Karolinska Institutet

Department of Medical Epidemiology and Biostatistics

Statistical methods for long-term follow-up of infectious diseases.

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Petrén, Nobels väg 12b, Karolinska Institutet, Solna

Fredagen den 14 mars, 2014, kl. 09.30

av
Anna Törner
MSc

Huvudhandledare:
Professor Paul Dickman
Karolinska Institutet
Institutionen för medicinsk epidemiologi och biostatistik MEB, Solna

Bihandledare:
Professor Åke Svensson
Stockholm Universitet
Matematistka Institutionen

Ann-Sofi Duberg, MD PhD
Universitetssjukhuset Örebro
Infektionskliniken

Fakultetsopponent:
Professor of Hygiene and Epidemiology
University of Athens Medical School and
Adjunct Professor of Epidemiology,
Harvard School of Public Health, Boston

Betygsnämnd:
Professor Lars Alfredsson
Karolinska Institutet, IMM, Solna

Docent Fredrik Granath
Karolinska Institutet
Institutionen för medicin (MedS), Solna

Docent Katja Fall
Universitetssjukhuset Örebro
Center för klinisk epidemiologi och biostatistik (KEB)

Stockholm 2014

Abstract

The overall aim of this work has been to investigate methodological issues connected to long-term follow-up of infections diseases. The work extended to prevalent cohorts in general. The common denominator for the main methodological efforts in these four papers is issues connected to selection bias. In the first three papers methods for visualizing selection bias in prevalent cohorts were explored and different approaches to adjust for this bias discussed. In the fourth paper, capture-recapture modeling was used to examine ascertainment level for liver cancer in the Swedish Cancer Register.

Study 1: In this study we investigated a novel approach to visualize and adjust for selection bias in prevalent cohorts. The method is an extension of the standard interval-based approach, where a risk estimate is calculated for disjointed time periods after inclusion in the cohort of interest. In the proposed method, observation time and events are cumulated, giving more power and more precise estimates which may be useful for studies with few events where it may be difficult to judge what is a true effect and what is random noise. The proposed method, cumulative SIR, is exemplified using data on hepatitis-C virus infection and the outcome liver cancer and non-Hodgkin lymphoma. The results using this novel approach were comparable to a standard approach with disjoint intervals. The results indicate that the method may be useful in situations with few events in the cohort. The method is only useful for cohorts where the risk of the studied outcome is fairly stable over time.

Study 2: Spurious observations have indicated that there may be a relationship between hepatitis C virus (HCV) infection and kidney cancer. In this study the relationship between HCV-infection and kidney cancer was investigated by use of disease registers. In addition the known association of HCV-infection and other forms of kidney disease was explored further. Methods for investigating selection bias explored in Paper I were used, in addition new ideas were investigated which were further developed in paper III. The relationship between HCV-infection and kidney cancer was not confirmed in this study, but the association of HCV-infection with other kidney-related diseases was investigated further.

Study 3: For cohorts that may have high hazard immediately after inclusion in the cohort, which then first decreases to later increase with follow-up time, the method of cumulative SIR must not be used. The cumulative properties will obscure the initial decrease and the method cannot give clear answers. In paper III we used restricted cubic splines to model to instantaneous failure rate (hazard). The shape of the hazard function may give an indication of the possible presence of selection bias in the cohort. The proposed method was exemplified using 1) data on HCV-infection where the outcome of interest was ‘kidney disease’ and 2) a cohort a patients with Monoclonal Gammapathy of Uncertain Significance (MGUS) and the outcome of interest ‘death’. The model was useful to study the shape of the hazard in the cohorts and the number of knots was adjusted to give a suitably flexible model, clearly showing the shape of the hazard without being too flexible.

Study 4: In this study we explored capture-recapture modeling, using a log-linear model to estimate ascertainment level of the Swedish Cancer Register (CR). We used a three-source model: CR, the National Patient Register (PR) and the Cause of Death Register (DR). Due to the limited degrees of freedom in available data, a full model can not be used. We chose to estimate a single two-way interaction between the most dependent registers (DR and PR) and a three-way interaction. This model will estimate the number of unreported cases of liver cancer to about 25% of the total number of cases in all three registers together, accounting for overlap. The analysis is likely to be biased by false positive cases identified in the PR and/or DR.

ISBN 978-91-7549-472-2