjustment caused ~11% effect attenuation; at IGF1R, serum urate adjustment represented up-to-18% attenuation of the effect of variants on ln(eGFR). These variant-trait associations were confirmed in data from larger consortia, validating robustness of the approach. Systematic mediation analysis in single population-based studies can identify mediators of the association between genetic variants and kidney function markers.

[1] Wuttke, M., et al., A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet, 2019. 51(6): p. 957-972.
[2] Pattaro, C., et al., The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. Journal of Translational Medicine, 2015. 13(1): p. 348.

Poster Sessions

WP112 Using autoregression to model novel viral outbreaks in human populations to manage medical resources.

Grant T.*, Crispino G.
StatisticaMedica ~ Dublin ~ Irelan

A simplified modelling approach for when new viruses enter the human population for a variety of transmission and recovery rates is considered. From time-to-time new viruses entire the human population such as avian flu, swine flu, and most recently SARS-CoV-19. Traditional models for the spread of an existing virus are based on compartmental models such as the SEIR models and their variants [1], which tend to have multiple parameters that are difficult to estimate for a new virus and require assumptions that are difficult to justify. The use of these complicated models can in many cases be eliminated by recognizing in the early phases of a new virus, the transition from the exposed to infected/infectious groups and subsequently to the removed groups are negligible relative to the transition from susceptible to infected. As such, the complex system of equations reduces to a model that is essentially an autoregressive model with a nonstationary constant [2]. Modelling this constant provides valuable estimates for traits central to the management of resources such as when will the peak demand occurs, without requiring assumptions about extra parameters. This work generates the deterministic counts for 3 R0 values (3, 6, and 9) and 3 values for the transition from the infected group to the removed group in the SIR model to produce outbreaks that peak between 50 to 275 days. These datasets are then analysed using the autoregressive framework in the context of a first outbreak of a new virus or contagious disease with the primary endpoint being the day of peak demand to demonstrate robustness across these parameter states. The estimations demonstrate the model predicts the peak within 4 days for time periods where pressure decisions are required. The models work well in predicting the time of peak needs of medical resources without requiring the assumption of model constants. In an actual outbreak, the model easily adjusts the prediction as time passes and more days of data is available.

[1] L. Lopez, X. Rodo, Results in Physics, 21, 2021, 103746-103761.
[2] R.H. Shumway, D.S. Stoffer, Time series analysis and its applications, Springer, 2000.

WP113 Profile likelihood confidence interval for the prevalence assessed by an imperfect diagnostic test

Hársfalvi P.*[1], Reiczigel J.[2]
[1]BiTrial Clinical Research ~ Budapest ~ Hungary, [2]University of Veterinary Medicine ~ Budapest ~ Hungary

Estimating a confidence interval (CI) for the true population prevalence is an essential task for epidemiological studies. As most diagnostic tests used for the estimation are prone to some misclassification, methods capable of adjusting the estimate based on the diagnostic sensitivity and specificity are also available. Lang and Reiczigel [1] and Flor et al. [2] proposed methods for constructing CIs for prevalence considering the diagnostic parameters as random variables instead of known values, advantageous especially when the samples for estimating sensitivity and specificity are small, being usually the case for the validation studies of screening tests. We propose a profile likelihood confidence interval (PLCI) for the prevalence of a disease if sensitivity and specificity of the diagnostic test are estimated from independent validation samples. We illustrate its attributes by presenting some practical examples for application and compare its performance with the CIs available for the same purpose. The proposed new CI has coverage probabilities comparable to those of the Lang-Reiczigel CI while its expected length is shorter, except when true prevalence is near 0.5. The Flor interval is comparable to the PLCI in terms of length but has much lower coverage probabilities.

[1] Z. Lang and J. Reiczigel, "Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity," Preventive Veterinary Medicine, vol. 113, pp. 13–22, 12 2014, doi: 10.1016/j.prevetmed.2013.09.015.
[2] M. Flor, M. Weiß, T. Selhorst, C. Müller-Graf, and M. Greiner, "Comparison of Bayesian and frequentist methods for prevalence estimation under misclassification," BMC Public Health, vol. 20, no. 1, pp. 1–10, 2020.

WP114 Sensitivity analysis for multiple exposure effects in generalized linear models with unmeasured confounders

Jeong B.[1], Lee D.*[1], Lee W.[2]
[1]Department of Statistics, Ewha Womans University ~ Seoul ~ Korea, Republic of, [2]Department of Public Health Science, Graduate School of Public Health, Seoul National University ~ Seoul ~ Korea, Republic of

In epidemiological studies, investigating the effect of multiple exposures on health outcomes of interest is an important topic. To deal with different types of target outcomes, generalized linear models and machine learning algorithms are widely used for analyzing multiple exposures. However, observational studies are prone to unmeasured confounding, so the estimates of multiple exposures effect can be substantially biased [1]. Sensitivity analysis is desirable to assess how sensitive the estimates of multiple exposure effects are to unmeasured confounding. In this study, we propose a sensitivity analysis to efficiently handle the multiple exposures and different types of non-normal outcomes. We develop a sensitivity model when the outcomes are non-normal or time-to-event variables. By solving the linear programming problem with quadratic and linear constraints, we provide the proper ranges of estimates of the single-exposure and joint-exposure effects on the non-normal outcome. Our analysis results are easily interpretable with visualized sensitivity intervals. Through extensive numerical studies for generalized linear models and the Cox proportional hazards model with different assumptions, we illustrate the usefulness of the proposed method. The proposed sensitivity analysis allows for statistical models with non-normal outcome variables and multiple exposures and is indispensable in the presence of unmeasured confounding. With domain knowledge, it can be utilized in various epidemiological areas.

[1] Weisskopf, Marc G., Ryan M. Seals, and Thomas F. Webster. "Bias amplification in epidemiologic analysis of exposure to mixtures." Environmental health perspectives 126.4 (2018): 047003.

WP115 Analytic error in national-level prevalence surveys of healthcare- associated infections: a systematic review

Pastaka A.[1], Kritsotakis E.*[2]
[1]Master of Biostatistics Program, Medical School and Department of Mathematics, National and Kapodistrian University of Athens ~ Athens ~ Greece, [2]Laboratory of Biostatistics, School of Medicine ~ Heraklion ~ Greece

INTRODUCTION & OBJECTIVES: Healthcare-associated infections (HAIs) constitute a major healthcare concern. Their burden is determined with national-level cross-sectional studies that estimate HAI prevalence in patient populations across a country, frequently using complex sampling designs. Complex design features, such as clustering and stratification, improve data collection and estimation efficiency, but require researchers to account for these features in statistical analysis. Moreover, inferences for subpopulations (e.g. elderly patients, those with pneumonia or certain pathogens) are often desired, which requires care as subpopulation sample size is random. However, healthcare researchers may frequently lack training in design-based or model-based methods of analysing complex survey data. We investigated the extent of analytic error in complex sample surveys of HAIs performed at national level.
METHODS & RESULTS: PubMed was searched for studies published between 2000 and 2022 reporting country-level estimates of HAI prevalence in routine care settings of hospitals, long-term care facilities, nursing homes, or health centres. In all, 59 studies involving 4,005,104 patients in 5,302 healthcare facilities in 34 countries were analysed. Most studies (n=52; 88%) reported point (1 day) prevalence estimates for any type of infection (n=55; 93%) in facility-wide settings (n=41; 70%) or intensive-care units (n=13; 22%). All but one studies were complex designs, involving either clustering (n=47; 80%) or clustering and stratification (n=11; 19%). Primary sampling unit was the entire healthcare facility in 53 studies (90%) and the ward in 5 studies (8%). One study drew a simple random sample of patients. Sample size calculations were reported in only 7 (12%) studies, 4 of which accounted for design effects. In statistical analysis, only 5/58 (8%) studies fully accounted for the complex sampling features (all design-based approaches), 48/58 (83%) ignored them, and 5/58 (9%) were unclear. Subpopulation inferences were conducted in 8 studies, of which only 2 (25%) used appropriate estimation methods, whereas 4 (50%) studies were unclear.
CONCLUSIONS: Considering the bias in point estimates and standard errors that occurs when design effects of complex data are not incorporated or sample weights are omitted from analysis, this investigation calls for improvement in the statistics and reporting of complex sampling surveys in healthcare epidemiology.

WP116 Clinically versus cytokine-defined genital inflammation and the risk of hiv infection

Osman F.*[3], Yende--Zuma N.[1], Niivashnee N.[3], Liebenberg L.[3], Passmore J.[2], Abdool Karim Q.[3], Abdool Karim S.S.[3], Mckinnon L.R.[4]
[1]SAMRC ~ Durban ~ South Africa, [2]University of Cape town ~ Cape Town ~ South Africa, [3]Centre For The Aids Programme Of Research In South Africa ~ Durban ~ South Africa, [4]University of Manitoba ~ Winnipeg ~ Canada

Studies indicate that elevated female genital tract inflammation increases the risk of HIV acquisition, but how inflammation is optimally defined both clinically and immunologically is still debated. The objective was to determine whether women who experienced clinical inflammation, defined by pelvic abnormalities or female genital tract (FGT) related adverse events (AEs), also had subclinical genital inflammation, or elevated cytokines, and to compare how these definitions predicted an increased risk of HIV acquisition. We studied 889 HIV-uninfected women from the CAPRISA 004 trial who were randomized to receive either tenofovir gel or hydroxyethylcellulose placebo and were followed between May 2007 and March 2010. Cox proportional hazard regression models accounting for repeated measures were used to evaluate the predictors of time to HIV infection. Time varying exposures included abnormal pelvic abnormalities and self-reported or clinically observed female genital tract (FGT) adverse event (AE). Linear mixed models were fitted to assess the effect of the frequency of pelvic abnormalities and FGT- related AEs on genital cytokine concentrations. FGT-related AEs were reported by 71.3% (634/889) of women, whilst 76.5% (680/889) of women had ever experienced an abnormal pelvic exam. In multivariable analysis, the number of FG-related AEs or pelvic abnormalities reported or observed, were strongly associated with an increase in pro inflammatory cytokines (IL_1α, IL_1β, IL_12P40, IL_18), chemokines (MIG, IL_8 CTACK, VEGF), growth factors (M_CSF, BASIC_FGF, LIF, HGF, G_CSF, IL_3, PDGF_ ββ, SCF), and adaptive cytokines (IL_17α and IL_2Rα). Also, a significant relationship between HIV infection and clinical inflammation was observed (aHR 1.76, 95% CI: 1.03-3.02, p=0.040). In the multivariable model, we further included an adjustment for genital inflammation defined by elevated levels of ≥5 cytokines and observed an increased risk of HIV infection (aHR 1.72, 95% CI: 1.01-2.96, p=0.047). Clinical and immunological definitions of inflammation may not always overlap; however, evidence suggest that both types of definition are important predictors to HIV risk in women. Combining these definitions could enhance the accuracy of predicting HIV acquisition. These findings have important implications for HIV prevention efforts and highlight the need for a holistic approach to assessing and addressing genital inflammation in women at risk for HIV.

[1] P.D. Allison, Survival Analysis Using SAS: A Practical Guide, SAS Institute, 1995, 138-157.
[2] K. Mlisana, N. Naicker, L. Werner, L. Roberts, F. van Loggerenberg, C. Baxter, J.A Passmore, A.C. Grobler, A.W. Sturm, C. Willaimson, K. Ronarcher, G. Walzl, S.S Abdool Karim, Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high- risk women in South Africa, J Infect Dis, 206(1), 2012, 6-14.

Poster Sessions

WP117 Diurnal eating patterns and their effect on body mass index in the italian population (inran-scai 2005-2006)

Lopez Sanchez L.[1], Palla L.*[2]
[1]Universitat Pompeu Fabra ~ Barcelona ~ Spain, [2]Universita' di Roma La Sapienza ~ Roma ~ Italy

Late food intakes have been linked to weight gain while early meals have been associated with weight loss and maintenance. However, the impact of temporal (diurnal) eating patterns (DEPs) including information on the time of food consumption throughout the day and the irregularity across surveyed days has not been studied in the Italian population. INRAI-SCAI is a cross-sectional nutrition survey conducted in 2005-2006 in a representative sample of the Italian population , collecting diet diaries over 3 days, including a questionnaire with demographic and anthropometric variables (BMI) . We derived the DEPs by Principal Component Analysis jointly on indices of average energy intake and irregularity of energy intake in the (a) 24 hour intervals and alternatively across (b) the reduced 6 time intervals corresponding to meal and in-between meal times. Using the covariance matrix, the first 5 DEPs explained 93% of the total variance, with the first DEP (47% variance) score increasing with energy intake at main meals. A mixed-effect model was applied including only adults, with BMI as outcome, accounting for the correlation of subjects within household (ICC=0.195, p<0.0001) and the main DEPs as exposures and sex, age, geographic area, average daily energy intake as confounders. This resulted in a positive association of BMI with the first DEP (b=0.75 per 100%score, p=0.009). A positive significant association also resulted between BMI and the third DEP (10% variance) whose score increased with energy intake at snack times outside main meals (b=0.89 per 100%score, p=0.013) and with the fifth DEP (6.4% variance) which mainly captured food intake at night and irregularity of intake at night (b=0.34 per 100%score, p=0.028). Despite the limitations of this cross-sectional analysis including possibility of reverse causality and the potential biases typical of self-reporting, the study indicates that at the time of the survey, in the Italian adult population BMI tended to increased not only with large energy intake at main meals and at snack times but also with energy intake and irregularity of intake at night, indicating the importance of accounting for and modelling diurnal eating patterns, beside the average daily intake, for obesity and weight management.
Yoshida J, Eguchi E, Nagaoka K, Ito T, Ogino K. Association of night eating habits with metabolic syndrome and its components: a longitudinal study. BMC Public Health. (2018) 18:1366. 10.1186/s12889-018-6262-3

Sette S, Le Donne C, Piccinelli R, Arcella D, Turrini A, Leclercq C; INRAN-SCAI 2005-6 Study Group. The third Italian National Food Consumption Survey, INRAN-SCAI 2005-06--part 1: nutrient intakes in Italy. Nutr Metab Cardiovasc Dis. 2011 Dec;21(12):922-32. doi: 10.1016/j.numecd.2010.03.001. Epub 2010 Jul 31. PMID: 20674305.
Palla L, Almoosawi S. Diurnal Patterns of Energy Intake Derived via Principal Component Analysis and Their Relationship with Adiposity Measures in Adolescents: Results from the National Diet and Nutrition Survey RP (2008-2012). Nutrients. 2019 Feb 17;11(2):422. doi: 10.3390/nu11020422. PMID: 30781551; PMCID: PMC6412640.

WP118 Logistic regression with covariate-dependent probability of misclassification

Reiczigel J.*[1], Klaschka J.[2], Hársfalvi P.[3]
[1]Department of Biostatistics, University of Veterinary Medicine Budapest ~ Budapest ~ Hungary, [2]Institute of Computer Science of the Czech Academy of Sciences ~ Prague ~ Czech Republic, [3]BiTrial Clinical Research ~ Budapest ~ Hungary

If dependence of a binary outcome on some covariates is modeled, misclassification of the outcome may affect the estimates. If misclassification rates are assumed to be constant, the model by Liu and Zhang (2017) allows joint estimation of the dependence on covariates and the misclassification rates [1]. In some cases, however, misclassification rates are not constant, but they also depend on some covariates. No model has been published for this situation yet. We propose a logistic model that can be applied if presence of a feature (seropositivity, disease, etc.) depends on some covariates and its detection probability (given it is present) depends on other covariates. The model enables estimating the true prevalence of the studied feature conditional on the covariates as well as the sensitivity (probability of detection) conditional on its covariates. Furthermore, likelihood ratio tests can be applied to test the following null hypotheses:
- sensitivity is constant, that is, does not depend on the given covariates,
- sensitivity is 1, that is, there are no false negatives,
the probability of outcome does not depend on the given covariates. Our model may also have applications in social sciences. To some sensitive survey questions, respondents are reluctant to answer honestly, and the degree of honesty may depend on certain covariates. In such cases the model makes it possible to estimate simultaneously the true proportion conditional on the covariates and the degree of response distortion conditional on its covariates. Preliminary simulations show that the model meets the above expectations. We hope that further simulation and application to real data will confirm the advantages of the model.

[1] H. Liu, Z. Zhang, Behaviormetrika, 44, 2017, 447-476.

Poster Sessions

WP119 Comparison of propensity score methods with multiple treatments: simulations and clinical application

Bernasconi D.*, Capitoli G., Galimberti S., Valsecchi M.G.
Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milano- Bicocca, Monza, Italy ~ Monza ~ Italy

Surgical resection is generally the first-line therapeutic option for patients with hepatocellular carcinoma (HCC). In this context, it is of interest to investigate which type of surgical technique is more effective to treat patients with a large tumor size (>5 cm): anatomic (according to liver segments), semi-anatomic or wedge (non-anatomic). We analysed data from a multicenter observational study where the performed surgical technique mainly depended on surgeon's choice. However, the distribution of several important prognostic factors was not well balanced between treatment groups. We aimed to estimate the marginal effect of the type of surgery on recurrence-free survival using propensity score -based methods. However, despite there is a huge literature on propensity score approached in the case of binary treatments, few papers focus on situations with multiple treatments and especially for survival outcomes. The following methods were considered for the estimation of the treatment effect expressed as a hazard ratio: Cox model adjusted for confounders (conditional treatment effect), propensity score matching for each pairwise treatment comparison, inverse probability weighing and matching based on the generalised propensity score cumulative distribution function (marginal treatment effect). In the three last methods, propensity score was estimated using multinomial logistic regression or generalized boosted model and the marginal treatment effect was obtained from a Cox model fitted on the weighted or matched samples and with treatment as the only covariate. Moreover, a Monte Carlo simulation study was set to compare the performance of the methods under different scenarios according to the number and the distribution of confounders, the event time and censoring distribution and the true marginal hazard ratios. Using propensity score based methods we found no evident advantages of a resection technique over the others in term of recurrence-free survival. According to the results from the simulation study, inverse probability weighting with propensity score estimated via multinomial logistic regression produced unbiased and more efficient estimates of the marginal treatment effect compared to the other methods considered.

[1] D. W. Brown, S. M. DeSantis, T. J. Greene, V. Maroufy, A. Yaseen, H. Wu, G. Williams, and M. D. Swartz. A novel approach for propensity score matching and stratification for multiple treatments: Application to an electronic health record–derived study, Statistics in medicine, 2020, 39(17):2308– 2323.
[2] D.F. McCaffrey, B.A. Griffin, D. Almirall, M.E. Slaughter, R. Ramchand, and L.F. Burgette. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Statistics in medicine, 2013, 32:3388-3414.

WP120 30 Day mortality in hospitalised covid-19 patients: a retrospective study for the province of bolzano (italy)

Bonetti M.*[1], Brigiari G.[2], Gregori D.[2], Melani C.[1]
[1]Observatory for Health, Health Department ~ Province of Bolzano ~ Italy, [2]Unit of Biostatistics, Epidemiology and Public Health, DCTVPH ~ Padova ~ Italy

Due to the rise in cases of COVID-19, even the small province of Bolzano, as in the other Italian regions, has to face the strong pressure at the hospital levels from February 2020. Data show differences during pandemic waves in terms of clinical outcomes. The aim is to analyse how the different waves have impacted on the 30-day mortality for hospitalised patients. Data from hospitalised patients were extracted from the Hospital Discharge Administrative Databases; the analysis included admissions to local hospitals for residents' patients with COVID-19 diagnosis from February 2020 until September 2022. Based on incidence rates, it has defined eight pandemic waves; sex, age, length of stay (LOS), admission to intensive and subintensive care, and vaccination have been collected and classified as time dependent for the Cox proportional hazards model, while 30-day mortality after first admissions as endpoint. The Cox model is able to estimate how the effects of the covariates change over time. The cohort includes 7,881 patients divided into different pandemic waves; 1,201 (15,2%) died within 30 days with a median age of 84 years. The median LOS decreased as the waves passed, from 13 days in the first waves to 7 in the last. The results of the Cox model show that women have a lower risk of death (HR = 0,729; IC95%: 0,647-0,821); as the age increases, the risk of death increases (HR greater than 1 in all waves), while the risk decreases associated with the reduction in LOS (HR less than 1 in all waves). Further considerations on the role of vaccines and intensive care are in progress. The study supports the hypothesis that the male sex and the increase in age represent a risk factor for mortality independently of pandemic waves. At the same time, data show that the impact of COVID-19 during the last periods seems to be less aggressive with respect to the first waves.

[1] W. Ageno, C. Cogliati, M. Perego M et al., Internal and Emergency Medicine, 16, 2021, 989–996.
[2] M. Buyse, P. Piedbois, Statistics in Medicine, 15, 1996, 2797-2812.

WP121 Importance of sex-stratification in biomarker identification of early-stage melanoma survival analyses

Chrysanthou E.*[2], Sehovic E.[2], Ostano P.[1], Chiorino G.[1]
[1]Fondazione Edo Ed Elvo Tempia ~ Biella ~ Italy, [2]University of Turin ~ Turin ~ Italy

Sex-related differences are observed in many traits, including physical, psychological and genetic ones, with males generally displaying higher variability than females, hence the existence of "the greater male variability" hypothesis. Cutaneous malignant melanoma (CMM) is the most aggressive type of skin cancer, with mortality, incidence, recurrence, and metastatic rate higher in men than in women [1]. Our aim is to identify sex specific survival biomarkers in early-stage CMM patients, by analyzing tumor transcriptomic profiles and creating optimal survival models least affected by the observed sex differences.
To decipher transcriptomic sex differences in the context of survival, we used the Leeds melanoma cohort gene expression dataset on 567 stage I, II CMM individuals. Sex, age and stage adjusted multivariate analysis was performed, revealing a number of significant genes. However, sex- stratification of samples (in 311 females and 256 males) and a repetition of the same analysis showed that most genes obtained from all samples together were only significant in females. A significant difference was also observed in gene expression variability (GEV), with females having a lower GEV than males. This was also observed in normal skin and nevi samples as well as the TCGA melanoma dataset. Sex-stratified multivariate Cox regression, penalized Cox regression and concordance index all demonstrated significant differences between the two sexes, with females having a larger number of significant genes and overall higher concordance index values. After adjusting the p-value of the multivariate Cox regression, no genes were significantly (0.05) associated with survival in males. For female specific survival biomarkers, 14 genes were selected by penalized Cox regression based on the lambda 1 standard error of the LASSO model. These genes are now in the process of being validated on an independent cohort of stage I,II melanoma patients. We focused on the sexual dimorphism in melanoma and concluded that substantial transcriptomic differences in the context of survival exist. GEV might be directly or indirectly related to these observed differences and should be further investigated. Finally, our results have highlighted the need of sex-stratification in order to identify clinically translatable biomarkers.

[1] Schwartz, M. R., Luo, L., & Berwick, M. Current epidemiology reports, 6, 112-118(2019)

WP122 Comparison of statistical methods for survival data with time-dependent covariates and competing risks

Dadas Ö.F.*, Köse T.
Ege University ~ Izmir ~ Turkey

Survival analysis is a common statistical tool used to analyze data in biomedical research. However, analyzing survival data with time-dependent covariates in the presence of competing risks can be challenging. In this study, we evaluated three statistical methods for analyzing survival data with time- dependent covariates in the presence of competing risks: joint model, landmark approach, and multi- state model. We compared the results of these methods using simulated data and real data.
We generated simulated data with time-dependent covariates and competing risks. We also used real data from a biomedical study to evaluate the performance of the three methods. We conducted the analyses using the R software. We found that all three methods performed well in analyzing survival data with time-dependent covariates in the presence of competing risks. The joint model provided more accurate estimates of the hazard ratios than the landmark approach and the multi-state model. The landmark approach was more robust to model misspecification than the joint model and the multi- state model. The multi-state model provided more information on the transition probabilities between different states than the joint model and the landmark approach.
In conclusion, our study demonstrated that all three methods are viable options for analyzing survival data with time-dependent covariates in the presence of competing risks. The choice of method depends on the research question and the characteristics of the data. We recommend that researchers evaluate the performance of each method on their specific data before making a decision on which method to use. Our study provides a useful resource for clinical researchers and biostatisticians seeking to improve the analysis of survival data.

1. Elashoff RM, Li G, LiN. (2008). A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics 64:762–71
2. Kalbfleisch, J., & Prentice, R. (2002). The statistical analysis of failure time data (2nd ed.). New York: Wiley.
3. Nicolaie, M. A., Van Houwelingen, J. C., De Witte, T. M., & Putter, H. (2013). Dynamic prediction by landmarking in competing risks. Statistics in medicine, 32(12), 2031-2047.
4. Proust-Lima, C., & Taylor, J. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach. Biostatistics, 10, 535–549.
5. R Core Team. (2021). R: A language and environment for statistical computing. Vienna: R Found. Stat. Comput.
6. Rizopoulos, D. (2012). Joint models for longitudinal and time-to-event data, with applications in R. Boca Raton, FL: Chapman & Hall/CRC.
7. Rizopoulos, D., Molenberghs, G., & Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biometrical Journal, 59(6), 1261-1276.

WP123 Analysis of treatment outcome in an observational study in pediatric leukemia with a propensity score approach

De Lorenzo P.*[2], Bernasconi D.P.[1], Conter V.[3], Valsecchi M.G.[1]
[1]School of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy, [2]Tettamanti Center, Fondazione IRCCS San Gerardo dei Tintori ~ Monza ~ Italy, [3]Department of Pediatrics, Fondazione IRCCS San Gerardo dei Tintori ~ Monza ~ Italy

When the effect of treatment on outcome is assessed in observational studies, rather than in randomized controlled trials (RCTs), one must be aware that the decision to treat or not to treat may depend on subject characteristics, thus leading to a bias estimation. In order to eliminate or reduce the bias, regression models adjusted by measured confounders can be applied. Alternative methods are based on the propensity score (PS) that is, the probability of treatment assignment given the observed covariates. The PS strategy allows to construct two balanced cohorts of treated and untreated subjects as in a hypothetical RCT, and comprises a variety of approaches, including the inverse probability of treatment weighting (IPTW) [1]. Since the true PS is unknown, it is estimated with a logistic regression model. Selecting which covariates to include among regressors is a critical step, as it affects the ability to reduce bias and improve efficiency of the adjusted estimator of treatment effect. We applied the IPTW method to analyse the effect of cranial radiotherapy in a cohort of 66 children with leukemia, treated in an international clinical trial [2]. According to protocol, all patients were supposed to undergo cranial radiotherapy, while 42% (n=28) did not. Standard Kaplan-Meier analysis showed that cranial radiotherapy did not improve the event-free survival. However, treated and untreated patients differed with respect to baseline characteristics and prognostic factors, including age at diagnosis, white blood cells at diagnosis, and enrolling centers. Thus, the unexpected result needed to be further investigated to avoid potential confounding. We applied a logistic model to estimate the probability of treatment, exploring different sets of regressors, and used its inverse in a weighted Kaplan-Meier analysis. This approach confirmed the naïve finding about cranial radiotherapy not improving outcome. The IPTW method allowed to achieve more convincing evidence on the debated role of cranial radiotherapy in the treatment of childhood leukemia. Selecting regressors for PS estimation is not straightforward in practice, especially in rare diseases, where it can be hard to distinguish true from potential confounders and variables related to treatment allocation but not necessarily to outcome.

[1] Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011 May;46(3), 399–424.
[2] Biondi A, Gandemer V, De Lorenzo P, Cario G, Campbell M, Castor A, Pieters R, Baruchel A, Vora A, Leoni V, Stary J, Escherich G, Li CK, Cazzaniga G, Cavé H, Bradtke J, Conter V, Saha V, Schrappe M, Grazia Valsecchi M. Imatinib treatment of paediatric Philadelphia chromosome-positive acute lymphoblastic leukaemia (EsPhALL2010): a prospective, intergroup, open-label, single-arm clinical trial. Lancet Haematol. 2018 Dec;5(12):641-652.

WP124 Assessment of impact on the survival outcome due to disclosure of immature survival data in oncology trials

Djidel P.*[1], Colin P.[1], Mazzei A.[1], Mishra K.[2], Appanna K.[2]
[1]Bristol Myers Squibb ~ Boudry ~ Switzerland, [2]Bristol Myers Squibb ~ Princeton Pike ~ United States of America

In confirmatory Oncology trials, often Progression-Free Survival (PFS) is the primary endpoint and Overall Survival (OS) is the key secondary endpoint tested hierarchically. When PFS based regulatory filing for drug approval occurs, either at an Interim or Final positive PFS analysis, OS data is often immature. The immaturity of OS data could be due to low information fraction and/or limited survival follow-up. However, such data, is often requested by Health Authorities and may be disclosed in reports available to public. Disclosure of treatment effect based on immature OS data could introduce bias in the final OS outcome. This can happen due to subjects that prematurely discontinued treatment or switched to another treatment during the trial, based on the perception of potential OS treatment effect from such disclosures. Here, we explore the impact of disclosing immature OS data on the final OS outcome relative to different levels data maturity and on-treatment subject proportion. We also study the effect of non- proportional hazard often observed in Immuno-Oncology trials. This can help assess the potential and extent of bias in the final OS readout and thereby quantify the risk associated with final OS outcome being positive. Preliminary results have demonstrated significant impact on power and average hazard ratio when the information fraction of OS is greater than 80% and when a meaningful number of participants switch treatment, for subsequent therapies. Further results will be presented by exploring different scenarios. This work will be further developed with evaluation of non-proportional hazard assumption and delayed effects as well.

1. Wayant C, Vassar M. A comparison of matched interim analysis publications and final analysis publications in oncology clinical trials. Ann Oncol. 2018 Dec 1;29(12):2384-2390. doi: 10.1093/annonc/mdy447. PMID: 30307531.
2. Lux MP, Ciani O, Dunlop WCN, Ferris A, Friedlander M. The Impasse on Overall Survival in Oncology Reimbursement Decision-Making: How Can We Resolve This? Cancer Manag Res. 2021 Nov 10;13:8457-8471. doi: 10.2147/CMAR.S328058. PMID: 34795526; PMCID: PMC8592394.
3. Tai TA, Latimer NR, Benedict Á, Kiss Z, Nikolaou A. Prevalence of Immature Survival Data for Anti- Cancer Drugs Presented to the National Institute for Health and Care Excellence and Impact on Decision Making. Value Health. 2021 Apr;24(4):505-512. doi: 10.1016/j.jval.2020.10.016. Epub 2020 Dec 8. PMID: 33840428.

Poster Sessions

WP125 Efficient estimation of the marginal mean of recurrent events in randomized clinical trials

Luca G.*[1], Giuliana C.[1], Henrik R.[2], Thomas S.[3]
[1]University of Padua ~ Padua ~ Italy, [2]Novo Nordisk ~ Copenaghen ~ Denmark, [3]University of Copenaghen ~ Copenaghen ~ Denmark

Recently, in the pharmaceutical industry there has been a renewed interest in studying recurrent events data more efficiently and correctly. In regard to this, the present work focuses on randomized clinical trials (RCTs), where it is often of interest to estimate the treatment effect of a new drug, that in our case is the difference of the marginal mean of the recurrent events between the treated and untreated subjects. In the RCT setting, this difference is an important estimand that can be estimated by comparing the simple marginal means estimator of each group, separately. In this work, we show how to improve the efficiency of such estimand theoretically, asymptotically and in practice by extensive simulations studies. The efficient estimator is a doubly augmented estimator that can be expressed as having an optimal censoring augmentation and an optimal RCT augmentation. These two augmentations can be dealt with separately because of their orthogonality. When applying working models to the two augmentations we are still guaranteed a variance reduction, asymptotically, even when the working models are misspecified. Both augmentations are easy to compute and rely on fitting standard models. The methods are applied to the LEADER study where we demonstrate that there are indeed important efficiency gains to be obtained. In each case tested in the simulations the double augmented estimator provides an efficiency's gain, although its amount varies in the different settings considered and depends on several factors like for example the number of recurrent events collected and the censoring distribution. Therefore we can conclude that the double augmented estimator is a good trade-off between the efficiency's amount reachble and the robustness of the results, thanks to its double robustness property. In regard to this, thiswork tries to answer the important request by the European Medicines Agency about efficient statistical estimation methods for the treatment effect on estimands of recurrent event endpoints

[1] Akacha, M., Binkowitz, B., Bretz, F., Fritsch, A., Hougaard, P., Jahn-Eimermacher, A., Mendolia, F., Ravn, H., Roger, J., Schlömer, P., et al. (2018). Request for chmp qual- ification opinion: Clinically interpretable treatment effect measures based on recurrent event endpoints that allow for efficient statistical analyses. Recurrent Event Qualification Opinion Consortium.
[2] Cortese, G. and Scheike, T. H. (2022). Efficient estimation of the marginal mean of recurrent events. Journal of the Royal Statistical Society Series C, 71(5):1787–1821
[3] Furberg, J. K., Rasmussen, S., Andersen, P. K., and Ravn, H. (2022). Methodological chal- lenges in the analysis of recurrent events for randomised controlled trials with application
to cardiovascular events in leader. Pharmaceutical Statistics, 21(1):241–267
[4] Marso, S. P., Daniels, G. H., Brown-Frandsen, K., Kristensen, P., Mann, J. F., Nauck, M. A., Nissen, S. E., Pocock, S., Poulter, N. R., Ravn, L. S., et al. (2016). Liraglutide and cardiovascular outcomes in type 2 diabetes. New England Journal of Medicine, 375(4):311– 322.

WP126 Comparison of asymptotic and re-randomization tests under non- proportional hazards scenarios with minimization

Kimura R.*[2], Nomura S.[3], Nagashima K.[1], Sato Y.[4]
[1]Biostatistics Unit, Keio University Hospital ~ Tokyo ~ Japan, [2]Graduate School of Health Management, Keio University ~ Kanagawa ~ Japan, [3]Department of Biostatistics and Bioinformatics, The University of Tokyo ~ Tokyo ~ Japan, [4]Department of Preventive Medicine and Public Health, Keio University School of Medicine ~ Tokyo ~ Japan

Pocock–Simon's minimization method has been widely used to balance treatment assignments across prognostic factors in randomized clinical trials (RCTs). Previous studies have shown that a statistical test that relies on asymptotic normality without adjusting stratification factors is conservative. In contrast, a re-randomization test without adjustment of these factors holds the type I error rate at the nominal level, resulting in a power gain. This study focuses on RCTs with the minimization method and survival endpoint. The objective is to examine the performances of various statistical tests under non proportional hazards (PH) scenarios in RCT with a minimization method referring to a previous study by Xu et al. [1] that compared a (stratified) log-rank test and adjusted Cox model under limited PH scenarios and limited sample sizes. The simulation scenarios include delayed, crossing, and diminishing effects. The following tests are considered: the log-rank test; Fleming–Harrington G^(ρ,γ) test; tests based on a maximum of G^(ρ,γ) statistics (MaxCombo); the Wald test based on Cox models; and difference in restricted mean survival time (dRMST) up to time point τ. We also assess the performances of these methods when randomization stratification factors are used for stratification and/or regression. Empirical type I error rates and power are the primary measures of interest. The type I error rates of the unadjusted asymptotic methods were conservative, and those of the adjusted Cox model and adjusted dRMST were inflated due to limited sample sizes. In contrast, the type I error rates of the re-randomization tests were always held at the nominal level. Subsequently, statistical power was compared between methods without inflated type I error rates. Under PH scenarios, the powers of re-randomization tests based on the adjusted Cox model were higher than those of other methods. Under delayed and crossing scenarios, the powers of re-randomization tests based on the unadjusted MaxCombo were higher than those of other methods. Under a diminishing scenario, those of the adjusted MaxCombo were the highest. Our findings may support practitioners in selecting appropriate tests for survival outcomes in the RCTs with a minimization method.

1. Xu Z, Proschan M, Lee S. Validity and power considerations on hypothesis testing under minimization. Stat Med 2016;35(14):2315-2327.

WP127 Stratified cox models under partly interval censoring

Ma J.*, Webb A., Manuguerra M., Hudson M.
Macquarie University ~ Sydney ~ Australia

The conventional semiparametric Cox model demands the proportional hazards (PH) assumption when covariates are not time dependent. Clearly, requiring all covariates to satisfy the PH assumption can be a strong condition in many applications. One remedy is to stratify the covariates that fail the PH assumption and then to allow stratum-specific baseline hazards, leading to the stratified Cox models; see, for example, Andersen et al. (1993) and Martinussen and Scheike (2006). In this talk, we study the stratified Cox models under the general partly interval censoring scheme where the follow-up times may include event times and right-, left- and interval-censoring times. If observed times contain interval censoring, the partial likelihood method cannot be implemented directly. We will present a maximum penalized likelihood approach to estimate regression coefficients and baseline hazards. We adopt penalty functions in the log-likelihood for two reasons: (i) to ensure smoothed baseline hazard estimates, and (ii) to reduce the dependence of the estimates on the number and location of knots used to approximate the baseline hazards. A large sample normality result will be provided with our proposed method, and this result enables inferences to be made not only on regression coefficients, but also on other quantities of interest uch as hazard or survival functions. A penalized likelihood method is developed to fit stratified Cox models, including estimation of the regression coefficients as well as the baseline hazards. We approximate the unknown baseline hazards using basis functions, such as M-splines, and our proposed approach finds the maximum penalized likelihood estimate of coefficients of the basis functions, where these baseline hazards are constrained to be non-negative. The large sample normal distribution can be used for performing inferences, for example hypothesis testing on Cox model regression coefficients or constructing confidence intervals for predictive survival curves. When compared with a commonly adopted approach where mid-point imputations replace the interval censored data and then applying the partial likelihood approach, our method can produce smaller biases and better coverage probabilities on regression coefficients and smaller mean integrated squared error on the baseline hazards estimates.

Andersen, P. K., Ørnulf, B., Richard, D. G. and Niels, K. (1993), Statistical models based on counting processes, Springer, New York.
Martinussen, T. and Scheike, T. H. (2006), Dynamic regression models for survival data, Springer, New York.

Poster Sessions

WP128 The performance of overlap weighting ps method for survival analysis when interaction in a subgroup exist

Nishikawa M.[1], Nishikawa T.*[2]
[1]*The Jikei University School of Medicine ~ Tokyo ~ Japan,* [2]*Yokohama City University ~ Kanagawa ~ Japan*

Propensity score (PS) is very often used to reduce the effects of confounding factors when estimating treatment and/or exposure effect in observational studies. In survival analysis, among several PS methods, PS matching and inverse probability of treatment weighting(IPTW) allow for the estimation of marginal hazard ratios(MHRs) with minimal bias (Austin, 2013). IPTW had been the most used weighting method before overlap weighting PS(OWS) was proposed and shown to be superior to IPTW in balancing. In application, subgroup analyses are also important to examine the homogeneity of the effect over entire population. Interactions of patients' characteristics and treatment are often observed. Conditional hazard ratios(CHRs) estimated with IPTW was shown to be biased(Austin, 2013). However, there are no simulation studies on the performance of OWS in time-to-event outcome. The objective of our research was to compare the performance of OWS with those of PS matching and IPTW for estimating MHRs and CHRs by simulation, taking subgroup interaction into consideration. In our simulation setting, there were 10 binary covariates including subgroup indicator. Their association with treatment allocation and outcome were assumed either combination of strong, moderate, or none. For each subject, the probability of receiving treatment was generated according to a logistic model, and the Bernoulli distribution was used to determine imbalanced/balanced presence or absence of treatment and/or covariate level. Survival times were generated using Weibull distribution, where the linear predictor was expressed with/without subgroup interaction terms. Uniform distribution was used for censoring times. True MHRs were set and estimated by similar way as Austin (2013). In the analyses, PS model contained all covariates. For estimating CHRs, all the covariates related to the outcome were included in the Cox proportional hazard model(CPHM), with/without interaction term. In the setting of subgroup interaction, recalculating IPTW/OWS by level of subgroup, using the original IPTW/OWS with adjusted CPHM, and contrast among 2 by 2 levels were explored. In each scenario, mean treatment effect (log-hazard ratio), bias, SD, MSE and coverage of the 95% CI(naïve, robust) were calculated. Results will be shown on presentation. Our finding will be very useful when using PS in survival analysis. Gayat et al. Pharmaceutical Statistics. 2012;11:222-229. Austin. Statistics in Medicine. 2013;32:2837-2849.

WP129 Evaluating surrogate endpoints for overall survival in rcts testing immune checkpoint inhibitors

Pagan E.*[1], Sala I.[2], Oriecuia C.[3], Specchia C.[4], Gelber R.[5], Conforti F.[6], Bagnardi V.[1]
[1]*Department of Statistics and Quantitative Methods, University of Milan-Bicocca ~ Milan ~ Italy,* [2]*Department of Medicine and Surgery, University of Milan-Bicocca ~ Milan ~ Italy,* [3]*Department of Clinical and Experimental Sciences, University of Brescia ~ Brescia ~ Italy,* [4]*Department of Molecular and Translational Medicine, University of Brescia ~ Brescia ~ Italy,* [5]*Dana-Farber Cancer Institute, Harvard Medical School, Harvard T. H. Chan School of Public Health, and Frontier Science and Technology Research Foundation ~ Boston ~ United States of America,* [6]*Department of Medical Oncology, Humanitas Gavazzeni ~ Bergamo ~ Italy*

Overall survival hazard ratio (HRos) is the gold-standard endpoint used to demonstrate the clinical efficacy of new cancer drugs in randomized clinical trials (RCTs). A reliable estimation of HRos requires large RCTs with long follow-up. To expedite drug approvals, the evaluation of new treatments in RCTs often relies on the assessment of their effects on surrogate endpoints, such as progression-free survival (PFS). Immune checkpoint inhibitors' (ICI) novel mechanisms of activating self-immunity against tumors could result in delayed clinical effects and long-term responders, and also in disease progression followed by tumor shrinkage (pseudo-progression), leading to the violation of the proportional hazard (PH) assumption on which the calculation of HR is based. The restricted mean survival time (RMST) was proposed as an alternative measure to account for deviation from PH assumption, and the modified PFS (mPFS) as a novel endpoint to omit pseudo- progressions from PFS [1]. Our aim was to compare the surrogacy value of PFS and mPFS for OS in RCTs testing ICIs, when the treatment effect is measured by the HR for OS, and by the HR and the ratio of RMST (rRMST) for PFS and mPFS. We systematically searched for phase II and III RCTs testing ICIs in patients with advanced solid tumors, up to December 2021. We reconstructed pseudo individual patient-level data using a validated algorithm [2]. For each treatment comparison we calculated the treatment effect measures. We assessed the trial-level correlation between (m)PFS treatment effect measures and HRos, in strata of immunotherapy strategy (i.e., ICI alone, ICI plus chemotherapy, ICI plus other treatment(s)), using weighted linear regression models. We identified 61 RCTs (67 treatment comparisons and 36,034 patients). HRpfs and HRmpfs had a strong surrogacy value in comparisons testing ICI plus chemotherapy (R^2=0.74 and R^2=0.81). HRpfs was the best surrogate in comparisons testing ICI as monotherapy, although having a moderate correlation (R^2=0.58). The value of potential surrogates for HRos was strongly affected by the type of treatment(s). Our results do not support the use of alternative endpoints, such as the mPFS, or treatment effect measures, such as the RMST.

[1] .X. Wang, H.X. Wu, L. Xie, et al. Exploration of modified progression-free survival as a novel surrogate endpoint for overall survival in immuno-oncology trials. J Immunother Cancer 2021; 9:e002114. doi:10.1136/jitc-2020-002114
[2] P. Guyot, A.E. Ades, M.J.N.M. Ouwens, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012; doi:10.1186/1471-2288-12-9

WP130 Outcome prognostication in acute brain injury using the neurological pupil index (orange) study.

Petrosino M.*, Citerio G., Rebora P., Galimberti S.
University of Milano-Bicocca ~ Milano ~ Italy

The standard care of Intensive Care Unit patients with acute brain injury involves the pupillary light reactivity evaluation as part of daily management having strong diagnostic and prognostic value. However, it is usually performed using a hand-held light source, which provides an unstandardized qualitative measurement. In this study, an automated infrared pupillometry, returning the NPi (Neurological Pupil index) [1], a composite numerical index ranging from 0 to 5 (abnormal values<3) of pupillary reactivity and global midbrain function, was used. The aim was to assess the prognostic efficacy of the NPi in predicting the risk of 6 months mortality. The ORANGE was an international, multicenter, prospective, observational study that enrolled 514 patients with acute brain injury requiring intensive care in 13 centres in Europe and the USA. Overall, NPi measurements were collected every 4 hours on the right and left eye of each patient, starting from ICU admission until day 7, for a total of 40,071 values. Multivariable adjusted extended Cox models with lowest NPi on both eyes as a time varying covariate were deployed. We found that a one-unit NPi increase was associated with a significantly lower hazard ratio (HR) of mortality (HR 0.55, 95% CI 0.50-0.62, p<0.0001). NPi<3 (HR 7.10, 95% CI 4.77-10.57, p<0.0001) and NPi between 3-4 (HR 1.70, 95% CI 1.13-2.56, p=0.0163) were associated to higher mortality, with NPi>4 as reference. As it is crucial to understand the implications of improving NPi from zero to a positive value, we extended our analysis considering the actual NPi(ti) and the preceding NPi(ti-1). Using two consecutive values of NPi >0 as reference category, the occurrence of two consecutive NPi=0 was associated with an HR of 13.92 (95% CI 8.94-21.67, p<0.0001). A deterioration of an NPi(ti-1)>0 to NPi(ti)=0 showed an HR of 8.37 (95% CI 2.52-27.87, p=0.0005). Notably, the hazard of death when NPi(ti-1)=0 improved to a NPi(ti)>0 was not statistically different from that of the reference category (HR 1.32, 95% CI 0.32-5.41, p=0.6994). Repeatedly low and deteriorating NPi to zero values were strongly associated with poor prognosis, while NPi improvement over time bring mortality risk back to a baseline level. Chen JW, Vakil-Gilani K, Williamson KL, Cecil S. Springerplus. 2014;3(1):548

WP131 Impact of censoring mechanisms in assessment of prognostic technologies

Rana D.*, Hawkins N.
Health Economics and Health Technology Assessment, School of Health and Wellbeing, University of Glasgow ~ Glasgow ~ United Kingdom

Methodologically rigorous survival analysis is vital to properly evaluate the predictive ability of prognostic technologies, particularly when its output informs cost-effectiveness modelling. However, censoring is a key challenge which arises when time-to-event data for study participants is incomplete. Interval censoring (event is observed but the exact event time is unknown) is common where continuous monitoring of the outcome is not possible, and disease diagnosis occurs between the last and current surveillance [1]. Instances where data have been inappropriately handled by assuming left (event is observed before a specific time point) or right censoring (event has not been yet observed but occurs beyond the specific time point) may create bias and mislead researchers and policymakers. We aim to evaluate the impact of the left, right and interval censoring mechanisms in assessing prognostic technologies. Retrospective data was used. The cohort consisted of 2,643 patients who underwent polypectomy at screening colonoscopy between May 2009 and December 2016 as a part of the Scottish Bowel Cancer Screening Programme. The patients were re-categorised as high-risk and low-risk for polyp recurrence according to the British Society of Gastroenterology (BSG) post-polypectomy surveillance guidelines 2020. We performed polyp-free survival analysis using the Cox proportional hazards regression assuming left, right and interval censoring in the data. Kaplan-Meier plots were drawn to visualise and compare the impact of censoring. Assumption of left censoring (at last negative colonoscopy) resulted in nearly all the patients having a polyp recurrence by six years; right censoring (at end of follow-up) resulted in patients remaining polyp-free from the date of last negative colonoscopy until the end of follow-up. Both scenarios result in bias: left censoring overestimates recurrence, and right censoring underestimates recurrence following the last colonoscopy. Interval censoring showed that most polyp recurrences in high-risk patients occur within the first three years. Our research demonstrates that censoring has the potential of biasing results if not handled appropriately. Interval censoring findings reflect BSG guidelines' surveillance recommendations, indicating that this method has the potential to reduce bias and should be considered when assessing the predictive ability of prognostic technologies. Turkson AJ, Ayiah-Mensah F, Nimoh V. Handling censoring and censored data in survival analysis: a standalone systematic literature review. International journal of mathematics and mathematical sciences. 2021 Sep 24;2021:1-6.

Poster Sessions

WP132 Pairwise cox modeling approach for causally interpretable average hazard ratio under nonproportional hazards

Shinozaki T.*, Chikamochi T.
Tokyo University of Science ~ Tokyo ~ Japan

Hazard ratios are a commonly used summary measure in clinical trials for survival analysis under proportional hazards. However, the interpretability of a single hazard ratio estimate becomes questionable under nonproportional hazards, which are prevalent in recent research due to various factors such as delayed treatment effects and heterogeneous subgroups. Nonproportional hazards also imply different susceptibilities to treatment among patients, causing selection bias and preventing causal interpretation. Nonetheless, it is possible to average time-varying hazard ratios into a well- defined summary measure for comparing two counterfactual distributions (Schemper, et al, 2009). We present a novel estimator for the causally interpretable average hazard ratio using a pairwise Cox modeling approach. We employ a pair-stratified Cox model for pairwise data and weight it by the double inverse probability of censoring weight for each pair to obtain a consistent estimator of the target estimand. Notably, this approach does not require the proportional hazards assumption. However, we use the pairwise Cox modeling approach to easily obtain a robust sandwich variance estimator that clusters individuals across pairwise data (Lin, 1994). To address our situations where the number of strata (i.e., combinations of every 2 data in the original sample) is much larger than the number of clusters (i.e., original sample size), we modify score residuals in the robust variance estimator using the "termination" technique. This technique immediately censors data at an event or censoring of its matched data, where the marginal Cox models fitted to terminated data are equivalent to pair-stratified Cox models (Shinozaki & Mansournia, 2019). We compare our proposed estimator with a rather heuristic estimator using weighted Cox models (Schemper, et al, 2009) in a simulation study. Our estimator is less biased and more stable, regardless of the proportionality of hazards. Furthermore, the confidence intervals achieved an almost nominal coverage rate. Even if the proportional hazards assumption is violated, hazard ratios remain useful summary measures of treatment effects when an adequate average is taken over time. Our proposed estimator allows for valid inference of that estimand using off-the-shelf software, even under nonproportional hazards. Lin, D.Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine, 13, 2233-2247.
Schemper, M., Wakounig, S., and Heinze, G. (2009). The estimation of average hazard ratios by weighted Cox regression. Statistics in Medicine, 28, 2473–2489.
Shinozaki, T. and Mansournia, M.A. (2019). Hazard ratio estimators after terminating observation within matched pairs in sibling and propensity score matched designs. International Journal of Biostatistics, 15. doi: 10.1515/ijb-2017-0103.

WP133 Joint modelling of longitudinal biomarkers and of risk of serious non-aids event in people living with hiv

Stamoulopoulos A.*, Thomadakis C., Pantazis N., Touloumi G.
Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens ~ Athens ~ Greece

Introduction and widespread of antiretroviral therapy (ART) has dramatically improved life expectancy of people living with HIV (PLWH); they still though experience increased morbidity and mortality compared to the general population, mainly due to serious non-AIDS events (SNAEs). The drivers behind this increased risk are likely multi-factorial. Development of dynamic prediction models combining longitudinal biomarkers' data with time to event data could contribute identifying PLWH at high risk. We investigated the association of longitudinal CD4 and serum albumin levels with time to first SNAE in PLWH on ART, within the joint modelling framework that allows simultaneous modelling of multiple, possibly correlated longitudinal markers and that of the hazard of a survival endpoint. Data were derived from the AMACS (Athens Multicenter AIDS cohort study). A joint modelling approach of the two markers and of the time to first SNAE was adopted. A multivariate mixed effects model was employed for the longitudinal evolution of the two markers using natural cubic splines of time with two knots at 3 months and 4 years for the CD4 counts and a cubic polynomial for the serum albumin (in both fixed and random effects of each marker, respectively). For the survival part, a cox proportional hazards model was utilized which included as covariates the current estimated ("true") value of both markers. Estimation procedure was carried out using MCMC algorithms (R: package JMbayes2). Model's predictive accuracy was assessed by calculating the expected prediction error (EPE) for various combinations of follow-up time points and of time windows within which predictions were evaluated, using a 5-fold cross validation. Both markers were inversely associated with increased risk for SNAE. Cross-validated EPEs displayed low values (observed range: 0.002-0.032) indicating a decent model predictive performance. Our results are in line with those previously published regarding the direction of the association of the two markers with study's endpoint. As SNAEs consist a complex endpoint, the utilization of multiple longitudinal markers is essential for the development of prognostic tools and joint models seem well- suited towards this end. Further work incorporating competing risks data is under investigation.

WP134 Adjusted kaplan-meier estimate for prediction of a decrease of covid-19 antibodies below laboratory cut-off

Stepanek L.*[3], Habarta F.[1], Mala I.[1], Marek L.[1], Stepanek L.[2]
[1]Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business ~ Prague ~ Czech Republic, [2]Department of Occupational Medicine, University Hospital Olomouc and Faculty of Medicine and Dentistry, Palacký University Olomouc ~ Olomouc ~ Czech Republic, [3]Department of Statistics and Probability, Faculty of Informatics and Statistics, Prague University of Economics and Business & Institute of Biophysics and Informatics, First Faculty of Medicine, Charles University ~ Prague ~ Czech Republic

Estimation of survival function, i.e., a probability of non-experiencing an event of interest by an individual until a given time point, is commonly approached using the Kaplan-Meier estimator. The Kaplan-Meier estimator of the survival probability for a given time point assumes that the event of interest experienced by an individual is irreversible. More specifically, an individual could not be considered a non-experiencing anymore if they have experienced the event. Also, they cannot be censored once they experience the event. However, there are situations when an event of interest could be spontaneously reversible in time, with no options for adequately or on-the-fly capturing the event's change in time. In this work, we address the described problem and adjust the Kaplan-Meier estimator to consider that not all of those individuals who experienced the event of interest have necessarily to stay in the state determined by the event; they could transit back to the state before the event experience. So, the event of interest is reversible in time, and its change is impossible to measure in real-time. Thus, since we cannot use multistate models, we derive exact formulas for the adjusted Kaplan-Meier estimator and its variance using the delta method. Finally, we apply the proposed estimation to 663 individuals' decrease of COVID-19 blood antibodies below laboratory cut-off. An individual's COVID-19 blood antibodies below the laboratory cut-off (i.e., seronegativity) may mean not only that the individual antibodies have exceeded the cut-off but decreased below it but also that the antibodies have been growing but have reached the cut-off yet. Thus, an individual whose COVID-19 antibodies are evinced as decreased below the laboratory cut-off could still reach the cut-off (i.e., seropositivity) afterward, having sufficient COVID-19 blood antibody level. The data gathered as a snapshot of a time point could underestimate true COVID-19 antibody blood levels in time using the traditional Kaplan-Meier estimator. The proposed adjustment enables handling such data and explains why other studies may indicate a level of COVID-19 antibody seropositivity in time higher than the actual data.

[1] Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53, 1958, 457--481.
[2] Štěpánek, L., Habarta, F., Malá, I., Štěpánek, L., Nakládalová, M., Boriková, A., & Marek, L. Machine Learning at the Service of Survival Analysis: Predictions Using Time-to-Event Decomposition and Classification Applied to a Decrease of Blood Antibodies against COVID-19. Mathematics, 11, 2023, 819--845.

WP135 Estimation of the cutoff value for continuous prognostic factors in survival data with competing risk

Woo S.Y.*
Samsung medical center ~ Seoul ~ Korea, Republic of

In clinical studies, biomarkers that are prognostic factors for specific diseases are often converted into categorical variables and analyzed. When estimating the cutoff value of prognostic factors in survival data without competing risk, methods for determining the cutoff value have been proposed using ROC (receiver operating characteristic) curve, test statistic, and concordance. In the situation of survival data with competing risk, a method of estimating the cutoff value with a method based on the ROC curve has been already proposed. However, few studies have been reported that estimating the cutoff value using methods related to test statistics and concordance in survival data framework with competing risk. Also, the performance of these methods used has never been compared. To estimate the optimal cutoff value of the baseline continuous biomarker for survival data without or with competing risk, test statistics based methods, concordance based methods, and ROC based methods were described. The performances of the methods for determining the cutoff value through various simulation scenarios were compared and applied to real data. The test statistic using IPCW tended to have a small bias in the cutoff value even when the true cutoff value was away from the center of the distribution of prognostic factors when there was no censoring rate. On the other hand, when the censoring rate was large, concordance based measure tended to have good performance even when the true cutoff value was away from the center of the distribution of prognostic factors.

Woo S et al. Determining cutoff values of prognostic factors in survival data with competing risks. Comput. Stat 2016;31:369–386.
Doyeong Yu and Wei-Ting Hwang. Optimal cutoffs for continuous biomarkers for survival data under competing risks. Communications in Statistics. 2019; 48: 1330–1345.

Poster Sessions

WP136 Construction of survival tree based on restricted mean survival time

Shimokawa A.*, Miyaoka E.
Tokyo Universito of Science ~ Tokyo ~ Japan

In recent years, restricted mean survival time (RMST) is paid attention as an evaluation index of clinical trials with time-to-event outcome. Compared to the Cox proportional hazards model, which is used under the proportional hazards assumption, the RMST has the advantage that it can be used even when the assumption does not hold, and the results obtained are easy to interpret (Royston and Parmar 2013). In addition, the estimation of RMST is straightforward in either a parametric or nonparametric framework. Incidentally, as one of the models for investigating the influence of covariates on response variables, the tree-structured model is used in a wide range of fields. A tree-structured model for time-to-event data is called a survival tree and is generally constructed using the log-rank test based splitting criterion (Leblanc and Crowley 1993). This criterion assumes that the survival functions between each node do not intersect. In fact, however, there are cases where this assumption is not satisfied, so we propose a survival tree construction method that focuses on the RMST. The classification and regression tree algorithm which consists of splitting, pruning, and selection steps is a popular method of constructing a survival tree. It is well known that the setting of the splitting criterion to construct the tree model is especially important. In this study, we propose a new splitting criterion that minimizes the prediction error of the RMST for the tree. That criterion consists of the sample proportion and the RMST estimator within each child node. We also propose a cost-complexity measure-based pruning criterion to minimize the prediction error of the RMST. From simulation studies, it was shown that the proposed method has better performance than the conventional method in terms of prediction of RMST, especially in the case of the proportional hazards do not hold. We proposed a new method to construct a survival tree based on the RMST. The performance of the proposed method and the conventional method were compared through simulations. In addition to this, we will show the results of the tree obtained from an actual data in the conference.

Royston P, Parmar M K B. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013, 13: 152.
Leblanc M, Crowley J. Survival trees by goodness of split. J Am Stat Assoc. 1993, 88: 457–67.

ISCB44

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

MILAN, ITALY | 27 – 31 AUGUST 2023

University of Milano-Bicocca
Building U6 Piazza dell'Ateneo Nuovo 1

Organising Secretariat
Promoest srl, Via G. Moroni 33, 20146 Milano (MI)
iscb2023@promoest.com

Mini Symposia 1-2 and ECB day

Mini Symposia 1

TEN YEARS STRATOS INITIATIVE – BRIEF SUMMARY OF PROGRESS AND PLANS FOR THE FUTURE

COORDINATORS: WILLI SAUERBREI AND FEDERICO AMBROGI

TMS1.1 Experience and progress with developing guidance for the analysis of key topics in observational research

Sauerbrei W.*[4], Abrahamowicz M.[1], Le Cessie S.[5], Huebner M.[2], Keogh R.[3], Carpenter J.[3]
[1]McGill University ~ Montreal ~ Canada, [2]Michigan State University ~ East Lansing ~ United States of America, [3]London School of Hygiene & Tropical Medicine ~ London ~ United Kingdom, [4]Medical Center - University of Freiburg ~ Freiburg ~ Germany, [5]Leiden University Medical Center ~ Leiden ~ Netherlands

The STRATOS initiative was launched at ISCB 2013 and the first STRATOS paper summarized the motivation, mission, structure and aims of this international initiative ([1], https://www.stratos- initiative.org/en). Providing accessible, evidence-based guidance for key topics in the design and analysis of observational studies is the main aim. Guidance is intended for applied statisticians and other data analysts with varying levels of statistical background and experience. The focus is on health sciences research, but the content is also relevant for applications of statistics in other empirical sciences. In 2013 the STRATOS initiative started off with seven topic groups (TGs) focusing on different aspects of study design and analysis methodology (1- Missing data, 2- Selection of variables and functional forms in multivariable analysis, 3- Initial data analysis, 4- Measurement error and misclassification, 5 - Study design, 6- Evaluating diagnostic tests and prediction models, 7- Causal inference). For their specific topic, each group provided a brief summary of the state of research, main issues, main aims and planned future research (Sauerbrei et al., 2014). Two further TGs were initiated in 2015 on the topics of Survival analysis (TG8) and High-dimensional data (TG9). Summaries are available on the STRATOS website. To coordinate the activities of the initiative, and to help improve standards of both methodological and applied research, we started several cross-cutting panels, that work on issues common to all TG's, including simulation, visualization, and most recently about open science. In this talk we will provide a short introduction illustrating the necessity of guidance for analysis of observational studies and outline experience and progress of the STRATOS initiative.

[1] Sauerbrei, W., Abrahamowicz, M., Altman, D. G., le Cessie, S., Carpenter, J., & STRATOS initiative. (2014). STRengthening analytical thinking for observational studies: the STRATOS initiative. Statistics in medicine, 33(30), 5413-5432.

TMS1.2 Level 1 guidance on conducting and reporting sensitivity analyses for missing data

Lee K.*[1], Mainzer R.[1], Carpenter J.[2]
Murdoch Children's Research Institute ~ Melbourne ~ Australia, [2]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom

Missing data are common in observational studies. When estimating a target parameter in the presence of missing data, the researcher (either implicitly or explicitly) makes assumptions about the unknown missingness mechanism. An important, but often overlooked step of the analysis is examining the robustness of estimates to alternative plausible assumptions about the missingness mechanism, and in particular conducting analyses that allow the missingness to depend on the missing values themselves, sometimes referred to as a "missing not at random" or a "delta-adjusted" analysis. We previously outlined a framework for handling and reporting the analysis of incomplete data in observational studies where we encourage researchers to think systematically about missing data and transparently report the potential effect on the study results. We extend this framework to the planning, conduct and reporting of sensitivity analyses which incorporate external information about how the missing values differ to those observed. We illustrate the process using a case study from the Avon Longitudinal Study of Parents and Children, providing practical guidance that can be tailored to the problem at hand. We hope this much needed guidance will make such sensitivity analyses more accessible to researchers, increasing its use in practice, and increasing the confidence in research findings from incomplete data. Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, Hogan JE, and Carpenter JR. Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework (2021). Journal of Clinical Epidemiology, 134; 79-88. https://doi.org/10.1016/j.jclinepi.2021.01.008. Tompsett DM, Leacy F, Moreno-Betancur M, Heron J, White IR. On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. Stat Med. 2018 Jul 10;37(15):2338-2353. doi: 10.1002/sim.7643. Epub 2018 Apr 2. PMID: 29611205; PMCID: PMC6001532.

TMS1.3 Aims of the new open science panel

Luijken K.*[1], Hoffmann S.[2], Boulesteix A.[2]
[1]University Medical Center Utrecht ~ Utrecht ~ Netherlands, [2]Institute for Medical Information Processing, Biometry, and Epidemiology ~ Munich ~ Germany

In recent years, the realization that published research findings across many disciplines are not as reliable as previously assumed has led to a "replication crisis" [1] or "statistical crisis in science" [2]. In response to this crisis, the scientific community, publishers, funders, and other stakeholders are increasingly encouraging open science practices. To help remove barriers that still exist in the adoption of open science practices, the STRATOS steering group started an Open Science panel. The aim of the Open Science Panel is to promote open science practices, both within the STRATOS initiative and by providing accessible guidance for the scientific community on ways to make research more transparent, reproducible, and credible. We plan to work on tutorials that help researchers in making their analysis code reproducible and dealing with "researcher degrees of freedom". In the long run we also intend to provide guidance on data sharing approaches that aim to find a compromise between preserving privacy protection and allowing reproducing results of statistical inference in a valid way. The panel is presently chaired by Sabine Hoffmann. This talk at ISCB is meant to give a brief overview of the topics the Open Science panel is working on, with the intention to motivate further researchers to enroll in the panel as a member.

[1] Patrick E Shrout and Joseph L Rodgers. Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. Annual Review of Psychology, 69:487–510, 2018.
[2] Andrew Gelman and Eric Loken. The statistical crisis in science data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. American Scientist, 102(6):460–465, 2014.

TMS1.4 Ongoing research towards state-of-the-art in variable and functional form selection for statistical models

Heinze G.*[1], Perperoglou A.[2], Sauerbrei W.[3]
[1]Medical University of Vienna ~ Vienna ~ Austria, [2]GSK ~ Stevenage ~ United Kingdom, [3]Medical Center - University of Freiburg ~ Freiburg ~ Germany

Topic group 2 (TG2) of the STRATOS initiative deals with multivariable model building, in particular with issues in building suitable descriptive regression models. In particular, our TG will provide guidance on strategies for variable selection and on the specification of the functional form of nonlinear effects of continuous covariates. Medical literature is still full of outdated statistical approaches, because there is a lack of awareness of possible pitfalls of commonly used methods even in very common scenarios. Contrary to that, a large number of new methods is proposed, but most of these methods are stuck in an early phase of development (see Heinze et al, 2023). There is insufficient evidence of them being fit for purpose. Often guidance for suitable methods is missing scientifically inferior procedures are used to analyse medical data, leading to questionable medical conclusions. These problems affect all STRATOS topics, but here we focus on multivariable modelling. In our overview paper (Sauerbrei et al, 2020) we identified seven areas where evidence should be created by well-designed comparison studies. These areas include: (1) investigating properties of variable selection strategies, (2) comparing spline procedures, (3) analysing variables with a spike at zero, (4) comparing multivariable procedures for variable and function selection, (5) clarifying the role of shrinkage to correct for bias induced by data-driven decisions, (6) evaluating approaches to post- selection inference, and (7) adapting model building strategies to large sample sizes. In this talk we mention recent activities inside and outside STRATOS to address these issues. Research is ongoing in almost all of these seven areas, but it will need further neutral studies exploring the empirical properties of existing methods in a wider range of problems, and studies that are able to uncover situations where established methods may fail and clarify which assumptions of a method are crucial and which are non-critical. We encourage researchers to perform, reviewers to appreciate, and biostatistical journals to publish such carefully planned method evaluation studies that are indispensable to create the evidence for defining a state-of-the-art in multivariable modelling.

Heinze G, Boulesteix AL, Kammer M, Morris TP, White IR; Simulation Panel of the STRATOS initiative. Phases of methodological research in biostatistics-Building the evidence base for new methods. Biom J. 2023 Feb 3:e2200222. doi: 10.1002/bimj.202200222.
Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell FE Jr, Royston P, Heinze G; for TG2 of the STRATOS initiative. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. Diagn Progn Res. 2020 Apr 2;4:3. doi: 10.1186/s41512-020-00074-3

Mini Symposia 1

TMS1.5 How to include time-varying exposures prone to measurement error in survival analyses

Proust--Lima C.*[1], Philipps V.[1], Deffner V.[2], Boshuizen H.[3], Freedman L.[4], Thiébaut A.[5]
[1]Inserm, Univ. Bordeaux ~ Bordeaux ~ France, [2]Department of Statistics, Ludwig-Maximilians-Universität ~ Munich ~ Germany, [3]Division of Human Nutrition and Health, Wageningen University & Research ~ Wageningen ~ Netherlands, [4]Gertner Institute of Epidemiology and Health Policy Research, Sheba Medical Center ~ Tel Hashomer ~ Israel, [5]Université Paris-Saclay, UVSQ, Inserm, CESP, High Dimensional Biostatistics Team ~ Villejuif ~ France

Epidemiologic studies often rely on a cohort of non-diseased individuals and their prospective follow- up to assess the association between an exposure and an event. When time-varying, although the exposure may change continuously with time, its updates usually occur at discrete and possibly irregular visits/questionnaires, and is subject to measurement error [1,2].
As part of the "measurement error and misclassification" topic group of the STRATOS initiative (TG4), we reviewed and assessed the methods adopted in the literature to model the association between a time-varying error-prone exposure measured intermittently and a time-to-event. Five methods were identified:
(i) the last observation carried-forward (LOCF) method which assesses the instantaneous risk of event according to the most recent value of the exposure, considered as error-free.
(ii) the regression-calibration technique (RC) [3] which models the available exposure data (truncated at the event time) in a mixed model and then includes the continuous-time error-free individual prediction into the Cox model.
(iii) the regression-calibration technique for external exposure that would be also available after the event (external RC).
(iv) the multiple imputation method (MI) [4] which models the available exposure data in a mixed model incorporating time-to-event information, and then samples error-free exposure trajectories to be included into the Cox model.
(v) a joint modeling technique [5] which simultaneously models the exposure and the associated risk of event (JM).
In simulations exploring various scenarios of visit frequency, association intensity, error magnitude and baseline risk, we found that LOCF was biased in all scenarios, as was RC to a milder extent. External RC and MI showed bias in a few rather extreme scenarios while JM systematically retrieved the association with no bias. The methods were also illustrated with the association between body mass index and postmenopausal breast cancer from biennial questionnaires in the French E3N cohort.
The measurement error and the sparse updates of time-vaying exposures should be carefully handled when evaluating associations with health events. Joint models and multiple imputation techniques are two relevant options while RC should be avoided due to the informative truncation of the exposure at the event-time.

1. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 1982;69(2):331-42.
2. Andersen PK, Liestøl K. Attenuation caused by infrequently updated covariates in survival analysis. Biostatistics 2003;4(4):633-49.
3. Albert PS, Shih JH. On Estimating the Relationship between Longitudinal Measurements and Time-to-Event Data Using a Simple Two-Stage Procedure. Biometrics. 2010;66(3):983-7.
4. Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas SK, Peeters A, Wolfe R. Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM). Biostatistics 2018;19(4):479-96.
5. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data [Internet]. New-York: Chapman and Hall/CRC; 2012

TMS1.6 Evaluating the impact of covariate measurement error on functional form estimation in regression modelling

Perperoglou A.*[1], Abrahamowicz M.[2], Gustafson P.[3], Kipnis V.[4], Thiébaut A.[5], Ferreira Guerra S.[2], Freedman L.[6], On Behalf Of Tg2 And Tg4 O.T.S.I.[7]
[1]GSK ~ London ~ United Kingdom, [2]McGill University ~ Montreal ~ Canada, [3]University of British Columbia ~ Vancouver ~ Canada, [4]NIH ~ Rockville ~ United States of America, [5]French Institute of Health and Medical Research ~ Paris ~ France, [6]The Gertner Institute ~ Tel Hashomer ~ Israel, [7]NA ~ NA ~ Germany

Covariate measurement error is a common problem in regression modelling, often leading to biased parameter estimation and incorrect conclusions about variable relationships. The problem may be amplified when the true functional form between the covariate and the outcome is not linear. Further research is needed to comprehend how measurement error distorts the functional form of relationships, and how various adjustment methods can address this issue.
This collaborative project combines expertise of Topic Group 4 of the STRATOS initiative focusing on measurement error and misclassification, and Topic Group 2, specialising in variable selection and functional form in multivariable analysis. The primary objective is to evaluate and compare methods for estimating the true relationship between a dependent variable and a covariate when it is subject to measurement error. The work will be split between distinct teams that focus on Data Generation and Evaluation, Bayesian Methods, Imputation, and SIMEX. The Data Generation team will create multiple datasets with independent realizations of a binary outcome value Y, a continuous covariate X* with error. Data will be generated using undisclosed distributions of actual values X, a logistic regression model linking Y to X, and a classical measurement error model connecting error-prone X* to X, with various functional forms, sample sizes, variance and other parameters. Method teams will work independently. The Bayesian team will develop a model requiring prior specifications for outcome, measurement, and exposure modules. The Imputation team will use two measurement error adjustment methods: Regression Calibration and Multiple Imputation . The SIMEX team will evaluate six approaches, combining flexible curve fitting methods and SIMEX-based correction methods. The efficacy of these methods will be continually compared, with underperforming techniques potentially excluded from further simulations. The talk will outline goals and general methods, present preliminary simulation results to assess impact of measurement error on un-penalized regression spline and fractional polynomial estimates of the functional form in logistic regression. Ultimately, the study aims to yield significant insights into the functionality of diverse methods. conclusions drawn will greatly contribute to the advancement of more precise and resilient statistical methods in biostatistics and associated fields.

Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Küchenhoff H, Tooze JA, Wallace M, Kipnis V, Freedman L (2020): STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1 - Basic theory and simple methods ent. Statistics in Medicine. https://doi.org/10.1002/sim.8532
Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Keogh RH, Kipnis V, Tooze JA, Wallace M, Küchenhoff H, Freedman L (2020): STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 2 - More complex methods of adjustment and advanced topics. Statistics in Medicine. https://doi.org/10.1002/sim.8531
Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell Jr. FE, Royston P, Heinze G for TG2 of the STRATOS initiative (2020). State of the art in selection of variables and functional forms in multivariable analysis - outstanding issues. Diagnostic and Prognostic Research, 4:3, 1-18.

TMS1.7 Statistical analysis of high-dimensional biomedical data: a gentle introduction

Ambrogi F.*[2], Rahnenführer J.[1], De Bin R.[3], Mcshane L.[4]
[1]Department of Statistics, TU Dortmund University ~ Dortmund ~ Germany, [2]Department of Clinical Sciences and Community Health, University of Milan ~ Milano ~ Italy, [3]Department of Mathematics, University of Oslo ~ Oslo ~ Norway, [4]Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute ~ Bethesda ~ United States of America

The goal of the High-dimensional Data (HDD) Topic Group of the STRATOS initiative (TG9) is to provide guidance amid the jungle of opportunities and pitfalls inherent in the analysis of high- dimensional biological and medical data. Methods for analysis of HDD are rapidly changing, and researchers across different fields, including biostatistics, bioinformatics, and bioengineering, contribute to their development. Advances in statistical methodology and machine learning methods have contributed to improved approaches for data mining, statistical inference, and prediction in HDD settings; however, adoption of these methods has sometimes gotten ahead of understanding of their proper application. The mission of TG9 includes identification of fundamental principles for analysis of HDD, explanation of available methods, and development of broadly accessible guidance on best practices in this complex and changing landscape. In this talk we present the first published work of TG9 [1]: a comprehensive review aiming to provide a solid statistical foundation for researchers, including statisticians and non-statisticians, who are new to research with HDD or simply want to better evaluate and understand the results of HDD analyses. Following that, we will describe new research topics currently being pursued by TG9. Specifically, guidance materials specific to HDD for sample size calculation, influence and choice of tuning parameters in machine learning applications, and use of plasmode data for simulations are under development. Simulation studies are especially challenging for HDD yet they are essential tools needed to perform evaluation and comparison of different methods. Proliferation of high-dimensional data in biomedical research has brought unprecedented opportunities to advance knowledge. In order to harness the power of the rapidly evolving repertoire of analysis methods to reveal useful insights from HDD, it is imperative that researchers have access to guidance on the methods available and their proper application.

[1] Rahnenführer, J., De Bin, R., Benner, A., Ambrogi, F., Lusa, L., Boulesteix, A.L., Migliavacca, E., Binder, H., Michiels, S., Sauerbrei, W. & McShane, L. Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. BMC Med 21, 182 (2023). https://doi.org/10.1186/s12916-023-02858-y;

Mini Symposia 1

TMS1.8 The slowly changing landscape of predictive modeling in biomedicine

Lusa L.*[1] , Kappenberg F.[2], Schmid M.[3], Sauerbrei W.[4], Rahnenführer -- For The Stratos Initiative J.[2]
[1]Department of Mathematics, Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska and Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia ~ Koper/, [2]Department of Statistics, TU Dortmund University ~ Dortmund ~ Germany, [3]Institute for Medical Biometry, Informatics and Epidemiology, Medical Faculty, University of Bonn ~ Bonn ~ Germany, [4]Institute of Medical Biometry and Statistics, Medical Center - University of Freiburg ~ Freiburg ~ Germany

The number of predictive models proposed in the biomedical literature is growing every year. In the last few years there has been an increasing attention to the changes occurring in the predictive modeling landscape, which are mostly related to the methods and data that are being used. Many influential editorials suggested that machine learning techniques are becoming more and more popular and that models are more often developed using complex data. Because of these changes they suggested that the existing best practice recommendations for design, conduct, analysis, reporting, impact assessment, and clinical implementation from the traditional biostatistics and medical statistics literature are no longer sufficient to guide the use of predictive models. We address the problem of quantifying these changes in practice, re-analyzing the data from a selection of systematic reviews on predictive models in biomedical fields. We re-analyzed the data from selected reviews published since 2020, reviewed at least 30 papers reporting developmental prognostic models, and per paper/per model data available for most of the information suggested in the CHARMS checklist[1]. We identified 8 reviews that analyzed 1499 predictive models published in 841 papers. Time trends were analyzed using the models published since 2005. The average number of patients and events increased greatly in the 2015-19 period, but remained rather stable afterwards. The largest studies were conducted mostly in the most recent years. The number of candidate and used variables did not increase dramatically, only few recent studies used very large numbers of variables. The use of internal validation and reporting of discrimination measures became more common, but external validation and calibration are rarely used. Information about missing values is still not reported in about half of the papers, imputation methods are becoming more common. The use of machine learning methods did not increase systematically in all reviews. Overall, most of the findings were heterogenous across reviews. Our findings indicate that changes in the predictive modeling landscape are less dramatic than expected and that poor reporting is still common; adherence to well established best practice recommendations from the traditional biostatistics literature is still needed. [1] K.G. Moons, JAH de Groot, W. Bouwmeester, Y. Vergouwe, S. Mallett, D.G. Altman, J.B. Reitsma, G.S. Collins, PLoS medicine, Vol11, Num10, 2014, e1001744.

Mini Symposia 1

TMS1.9 Counterfactual prediction for personalized healthcare using observational data

Van Geloven N.*[1], Steyerberg E.[1], Wang J.[2], Didelez V.[3], Keogh R.[4]
[1]Leiden University Medical Center ~ Leiden ~ Netherlands, [2]Utrecht University ~ Utrecht ~ Netherlands, [3]Leibniz Institute for Prevention Research and Epidemiology - BIPS ~ Bremen ~ Germany, [4]London School of Hygiene and Tropical Medicine ~ London ~ United Kingdom

Members of Stratos topic groups 6 and 7 (plus 4 and 8) recently joined forces in an intensive week working on counterfactual prediction, or prediction under (hypothetical) interventions. Predictions under interventions provide estimates of a person's risk of an outcome if they were to follow a particular treatment strategy, taking into account other patient characteristics that are predictive of the outcome. Estimating and validating interventional predictions in randomized controlled trials is usually limited by small sample sizes. Using observational data for this task requires expansion of both prediction and causal inference methods. In this talk we describe the methodological gaps identified at the workshop and new projects that arose from it. Prediction under interventions is a promising yet challenging area of new research. We will elicit input from the other TG's regarding further challenges that may be expected when estimating interventional predictions from different randomized or observational data sources.

TMS1.10 Developing recommendations to handle patient reported outcome data in oncology cancer trials: sisaqol-imi

Le Cessie S.*[1], Goetghebeur E.[2], Thomassen D.[1]
[1]Leiden University Medical Center ~ Leiden ~ Netherlands, [2]University of Ghent ~ Gent ~ Belgium

This talk is on behalf of work package 3 of the SISAQOL-IMI consortium. SISAQOL-IMI is an international project, led by the European Organisation for Research and Treatment of Cancer (EORTC) and Boehringer Ingelheim. The aim of this four year project is to establish international standards in the analysis of patient reported outcomes (PRO) and health-related quality of life data in cancer clinical trials.
This is done by seeking consensus internationally and across stakeholders (industry, academics, patients, trial organizations, regulators, etc) . STRATOS is involved in this large project, in particular in the work package on single arm trials and other non randomized studies. We are now halfway through this project and currently a set of suggested recommendation for PROs in single arm oncology studies is piloted on case studies. arm oncology studies is piloted on case studies. The recommendations consider the aims of PROs in single arm studies, formulation of research questions in the context of the ICH E9(R1) estimand framework (1), choosing appropriate outcome measures, addressing the absence of a randomised control group, and handling intercurrent events and missing data. In this talk we will give an update on the current status of our work and discuss questions and challenges in the design and analysis of single arm PRO studies.

1. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

TMS1.11 Comparing quality of life while alive between treatment and controls: real world analysis in trials

Goetghebeur E.*[1], Reynders D.[1], Thomassen D.[2], Le Cessie S.[2]
[1][1]Ghent University ~ Ghent ~ Belgium, [2]Leiden University Medical Center ~ Leiden ~ Netherlands

The EU IMI_SISAQOL project
https://www.imi.europa.eu/projects-results/project-factsheets/sisaqol-imi
develops guidelines for evaluating the effect of novel treatments on PROMs in oncology populations. STRATOS helps develop statistical guidelines as a partner in this project. When mortality is high, we consider `time to death' in combination with `Quality of life (QOL)- while- alive' as the bivariate real-world outcome. From a statistical perspective, comparing treatments in terms of this outcome is wrought with major challenges. First, death is a terminal event for any patient-reported outcome, and especially for quality of life. While obvious, implications for statistical analysis and interpretation are often ignored. This happens for instance, when parameters in mixed models for QOL measures taken repeatedly over time get a direct interpretation which implicitly imputes QOL-after-death. Averaging derived predictions among the living at a given time t is then warranted but GEE models target the real world estimand more directly. Second, in single arm studies as in RCTs, the question arises `in which populations to compare QOL(t*) (at a chosen time t*) with and without treatment'. To this end, we identify a control group with overlapping entry and exclusion criteria where a common `core set' of confounder variables is measured at baseline t0, and the positivity assumption holds. The combined propensity of treatment at t0 and propensity of censoring avoidance at t* enable adjustment for measured confounders and explainable censoring, respectively. An extra between study variation on the standardized QOL(t*), then yields a `real world' Confidence Interval for the population Average Treatment Effect (among the Treated), The final challenge arises when at disease progression or change of treatment one stops measuring QOL. Besides death as an intercurrent event, and administrative censoring there are missing QOL data, with missingness likely dependent on time to subsequent death. We argue in favour of a well understood real world analysis and present challenges and solutions in the setting of a single arm study seeking an external control group. We illustrate how the three challenges can be handled jointly in our case study.

Brenda F. Kurland, Laura L. Johnson, Brian L. Egleston and Paula H. Diehr Longitudinal Data with Follow-up. Truncated by Death: Match the Analysis Statistical Science 2009, Vol. 24, No. 2, 211–222 DOI: 10.1214/09-STS293

Mini Symposia 2

NOVEL APPROACHES TO COMPLEX DATA AND PREDICTIVE MODELING IN HEALTHCARE RESEARCH

COORDINATORS: EMANUELE DIANGELANTONIO AND FRANCESCA IEVA

TMS2.1 Predicting emergency admissions in scotland

Vallejos C.A.*
University of Edinburgh ~ Edinburgh ~ United Kingdom

Emergency admissions (EA), where a patient requires urgent in-hospital care, are a major burden for individuals and healthcare systems. The development of risk prediction models can alleviate this problem by supporting primary care interventions and public health planning. We introduce SPARRAv4, a predictive score for EA risk that will be deployed nationwide in Scotland. SPARRAv4 was derived using supervised and unsupervised machine-learning methods applied to routinely collected electronic health records from approximately 4.8M Scottish residents. We demonstrate improvements in discrimination and calibration with respect previous scores deployed in Scotland, as well as stability over a 3-year timeframe. Our analysis also provides insights about the epidemiology of EA risk in Scotland, by studying predictive performance across different population sub-groups and types of admission, as well as by quantifying the effect of individual input features. Finally, we discuss broader challenges including reproducibility and how to safely update risk prediction models that are already deployed at population level. SPARRAv4 represents a real-world use of machine learning through a risk score that fitted and deployed at national level, and widely available in clinical settings. In an era where healthcare systems are under high stress, we expect that the availability of robust and reproducible risk prediction scores such as SPARRAv4 will contribute to the design of proactive interventions that reduce pressures on healthcare systems and improve healthy life expectancy.

Liley et al (2021) medRxiv. doi: 10.1101/2021.08.06.21261593

TMS2.2 Time: the next frontier in machine learning for healthcare

Mihaela van der Schaar
University of Cambridge

In this talk, I aim to illuminate the underemphasized yet critical dimension in machine learning for healthcare as well as biostatistics: time. I contend that time harbors the potential to revolutionize machine learning methodologies, particularly within healthcare. This presentation underscores the opportunities and challenges that emerge from integrating temporal dynamics into machine learning models, enriching prediction accuracy, inference robustness, causal inference and conceptual understanding.

TMS2.3 Rage against the machine learning

Marteen van Smeden
University Medical Center Utrecht

Medicine appears to be at the start of a new era due to the many promising developments in machine learning (ML). Part of that promise is that ML will improve upon traditional statistical modelling. Another promise is that ML will outperform medical doctors. In this talk I will highlight a few medical settings where ML seems to have lived up to its promises, and some where it has not. A couple of challenges for fair comparisons between machine learning, traditional statistical modelling and doctors will be identified and discussed. Finally, I will speculate on the future role of machine learning, traditional statistical modelling and medical doctors in algorithm based medicine.

TMS2.4 Regression and ml approaches for evaluation of biomarkers with application to primary biliary cholangitis

Bernasconi D.P.*[1], Gerussi A.[2], Valsecchi M.G.[1]
[1]Bicocca Bioinformatics Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milano- Bicocca ~ Monza ~ Italy, [2]Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca ~ Monza ~ Italy

The discovery and validation of new biomarkers that have the potential to improve outcome prediction and personalized therapeutic approaches is now an important aspect in medicine. Prediction and validation need novel statistical approaches that may also rely on intensive computational methods, machine learning and artificial intelligence. In primary biliary cholangitis (PBC), a rare female- predominant chronic autoimmune liver disease, patients are treated with Ursodeoxycholic acid (UDCA) to slow down disease progression. The identification of biomarkers associated with treatment response and disease progression is important to target the need of additional treatments. Two clinical risk scores, based on bilirubin and alkaline phosphatase (ALP), are currently used to identify patients at higher risk of liver transplant or death after 12 months of UDCA. Our aim is to explore the role of potential new markers to define prognosis using both statistical methods and an unsupervised machine learning approach. We analysed data from a large European registry on PBC with 2129 patients to explore the role of gamma-glutamyl transferase (GGT) as a potential new marker for liver-transplant-free survival. Using classical survival analysis techniques and the time-dependent ROC curve methodology, we showed that GGT adds discriminatory power to the previous risk scores and, with proper cut-off values, improves ALP-based prognostic stratification. In a second study, we pooled data from three broad PBC registries (Italian, Japanese and western populations) and we applied unsupervised k-medians clustering, a fully data-driven machine learning (ML) algorithm, to search for new potential biochemical markers supporting disease sub-phenotyping and risk stratification. The sample of more than 12,000 was used to derive the model and to validate it with respect to the observed liver-transplant-free-survival and to the existing clinical scores. Change of albumin from diagnosis to 1 year of treatment with UDCA was identified as a novel useful biomarker for disease prognostication. In the two studies presented, statistical survival analysis techniques and data-driven ML algorithms were used to improve risk stratification and to generate hypotheses on potential new markers in patients affected by PBC.

[1] Gerussi A, Bernasconi DP, O'Donnel SE et al. Measurement of Gamma Glutamyl Transferase to Determine Risk of Liver Transplantation or Death in Patients With Primary Biliary Cholangitis, Clinical Gastroenterology and Hepatology 2021;19:1688–1697
[2] Gerussi A, Verda D, Bernasconi DP, et al. Machine learning in primary biliary cholangitis: a novel approach for risk stratification, Liver International. 2022;42:615–627

ECB

EARLY CAREER BIOSTATISTICIANS (ECB)

COORDINATORS: EARLY CAREER BIOSTATICIAN COMMITTEE

ECB.1 Sustaining a culture of reproducibility in research: a personal credo for early career biostatisticians

Edefonti V.*
Branch of Medical Statistics, Biometry, and Epidemiology "G. A. Maccacaro", Department of Clinical Sciences and Community Health, Università degli Studi di Milano and Fondazione IRCSS Ca' Granda Ospedale Maggiore Policlinico ~ Milano ~ Italy

Full replication remains the ultimate standard by which we evaluate scientific claims. Computational science has led to exciting developments, but the nature of the work has revealed limitations in our ability to evaluate published findings. Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible [1]. I take the chance of using the Early Career Biostatistician Day to share some personal experiences grappling with how to operationalize reproducibility while balancing its demands against other priorities. First, I will introduce the replication crisis in science and have a look to its causes. Second, I will revise the terminology, including the main concepts of replication and reproducibility and put them in the context of the different research fields. Third, while discussing remedies to the general replication crisis, I will deal with reproducibility in computational science, providing examples from my major field of expertise, epidemiology, where it is often impossible to fully replicate findings. Finally, I will give practical advice on how to make research reproducible, including publishing a cleaned-up version of the code along with the data sets in a durable non-proprietary format, within the software environment necessary to execute the code. I will explore literate statistical analysis tools to publish data analyses in a single document, a compendium, that allows others to easily execute the same analysis to obtain the same results. Reproducibility is an ideal that no researcher would dispute "in the abstract", but when aspirations meet the hard reality of the academia, reproducibility often "loses out". I will try to show that it is advantageous for early career biostatisticians to dedicate time and effort to developing a culture of reproducibility. Simply bringing the notion of reproducibility to the forefront and making it routine will make a difference.

[1] R.D. Peng, Science, 334 (6060), 2011, 1226-1227.

ECB.2 A network of mentors: leveraging the power of networking and mentoring to accelerate your career

Wjesuriya R.*
Clinical Epidemiology & Biostatistics (CEBU), Murdoch Children's Research Institute ~ Melbourne ~ Australia

Mentoring and networking are both very powerful forms of social learning. Both can help us tap into shared knowledge and obtain support and opportunities, thus providing us with the necessary tools and confidence to advance our careers. While we all engage in networking and mentor-mentee relationships at different stages of our careers, we rarely put dedicated and conscious effort to combine the two to build a "network" of mentors and reap the compounded benefits it can offer. In this talk I will share my personal quest of building a network of mentors early in my career, and the lessons I have learnt in seeking and maintaining fruitful mentor-mentee relationships. Seeking and maintaining a network of mentors provides multiple opportunities, and different perspectives and is a valuable investment of time for accelerating a scientific career.

ECB.3 Unreported rct results: should biostatisticians care?

Ter Schure J.*
Department of Epidemiology & Data Science, Amsterdam UMC ~ Amsterdam ~ Netherlands

Roughly 50% of Randomized Clinical trials (RCTs) never report results [1]. This leads to publication bias and wastes research effort, funding, patient trust, and other resources. Should biostatisticians care? And what can they do about this?
Many biostatisticians work as consultants and ethics committee members, influencing RCT design. Information from similar RCTs can help improve new RCTs, e.g. summary information on recruitment rates, event incidence, or background variation. So biostatisticians should care about complete registration and reporting of completed and ongoing RCTs. Trough their involvement in RCT design, they might also encourage or even enforce complete reporting in trial registries as well as open publications.
Can we improve reporting by embedding existing publications, registry reporting, or protocol registrations in our consultancy?

[1] TranspariMED 2020 CLINICAL TRIAL TRANSPARENCY IN THE NETHERLANDS https://www.transparimed.org/_files/ugd/01f35d_7f02a5ffd53c4429bc21e4d7b17dddа0.pdf?index=true

ECB.4 Are statistical and scientific assessments of rapid self-test diagnostics reliable?

Hillier B.*, Baldwin S., Scandrett K., Agarwal R., Davenport C., Deeks J.
University of Birmingham ~ Birmingham ~ United Kingdom

Following the widespread use of rapid Covid self-tests, there has been a recent emergence of self-test diagnostics for multiple health conditions. Self-tests increase access and speed of testing, and can make users feel as though they are gaining control over their own health. However, self-tests may cause unintended harm: their performance may be inadequate; they may be used or interpreted inappropriately; or recommended health actions following testing may not be understood or adhered to. We aim to survey self-tests currently available, evaluate their claims, the evidence supporting their performance, and consider whether potential harms in using these tests are a concern to the public. We surveyed high-street supermarkets and chemists selling self-tests, within 10 miles of the University of Birmingham. Test kits included home diagnostics for HIV, cancer, nutritional deficiencies, menopause and sperm mobility. We purchased test kits, then analysed statements on the packaging and Instructions For Use (IFU) documents. We requested Clinical Study Reports (CSRs) from distributors to analyse the evidence used in obtaining regulatory approvals, and assessed the validity of statements on statistical performance, study design, methods and reporting. We systematically assessed the statistical and clinical claims on the packaging, IFUs and CSRs, covering details of: the intended use of tests (in whom and when should tests be performed); the benefits and harms of using tests; statistical claims; the interpretation of positive and negative results; and associated health advice. Early findings show that most tests claim to be 95%-99% accurate. Most CSRs are based on undescribed samples with no details on setting or patient characteristics. Test accuracy claims are primarily determined by laboratory evaluation rather than by their performance in at-home settings. Few CSRs compared tests with appropriate reference standards or with alternatives such as routine hospital laboratory tests. Statements on accuracy were made without consideration of prevalence, and without statements on positive and negative predictive values. The scientific and statistical claims on self-tests are potentially misleading to the public. The current regulations on medical devices seem inadequate. It is important for public health to ensure that statistical standards are implemented when evaluating tests.

1. Royal Statistical Society. Diagnostic Tests Working Group report. Jun 2021. https://rss.org.uk/RSS/media/File-library/Policy/2021/RSS-Diagnostic-tests-report-FINAL.pdf.

ECB.5 Challenges in extracting and processing ehr data for dynamic prediction models

Albu E.*
Department of development and Regeneration, KU Leuven , Belgium ~ Leuven ~ Belgium

EHR (Electronic Health Record) data are routinely collected with the purpose of supporting the healthcare processes; they follow the clinical workflow allowing doctors to make decisions during the patient follow-up. While these data are not purposely collected for research, careful extraction and pre- processing can deem EHR data useful for various research questions.
We are currently conducting research for dynamic prediction of CLABSI (Central-Line Associated Bloodstream Infections) based on rich hospital-wide EHR data extracted from the university hospital in Leuven, Belgium (UZ Leuven). We have extracted data regarding medical devices (catheters), laboratory results, medication, comorbidities, care items and administrative patient data from different software systems for both ICU (Intensive Care Units) and hospital wards and faced challenges on both the extraction process and the pre-processing steps to bring the data in a format appropriate for prediction modelling. The "lessons learned" during this process might prove useful to other researchers conducting their work on similar data sources. Close collaboration with the hospital IT team responsible for data extraction allowed us to get a deep understanding of the extracted items, the healthcare processes that generated the data, the missingness patterns due to extraction or registration patterns. We contributed to improving the quality of the data extraction process resulting in better completeness of research data and better documentation. Pre-processing the extracted data for dynamic prediction modelling differs from cross-sectional prediction in terms of: defining time steps for prediction; defining the start and end time of real-time prediction; meaningful prediction update times. Moreover, it differs from research for inferential studies in terms of clinically relevant concepts and clinically relevant timestamps, harmonization of clinical concepts between data sources, reproducibility and code efficiency. There is no standard to date for extracting and pre-processing EHR data for research, and we argue that there should be no standard, but a set of guidelines. The key guidelines we propose are: (1) simplified; easy to understand; well documented; complete and qualitative data extraction and (2) time efficient and reproducible pre-processing workflow with appropriate considerations for the research question.

ECB.6 Early career biostatisticians

Hunt A.*
University of Liverpool ~ Liverpool ~ United Kingdom

I am 6 months into my career as a research associate and 4 months (as of March 2023) into my part time PhD. My career as a Biostatistician is very early however, I have learnt a lot of useful skills and life hacks within my short time. My presentation focuses on our mental health, dealing with the ups and down that comes with the job and imposter syndrome. I currently work on three projects whilst completing my PhD part time and this requires a lot of time management to ensure that I complete the tasks needed to meet my deadlines. I have adapted my working day around different techniques to keep me motivated, on target and most importantly off my phone. Office research can often lead to procrastination a lot easier than a research job in a lab, or in lecture theatre, therefore, I felt it was important to adopt a schedule for my days at my desk. With this came different methods of writing, reading and learning skills that work best with my personality. Balancing a PhD with work can sometimes be difficult, however, speaking openly with my supervisors and delegating my workload has taught me to feel comfortable and confident with the research I produce. A huge hurdle for me as an early career biostatistician was imposter syndrome. My life very quickly moved from learning a Masters in Health Data Science, whereby I would be in university 5 days a week learning from Professors and Doctors who are extremely competent and knowledgeable in their subject areas. To becoming a researcher like them. It was a big challenge to train myself to believe that I was good enough and did have the knowledge and skillsets to complete my jobs independently. I did this through a variety of methods; reading material, chatting to my peers and using the resources provided by my University.

Early Career Biostatisticians (ECB)

ECB.7 Tackling imposter syndrome

O"Donnell A.*
University College Cork ~ Cork ~ Ireland

Imposter syndrome is one of the most frequently mentioned topics when it comes to PhDs and early career researchers. First described in 1978 [1] the imposter phenomenon is something that affects many people at various stages in their career journey and is discussed widely on social media platforms. In the medical profession alone it is estimated that approximately 50% have experienced the impostor phenomenon [2]. Often, at conferences, a more senior researcher will give a talk on how they overcame their imposter syndrome, or other times, how it hasn't left them even late in their career. What is rarely mentioned, however, is how we can actually tackle our imposter syndrome, right now. As an individual who has taken several large career shifts in the short period of time since completing my undergraduate degree I have had a lot of first-hand experience with imposter syndrome and have developed some useful techniques to help with the feeling as well as discovering what definitely does not work for me. In this talk, I will discuss how I deal with imposter syndrome with respect to my research, teaching and professional career and beyond.

[1] P. R. Clance and S. A. Imes, "The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention.," Psychotherapy: Theory, Research & Practice, vol. 15, no. 3, pp. 241-247, 1978.
[2] M. Thomas and S. Bigatti, "Perfectionism, impostor phenomenon, and mental health in medicine: a literature review," Int J Med Educ, vol. 11, no. 1, pp. 201-213, 2020.

ECB.8 Navigating the world of biostatistics

Lee K.*
Murdoch Children's Reseach Institute and University of Melbourne ~ Melbourne ~ Australia

Careers in biostatistics can take a variety of different forms, and typically consist of some combination of collaborating on applied research studies, conducting methodological research, and teaching. These three activities are critical to sustaining this vital discipline that underpins medical and public health research. A successful career as a biostatistician is underpinned by developing strong research collaborations, both with applied researcher teams and with other biostatisticians, as it is the sharing and discussion of ideas that leads to the best research. In this talk I will take the audience on my journey through my career in biostatistics, and provide some insight into the decisions and activities that have helped me along the way. Importantly, I describe the collaborations that I have established which have led me to where I am.

ISCB44

44th Annual Conference of the International Society for Clinical Biostatistics

Joint conference with the Italian Region of the International Biometric Society

MILAN, ITALY | 27 – 31 AUGUST 2023

University of Milano-Bicocca
Building U6 Piazza dell'Ateneo Nuovo 1

Organising Secretariat
Promoest srl, Via G. Moroni 33, 20146 Milano (MI)
iscb2023@promoest.com


UNIVERSITÀ DEGLI STUDI MILANO BICOCCA
ISCB44
MILAN, ITALY | 2023