Statistical methods for register based studies with applications to stroke

Anita Lindmark

Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för avläggande av filosofie doktorsexamen framläggs till offentligt förvar i Hörsal E, Humanisthuset, fredagen den 21 oktober, kl. 9:30. Avhandlingen kommer att förvaras på engelska.

Fakultetsopponent: Professor, Saskia le Cessie, Department of Clinical Epidemiology, Leiden University Medical Center.
Abstract
This thesis adds to the area of register based research, with a particular focus on health care quality and (in)quality. Contributions are made to the areas of hospital performance benchmarking, mediation analysis, and regression when the outcome variable is limited, with applications related to Riksstroke (the Swedish stroke register).

An important part of quality assurance is to identify, follow up, and understand the mechanisms of inequalities in outcome and/or care between different population groups. The first paper of the thesis uses Riksstroke data to investigate socioeconomic differences in survival during different time periods after stroke. The second paper focuses on differences in performance between hospitals, illustrating the diagnostic properties of a method for benchmarking hospital performance and highlighting the importance of balancing clinical relevance and the statistical evidence level used.

Understanding the mechanisms behind observed differences is a complicated but important issue. In mediation analysis the goal is to investigate the causal mechanisms behind an effect by decomposing it into direct and indirect components. Estimation of direct and indirect effects relies on untestable assumptions and a mediation analysis should be accompanied by an analysis of how sensitive the results are to violations of these assumptions. The third paper proposes a sensitivity analysis method for mediation analysis based on binary probit regression. This is then applied to a mediation study based on Riksstroke data.

Data registration is not always complete and sometimes data on a variable are unavailable above or below some value. This is referred to as censoring or truncation, depending on the extent to which data are missing. The final two papers of the thesis are concerned with the estimation of linear regression models for limited outcome variables. The fourth paper presents a software implementation of three semi-parametric estimators of truncated linear regression models. The fifth paper extends the sensitivity analysis method proposed in the third paper to continuous outcomes and mediators, and situations where the outcome is truncated or censored.

Keywords
Registers, quality of care, socioeconomic status, hospital performance, stroke, mediation, sensitivity analysis, truncation, censoring