Andersen PK, Perme MP, van Houwelingen HC, Cook RJ, Joly P, Martinussen T, Taylor JMG, Abrahamowicz M, Therneau TM for the STRATOS TG8 topic group (2020): Analysis of time-to-event for observational studies: Guidance to the use of intensity models. Statistics in Medicine. DOI:10.1002/sim.8757

Lay Abstract

This paper provides guidance for researchers with some mathematical back- ground on the conduct of time-to-event analysis in observational studies. Time- to-event data appear frequently in observational and clinical studies when sub- jects are followed over time for the occurrence of an event of interest, such as death, disease incidence, or disease recurrence. Such data are characterized by incomplete observation because, at the end of data collection, the event will not have occurred for all subjects and, thus, for some subjects the only information available about event occurrence is that it did not occur during the time interval over which the subject was observed. This is known as censoring. A very useful quantity in survival analysis, that is in the focus of our paper, is the intensity (or hazard) function. It answers the question: what is the probability - given the past information available till today - of the event occurring tomorrow. The paper discusses basic concepts like time axis, event definition and censoring. Intensity models are introduced, with special emphasis on the Cox proportional hazards regression model. Check lists are provided that may be useful both when fitting the model and assessing its goodness of fit and when interpreting the results. Special attention is paid to how to avoid problems with so-called immortal time bias. This bias occurs when, erroneously, the exposure status of a subject today is characterized based on information that only appears later and it may be avoided by properly addressing the time order in which information unfolds. Technically, this is done by introducing explanatory variables that evolve over time (time-dependent covariates). Prediction based on hazard models and difficulties when attempting to draw proper causal conclusions from such models are discussed. Finally, a series of examples are presented where the methods and check lists are exemplified. Computational details and implementation using the freely available R software are documented in Supplementary Material. The paper was prepared as part of the STRATOS initiative.