



UMEÅ UNIVERSITET

Estimation of hazard ratios from observational data with applications related to stroke

Guilherme Wang de Faria Barros

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt försvar i Hörsal NBET.A.101, Norra Beteendevetarhuset, fredagen den 2 februari, kl. 10:00.

Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Associate Professor Daniel Nevo, Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel.

Department of Statistics, USBE

Organization

Umeå University
Department of Statistics
Umeå School of Business,
Economics and Statistics

Document type

Doctoral thesis

Date of publication

09 January 2024

Author

Guilherme Wang de Faria Barros

Title

Estimation of hazard ratios from observational data with applications related to stroke

Abstract

The objective of this thesis is to examine some challenges that may emerge when conducting time-to-event studies based on observational data. Time-to-event (also called survival) is a setting that involves analyzing how different factors may influence the length of time until an individual experiences the event of interest. This type of analysis is commonly applied in fields such as medical research and epidemiology. In this thesis, which focuses on stroke, we are interested in the time to a recurrent stroke or the death of a patient who survived a first stroke.

Hazard ratios are one of the main parameters estimated in time-to-event studies. Hazard ratios involve comparing the risk of experiencing the event between two groups, usually a treated group and an untreated group. They can also involve other factors, such as different age groups. Hazard ratios can be estimated from the data by using the Cox regression model.

Observational data, in contrast to experimental data, involves data collected without any intervention or random assignment of treatment to the individuals. Confounders, that is, variables that distort or obscure the true relationship between treatment and outcome, are always present and need to be controlled for. National registers are an important source of observational data. A national registry is a centralized database or system that collects, stores, and maintains information about a specific population or group of individuals within a country. Sweden is known for its detailed and complete national registers. In this thesis, data from the Swedish Stroke Register (Riksstroke) is used to study factors related to stroke.

In time-to-event studies involving observational data, several challenges may arise for the researcher during data analysis. Some individuals may not experience the event during the observation period and thus the information about their time until the event is incomplete. These individuals are considered as censored. Some individuals may experience another event rather than the one of interest, a competing risk. Additionally, models must be properly constructed, with researchers selecting variables and determining the suitable functional form.

Four papers are included in the thesis. Paper I demonstrates how to handle competing risks in survival analysis. The study involves comparing individuals with and without standard modifiable risk factors and their risks of a recurrent stroke or death using data from the Swedish Stroke Register. The estimation of marginal hazard ratios is a common theme in the other three papers. All involve simulation studies in order to extend methods and explore best practices when estimating marginal hazard ratios. Paper II explores non-parametric methods that can be used as alternatives to more traditional parametric methods when balancing datasets in order to estimate a marginal hazard ratio. A case study was also conducted using data from the Swedish Stroke Register involving the prescription of anticoagulants at hospital discharge after a stroke.

Paper III is about how censoring affects marginal hazard ratio estimation, even with perfect balancing of the dataset. We study this issue, taking into consideration varying effect sizes and censoring rates. A procedure to attenuate the problem is also studied. Paper IV concerns covariate selection in the case of high-dimensional data. High-dimensional data involves cases in which the number of covariates in the study is comparable to the number of individuals, and therefore covariate selection methods are needed. In the paper, we explore some of these methods and suggest a best-performing procedure. As Paper II, Paper IV involves a case study of anticoagulant prescription using data from the Swedish Stroke Register.

Keywords

survival analysis, causal inference, hazard ratios, marginal hazard ratio, stroke, balancing

Language

English

ISBN

print: 978-91-8070-240-9
pdf: 978-91-8070-241-6

ISSN

1100-8989

Number of pages

19 + 4 papers