

Pseudo observations in survival analysis

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.

Instructors:

Per Kragh Andersen, professor, Biostatistics, University of Copenhagen, Denmark

Henrik Ravn, senior statistical director, Novo-Nordisk A/S, Denmark