Statistical methods for register based studies with applications to stroke

Anita Lindmark

Contents

List of papers

The thesis is based on the following papers:

I. Lindmark, A., Glader, E-L., Asplund, K., Norrving, B. and Eriksson, M. (2014). Socioeconomic disparities in stroke case fatality: observations from Riks-Stroke, the Swedish stroke register. *International Journal of Stroke*, 9(4): 429-436.

II. Lindmark, A., Van Rompaye, B., Goetghebeur, E., Glader, E-L. and Eriksson, M. (2016). The importance of integrating clinical relevance and statistical significance in the assessment of quality of care - illustrated using the Swedish stroke register. *PLoS ONE*, 11(4): 1-13.

III. Lindmark, A., de Luna, X. and Eriksson, M. (2016). Sensitivity analysis for unobserved confounding of direct and indirect effects using uncertainty intervals. Manuscript under revision. Resubmission invited by *Statistical Methods in Medical Research*.

IV. Karlsson, M. and Lindmark, A. (2014). `truncSP`: an R package for estimation of semi-parametric truncated linear regression models. *Journal of Statistical Software*. 57(14): 1-19.

V. Lindmark, A. (2016). Uncertainty intervals for unobserved confounding of direct and indirect effects with extensions to censoring and truncation. *Manuscript*.

Abstract

This thesis adds to the area of register based research, with a particular focus on health care quality and (in)equality. 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 semiparametric 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

Populärvetenskaplig sammanfattning

Sverige har en lång tradition av folkbokföring genom t.ex. de kyrkböcker som förts av församlingarna. Idag sker datainsamling i form av offentlig statistik, uppgifter som samlas in av olika myndigheter såsom Statistiska centralbyrån och Socialstyrelsen på uppdrag av den svenska staten. De insamlade uppgifterna görs tillgängliga för allmänheten genom t.ex. rapporter och forskare kan ansöka om att få använda data för specifika forskningssyften. Det finns även närmare hundra nationella kvalitetsregister som inte är en del av den offentliga statistiken utan har bildats för att följa upp och säkra kvaliteten hos olika delar av den svenska sjukvården. Riksstroke är ett kvalitetsregister med syfte att kontrollera och stödja kvalitetsutvecklingen inom den svenska strokevården. Ett viktigt led i detta arbete är att använda lämpliga statistiska metoder så att felaktiga slutsatser undviks och resurser används effektivt.

En central fråga för kvalitetssäkring är huruvida det existerar socioekonomiska skillnader (skillnader på grund av t.ex. utbildning eller inkomst) inom olika aspekter av strokevården. Den första artikeln i denna avhandling undersöker socioekonomiska skillnader i överlevnad inom olika tidsperioder efter stroke. Resultaten visar på skillnader beroende på inkomst, utbildningsnivå samt huruvida patienten bor ensam eller inte. När skillnader såsom dessa påvisas är en naturlig följdfråga, vad beror de på? För att undersöka vilka orsaksmekanismer som ligger bakom ett samband krävs sofistikerade statistiska metoder med antaganden som är svåra att testa. De tredje och femte artiklarna i avhandlingen handlar om metoder för att undersöka hur känsliga resultaten är för avvikelser från dessa antaganden.

Ofta bedöms kvalitet inte bara på patientnivå utan även på sjukhusnivå med utgångspunkt från mått såsom dödlighet, antal patienter som får en viss behandling, osv. Avhandlingens andra artikel handlar om hur man kan jämföra sjukhus på ett rättvist sätt (utifrån t.ex. den typ av patienter som de behandlar) och hur resultaten av sådana jämförelser påverkas av bland annat den statistiska bevisnivå man använder sig av.

Insamling av data kan vara behäftat med olika problem som komplicerar analyser och gör resultat otillförlitliga. Ett sådant problem är trunkering vilket innebär att data inte går att observera över eller under en viss nivå. Avhandlingens fjärde artikel beskriver mjukvara framtagen för analys av data då utfallet av intresse är trunkerat.

Preface

The whole of life is just like watching a film. Only it's as though you always get in ten minutes after the big picture has started, and no-one will tell you the plot, so you have to work it out all yourself from the clues.

- Terry Pratchett, Moving pictures

I have always quite liked this aspect of life, working it out on my own from the clues. But, for life in general and my PhD-studies in particular, I have learned that it is easier if you have other people to help you. This has been a long journey, and one I could not have completed alone.

First and foremost I would like to thank my main supervisor Marie Eriksson. Without you I do not think that I would have begun this journey, and I definitely would not have finished it! Thank you for always being positive and supportive and for helping me put things into perspective. Thank you also to my co-supervisors, Xavier de Luna for your methodological expertise, research ideas, and asking relevant questions, and Eva-Lotta Glader for your invaluable medical knowledge.

I have had the privilege of collaborating with many different researchers during my PhD-studies. This has been a great source of knowledge as well as new research topics and I would like thank all of my co-authors: Kjell Asplund, Bo Norrving, Bart Van Rompaye, Els Goetghebeur and Maria Karlsson. Thank you also to Riksstroke, the participating hospitals, and the secretariat.

Sharing an office can be a blessing or a curse. That is why I am so glad that I got to share the "PhD-office" with such great companions over the years. Inga, Hilda, Laura, and Gabriel thank you for the laughter and support! Philip, this time would have been so much less enjoyable without your company and my collection of in-jokes would be much smaller. Thank you for being a friend!

Thank you all of my colleagues at the Department of Statistics, past and present. I feel so lucky and proud to have been part of such a wonderful workplace!

I would like to thank my friends for reminding me that there is a life outside of the office. Miriam, I am thankful that I have you in my life. You are the kindest and most generous person I know, and different from me in all the best ways. You are a great travel companion and without you I would have experienced much less.

And last but not least I want to thank my family, my brothers Anders, Henrik and Richard and their families, and especially my parents. Evy for all your support, reminding me that eating something often makes all the difference when things feel impossible and for all the practical solutions that you come up with. Tack mamlan! Bertil for being kind and patient, and a berry picking machine!

Umeå, September 2016
Anita Lindmark

1   Introduction

Sweden has a long tradition of keeping population records, dating back to 1686 when parish registration was made mandatory through the Church Ordinance of that year (Statistics Sweden, 2016). Today there is a large number of nationwide registers constituting an important resource for research on areas such as community planning, economy, labour, and health and welfare. One such register is Riksstroke, the Swedish stroke register, which monitors and supports improvements of the quality of stroke care in Sweden (Asplund et al., 2011).

There are various methodological issues concerned with register based research. An important aspect of the quality assurance work performed by health care registers such as Riksstroke is making sure that individual hospitals perform at an acceptable level. This raises methodological questions regarding e.g. how precisely hospital performance can be measured, how to take into account differences in patient composition between hospitals, and the properties of the decision process used to determine whether or not a hospital performs at the desired level. Paper II of this thesis illustrates the diagnostic properties of a decision process for benchmarking hospital performance and highlights the importance of balancing clinical relevance with the statistical evidence level used.

Quality assurance is also concerned with health care equality and to that end an important research question is the identification of differences in care and outcome between population groups (defined by e.g. age, sex or socioeconomic status). Paper I of the thesis uses Riksstroke data to investigate socioeconomic differences in survival after stroke. If such differences are identified the next step is to ascertain why they exist. Are there factors due to the process, i.e. the health care itself, or are there qualities within certain population groups that cause them to be more vulnerable? Mediation analysis attempts to uncover causal mechanisms by decomposing the effect of an exposure (e.g. low education) on an an outcome (e.g. survival after stroke) into components where the exposure directly affects the outcome and components where the exposure has an indirect effect on the outcome (e.g. low education affecting life style factors that in turn have an effect on survival). Estimating these direct and indirect effects relies on untestable assumptions and the results of a mediation analysis should be reported together with an analysis of how sensitive they are to violations of these assumptions.

Paper III considers mediation analysis based on binary probit regression and suggests a sensitivity analysis method which is then illustrated using Riksstroke data.

Data registration is not always complete and sometimes data on a variable are unavailable above or below some value, 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 situations when the estimation of a linear regression model is made difficult by the fact that the outcome variable is truncated or censored. Paper IV presents a software implementation of three semi-parametric estimators of truncated linear regression models. Paper V extends the sensitivity analysis method proposed in Paper III to e.g. continuous outcomes and situations where the outcome is truncated or censored.

The thesis is structured as follows. An introduction to the Swedish system of national registers in general and Riksstroke in particular is given in Section 2. Concepts relevant to benchmarking hospital performance are presented in Section 3. An overview of mediation analysis is given in Section 4, followed by an introduction to truncation and censoring in Section 5. The papers in the thesis are then summarized, followed by concluding remarks.

2   Swedish national registers

Swedish official statistics are regulated by law and 27 different authorities have been appointed to keep records on statistics for public information, planning and research purposes in specified areas. The largest of these is Statistics Sweden which is the administrative agency responsible for non-sectoral statistics within four main areas: population and welfare, economic statistics, national accounts, and regions and environment (for more information see www.scb.se). Official statistics within the areas of health and medical care and social services are kept by the Swedish National Board of Health and Welfare (see www.socialstyrelsen.se). The official statistics are made available to the public through publication of reports and through aggregated data available on authority web pages. De-identified individual level data can be made available to researchers for specific projects, subject to ethical and secrecy vetting.

In addition to the official statistics regulated by Swedish law, there are 96 national quality registers with the purpose of monitoring and improving health care within different areas, such as diabetes and breast cancer (for a complete list see www.kvalitetsregister.se). These registers contain patient level data on e.g. baseline characteristics, medical interventions, and patient outcomes. Data are registered by health care workers as part of the clinical work and are made available through e.g. annual reports. Although the registries do not collect data for the main purpose of research de-identified micro level data can be made available for research projects after ethical vetting.

All individuals registered as part of the Swedish population are assigned a unique personal identification number. This allows different registers to be linked at an individual level. In Papers I-III of this thesis Riksstroke data was linked with data from the Swedish Cause of Death Register kept by the Swedish National Board of Health and Welfare to obtain patient death dates. In Paper I data on socioeconomic status was retrieved from the LISA database (Longitudinal integration database for health insurance and labor market studies) managed by Statistics Sweden, which integrates data from registers in the labor market, educational and social sectors. The data linkage was performed at Statistics Sweden and the Swedish National Board of Health and Welfare and de-identified data returned.

2.1 Riksstroke

Stroke is a disease that affects the brain, caused by poor blood supply due to a clot or bleed. It is the second leading cause of death globally (World Health Organization, 2014) and the main cause of disability in adults. Approximately 30 000 people in Sweden are diagnosed with stroke each year (Riksstroke, 2016).

About 85% of stroke cases in Sweden are ischemic strokes, usually caused by a blood clot forming locally in the brain's blood vessels or by the migration to the brain of clots formed due to arteriosclerosis (narrowing) of vessels in the neck. Intracerebral hemorrhage (about 10% of stroke cases) is caused by the rupture of a blood vessel inside the brain. The remaining 5% of cases are made up of subarachnoid hemorrhages, bleeding in the space surrounding the brain.

Poor blood flow causes death of cells in the brain tissue, leading

to symptoms related to the part of the brain that has been damaged. Typical symptoms are sudden weakness, unilateral numbness or loss of control of the face, arm, or leg, and sudden difficulty speaking or understanding speech. About one fifth of patients die within 90 days of stroke (18% in 2014, Riksstroke, 2014c) and many stroke survivors experience loss of functionality such as impaired speech or inability to perform daily activities.

The incidence of stroke increases with age and more than 80% of stroke cases are individuals over the age of 65 (Riksstroke, 2016). The risk factors for stroke are largely lifestyle related, such as smoking, heavy alcohol consumption, a diet that is low in fruits and vegetables, high levels of stress and inactivity. These lifestyle factors increase the risk of other medical conditions such as hypertension, diabetes, atrial fibrillation, high blood lipid levels and sleep apnea, which in turn increase the risk of stroke. Since an estimated 1.8 million Swedes suffer from hypertension (Nyström and Engfeldt, 2014) it is the single most important risk factor for stroke at the population level. Atrial fibrillation leads to a five-fold increase of the risk of stroke, making it the most important risk factor at the individual level (Wester and Norberg, 2014).

Riksstroke was established in 1994 to improve the quality of the Swedish stroke care (Asplund et al., 2011). It registers acute cases of ischemic stroke and intracerebral hemorrhage and all hospitals in Sweden that admit acute stroke patients are covered, 72 hospitals in 2014 (Riksstroke, 2014b). Data is collected during the acute phase, i.e. upon admission to hospital, and at follow-up 3 months and 1 year after the stroke. In 2014, Riksstroke covered 90% of acute stroke cases and about 88% of those registered during the acute phase were followed up at 3 months (Riksstroke, 2014c) and 78% at 1 year after stroke (Riksstroke, 2014a).

In addition to baseline patient characteristics such as sex, age, and risk factors, Riksstroke also registers indicators of quality of care. These can be divided into process and outcome indicators. Process indicators are indicators of what is done by the health care, such as treatments in the acute stage and after, follow-up appointments and secondary preventative measures (e.g. prescription of antihypertensive medications and anticoagulants). Outcome indicators map the results of these processes, e.g. survival, loss of function, etc.

The Regional Ethical Review Board in Umeå has approved Riksstroke in general (reference number 95-168) and its use for statistical method development in particular (2012-321-31M). All patients and their next of kin are informed about the registration, the aim of the register, and that data may be used for developing and ensuring the quality of stroke treatment, for compiling statistics and for health care research purposes. They are informed of their rights to decline participation (opt-out consent). No additional consent is collected for specific research projects.

Participation in Riksstroke involves little direct risk for the patients. The biggest risk is the potential vulnerability of the data to intruders. To ensure secure data handling and to minimize the risk of security breaches all data files are kept on servers at ICT Services and System Development (ITS) at Umeå University. Identification of individual patients is removed from working files, and files merged with other registers. Data are always presented at group level without the possibility to identify individuals.

3   Benchmarking hospital performance

One of the goals of Riksstroke is an equal and high quality stroke care throughout Sweden. To facilitate this, target levels have been implemented for 13 quality indicators. There are two tiers, moderate and high target fulfillment, and levels are set based on published international data, the current distribution among hospitals and a general assessment of how Sweden as a nation is ranked internationally. Hospital fulfillment of these targets is reported in published reports and on the Riksstroke web page (www.riksstroke.se). Similar quality reporting systems are in place for other Swedish quality registers.

Several quality assessment and assurance systems have been implemented worldwide, e.g. the World Health Organization's Performance Assessment Tool for Quality Improvement in Hospitals (PATH) (Veillard et al., 2005; World Health Organization Regional Office for Europe, 2016). In addition to using these tools as a guide for internal quality improvement results can be the basis for decisions by policy makers, allocation of resources and patients' choices of care providers.

There are many important aspects that need to be considered to reliably assess hospital performance. Two main points are the precision

with which hospital performance is measured and the properties of the decision rule used to classify hospitals into different performance levels. How precisely hospital performance is measured depends on e.g. the quality indicator used, hospital size and the patient case-mix, i.e. what type of patients that are treated at the hospital. A hospital that treats elderly patients or patients with more severe stroke may exhibit worse outcomes without this being the result of poorer quality of care.

3.1   Case-mix adjustment

The goal of case-mix adjustment (risk-adjustment, adjustment for confounding) in the context of benchmarking hospital performance is to account for differences in patient composition between hospitals and facilitate fair comparisons.

Ideally one would like to compare the outcome for a given patient in all the different hospitals where he/she could have received treatment. The *potential outcome* or *counterfactual* framework (Neyman, 1923; Rubin, 1974, 1977) has been developed to formalize such concepts. Let us denote the potential outcome for a patient under the care level of a specific hospital h as $Y(h)$, $h = 1, \ldots, m$. The problem of course is that each patient is treated by only one of these m hospitals, so the only potential outcome that we observe for a given patient is the one for the hospital where the patient was actually treated. The others will be unobserved (or "counter to fact"). The potential outcomes can however be estimated from observed data if we can take into account all the patient characteristics $\mathbf{X}$ that may influence both the outcome and the choice of hospital, commonly referred to as confounders. Specifically, the potential outcomes can be estimated from observed data if we can assume that the potential outcome for a given patient is independent of where they were actually treated given his/her characteristics. In mathematical notation this can be stated as $Y(h) \perp\!\!\!\perp H | \mathbf{X} = \boldsymbol{x}$ for all hospitals h, and all $\boldsymbol{x}$ in $\mathcal{X}$ (the support of $\mathbf{X}$). We also need to make a *positivity* assumption, stating that there is a positive probability for each individual, given his/her characteristics, of being treated by hospital h, $h = 1, \ldots, m$.

Estimation is often performed through regression modeling, where the outcome in question is related to patient characteristics and treatment hospital. Different modeling approaches can be employed to target

the potential risk for a patient under the care level of a specific hospital. Suppose that we have a binary quality indicator, e.g. whether or not the patient dies within 90 days after stroke, and that we use a logistic regression model to estimate the patient specific risk. In a *fixed effects model* hospital effects are added by including hospital indicators, $\mathbb{I}(H = h) = 1$ if the patient was treated at hospital h and 0 otherwise, $h = 1, \ldots, m$:

$$\mathbb{E}(Y|\mathbf{X} = \boldsymbol{x}, H) = \text{expit}\left(\boldsymbol{x}^T \boldsymbol{\beta} + \sum_{h=1}^{m} \psi_h \mathbb{I}(H = h)\right), \qquad (1)$$

where $\text{expit}(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$, $\boldsymbol{\beta}$ is a vector of regression parameters for the patient specific covariates, and ψ_h is the hospital effect of hospital h. The model parameters can be estimated using maximum likelihood (ML). Given that $\mathbf{X}$ is sufficient to adjust for hospital-outcome confounding, the potential outcome for a patient under the care level of hospital h can be obtained by replacing the effect of the hospital where the patient was actually treated in (1) with that of hospital h:

$$Y(h) = \mathbb{E}(Y|\mathbf{X} = \boldsymbol{x}, H = h) = \text{expit}\left(\boldsymbol{x}^T \boldsymbol{\beta} + \psi_h\right).$$

A disadvantage to the fixed effects model is that when there are many hospitals in the population of interest a large number of hospital effects need to be estimated and there may be problems with overfitting. In addition, if some of those hospitals have a small number of patients results of the estimation may be biased (Peduzzi et al., 1996).

In a normal *mixed effects model*, ψ_h in (1) is included as a random effect, considered as drawn from a normal distribution, e.g. $\psi_h \sim N(\mu_\psi, \sigma_\psi)$. Then we only need to estimate two parameters to capture the hospital effects, μ_ψ and σ_ψ, and problems with overfitting are avoided. A potential drawback to the normal mixed effects model is a tendency for estimated hospital effects to shrink towards the overall mean hospital effect μ_ψ (Normand et al., 1997; Austin et al., 2003; Kalbfleisch and Wolfe, 2013). This shrinkage may make it more difficult to detect outlying hospitals and may be especially pronounced for small hospitals as these have less influence on the estimation of μ_ψ and σ_ψ.

To address some of the issues with fixed and mixed effects models an alternative modeling approach is to use a fixed effects model with a Firth correction (Firth, 1993). This means that a penalized likelihood is used

in the ML procedure, reducing the bias of the ordinary ML estimator. Further, since this penalization is based on a distribution with heavier tails than the normal distribution, although the resulting estimates will be subject to shrinkage this problem tends to be smaller than for the mixed effects model (Varewyck et al., 2014).

As a side note, the fixed and mixed effects models described in this section are *conditional* (cluster-specific, subject-specific) models, i.e. the probability of the outcome is modeled as a function of the patient-specific covariates and hospital-specific parameters. This is of course appropriate here since we want to model the hospital effects explicitly in order to predict a patient's outcome had he/she been treated under the care level of a specific hospital in the population. A different approach to modeling clustered data is to use a *marginal* (population-averaged) model. In Paper I, where the purpose was not hospital comparisons but we wanted to adjust for treatment hospital, a marginal approach, generalized estimating equations (GEE) (Liang and Zeger, 1986; Diggle et al., 2002, pp. 146-147) was used. GEE does not model hospital effects explicitly but rather models the average outcome among patients with common characteristics and accounts for within hospital correlation by adjusting the covariance matrix of the estimated model parameters.

3.2 Performance measure

To compare hospital performance on a given quality indicator we need a summary measure that captures the overall performance for each hospital. Such measures are often based on *standardization*, the two most common types being *direct* and *indirect* standardization (Nicholl et al., 2013). Direct standardization tries to capture the average outcome if all patients in the population had received the level of care of a specific hospital h and compares this to the overall population average. Indirect standardization tries to capture the outcome if the patients of hospital h had received the "average level of care" and compares this to the observed average outcome of hospital h. Standardization can be based on predicted values from regression models, such as those described in the previous section (Roalfe et al., 2008; Varewyck et al., 2014).

Continuing the 90 day mortality example from the previous section, if the case-mix adjustment is complete so that the hospital effects in (1) express the causal effects of hospital on the outcome, the directly

standardized risk of hospital h can be estimated by averaging the $Y(h)$ over the population:

$$\hat{\mathbb{E}}\left[Y(h)\right] = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}(h) = \frac{1}{n}\sum_{i=1}^{n}\text{expit}\left(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \hat{\psi}_h\right),$$

where n is the size of the study population. The directly standardized risk of each hospital is then evaluated in relation to the overall average, $\frac{1}{n}\sum_{i=1}^{n}Y_i$, or some function thereof.

In indirect standardization the expectation calculated is the expected risk if the patients at hospital h had experienced the average treatment level across all hospitals:

$$\frac{1}{m}\sum_{h^*=1}^{m}\hat{\mathbb{E}}\left[Y(h^*)|H=h\right] = \frac{1}{mn_h}\sum_{h^*=1}^{m}\sum_{i=1}^{n_h}\text{expit}\left(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \hat{\psi}_{h^*}\right),$$

where n_h is the number of patients treated at hospital h. This is then compared with the observed overall risk at hospital h, $\frac{1}{n_h}\sum_{i=1}^{n_h}Y_i$, as either a ratio (e.g. the standardized mortality ratio, DeLong et al., 1997) or a difference (excess risk, Goetghebeur et al., 2011).

In direct standardization each hospital is evaluated against the same reference population, while in indirect standardization each hospital is evaluated based on its own patient case-mix. Indirect standardization may thus be more appropriate for evaluation of specialized hospitals that tend to treat a certain type of patients, but it does not allow direct hospital comparisons when the case-mix differs substantially between hospitals (Shahian and Normand, 2008). Direct standardization does not have this lack of comparability problem, as hospitals are evaluated against a common population and may be more interesting if we want/expect all hospitals to provide good health care on the overall patient population. However, direct standardization is less appropriate when a hospital tends to treat patients with a very different case-mix compared to the reference population. Since stroke is an acute illness where patients are brought to and treated at the nearest hospital differences in case-mix between hospitals tend to be small, and the ambition is to have an equal stroke care where all hospitals provide high quality care for the entire patient group. Therefore, Riksstroke uses direct standardization as the basis for hospital comparisons, and this method is also used for the evaluations in Paper II of the thesis.

3.3 Decision rule

Finally, we need to make a decision about whether or not a given hospital performs at the appropriate level based on the performance measure. Often, hospitals are evaluated in relation to each other, e.g. through league tables where a fixed proportion at the bottom and/or top of the table are classified as outlying. A disadvantage to using relative rankings as a classification method is that the same proportion of hospitals will be flagged as outlying regardless of the general quality of care.

Another frequently used method of reporting hospital performance is through funnel plots, where hospital specific performance measures are displayed around the overall mean/proportion (see Figure 1). To mark the boundary for outlying performance confidence limits (often 95%) ordered according to a factor that affects the precision of estimation,
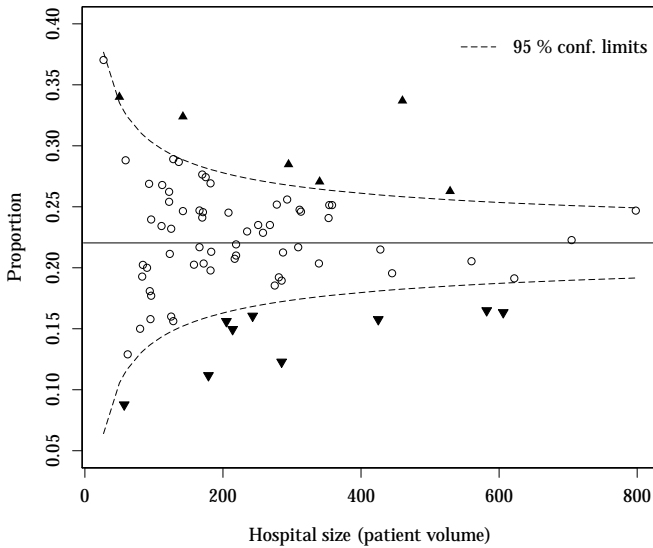

Figure 1: A funnel plot. The solid line represents the overall proportion. Outliers are marked with triangles.

e.g. hospital size, are displayed. This means that hospitals are classified as outlying if they fall outside the upper or lower confidence limits. For a detailed discussion of the advantages and disadvantages of funnel plots see e.g. Spiegelhalter (2005); van Dishoeck et al. (2011).

Table 1: Error matrix for classifying hospitals into the levels outlying or acceptable. The calculations of sensitivity, specificity, and positive and negative predictive values are displayed in the margins.

		True performance level:		
		Outl.	Acc.	
Result of benchmarking:	Outl.	A. Correct classification as outlier	B. False classification as outlier	$\text{PPV} = \frac{\#A.}{\#A.+\#B.}$
	Acc.	C. False classification as acceptable	D. Correct classification as acceptable	$\text{NPV} = \frac{\#D.}{\#C.+\#D.}$
		$\text{Sens.} = \frac{\#A.}{\#A.+\#C.}$	$\text{Spec.} = \frac{\#D.}{\#B.+\#D.}$	

The successfulness of a given benchmarking method can be summarized by the frequency with which hospitals are correctly classified. Suppose that we have two performance levels: outlying and acceptable. *Sensitivity* corresponds to the probability of a hospital being classified as outlying given that it truly exhibits outlying performance, while *specificity* is the corresponding probability for an acceptable classification. A high sensitivity thus corresponds to a low type II error rate and a high specificity to a low type I error rate. The *positive* and *negative predictive values* (PPV and NPV) measure the probability that the classification given to a hospital reflects its true performance level. Table 1 shows how these measurements are calculated based on the frequency of correct and incorrect classifications, where B. corresponds to a type I error and C. to a type II error.

The relative importance of the properties of a benchmarking method depends on the purpose of the benchmarking. Often a high level of statistical evidence is required before flagging a hospital as outlying, e.g. that a hospital falls outside the 95% confidence bounds in a funnel plot. Requiring a high level of statistical evidence means that fewer hospitals are erroneously flagged, i.e. high specificity, but that the power to detect outliers is lower, i.e. low specificity. If results are to be made publicly available this may be desirable to avoid undue damage to a hospital's reputation, but in many situations a more even balance between sensitivity and specificity may be desired.

Aside from the issue of which statistical evidence level to use, an important consideration is practical vs. statistical significance, since

the results of the benchmarking are often used to guide decisions, such as allocation of resources. The benchmark that hospitals are evaluated against should be set at such a level that hospitals that are flagged as outlying actually deviate to a clinically relevant degree from the norm.

4   Mediation analysis

The purpose of mediation analysis is to investigate the causal mechanisms behind the effect of an exposure on an outcome by decomposing it into an *indirect* effect and a *direct* effect. An indirect effect is the effect of the exposure (e.g. living alone) on the outcome (e.g death or disability at 3 months after stroke) that goes through some intermediate variable (e.g. stroke severity), called a *mediator*. A direct effect is the effect of the exposure on the outcome that does not go through the mediator. This can be illustrated through a directed acyclic graph (Lauritzen, 1996) such as the one in Figure 2, where Z is an exposure, Y an outcome, and M a mediator of the exposure-outcome relationship. The arrow from Z to Y corresponds to the direct effect, while the path from Z to Y that passes through M corresponds to the indirect effect.
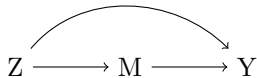

Figure 2: A directed acyclic graph showing the relationships between exposure Z, mediator M, and outcome Y.

To formally define direct and indirect effects we once again turn to the potential outcomes framework and use counterfactuals formulated by Robins and Greenland (1992) and Pearl (2001) for mediation analysis. Let us denote the potential value of the mediator for an individual under exposure level z as $M(z)$. Since the outcome depends on the exposure both directly and through its effect on the mediator we have the potential outcome $Y(z, m)$, the potential outcome if the exposure Z were set to the value z and the mediator M were set to the value m. We make a consistency assumption that for an individual with observed exposure $Z = z$ the observed mediator is given by $M = M(z)$, and the observed outcome is given by $Y = Y(z, M(z))$.

Direct and indirect effects can be defined in different ways (see e.g. Robins and Greenland, 1992; VanderWeele, 2015), the most common of which will be presented here, on the mean difference scale. The definitions presented are based on a binary exposure, taking value 1 if an individual is exposed and 0 if unexposed, but can be generalized to contrasting, e.g., any two levels of a continuous exposure.

The *controlled direct effect* is defined as

$$\text{CDE}(m) = \mathbb{E}\left[Y\left(1, m\right) - Y\left(0, m\right)\right], \tag{2}$$

which expresses the effect of an intervention that sets the mediator at some fixed level for the whole population, i.e the effect of Z on Y when fixing the value of the mediator at some value m for all individuals in the population. For example, let $Z = 1$ if an individual lives alone and 0 if that individual cohabits, Y be death or disability at 3 months after stroke, and $M = 1$ if an individual has a severe stroke (e.g. has a lowered level of consciousness at arrival to hospital) and 0 otherwise (the individual is fully conscious at arrival). $CDE(0)$ would then be the effect of living alone on death or disability if all individuals in the population were fully conscious at arrival to hospital.

The *natural direct effect*, NDE, is defined as

$$\text{NDE} = \mathbb{E}\left[Y\left(1, M(0)\right) - Y\left(0, M(0)\right)\right], \tag{3}$$

i.e., the mean difference in the outcome under exposure and the outcome in the absence of exposure when the mediator is allowed to vary as it would naturally if all individuals in the population were unexposed. Continuing the previous example, the NDE corresponds to the effect of living alone on death or disability if the stroke severity for each individual in the population were fixed to the value it would have taken if that individual was cohabitant.

The *natural indirect effect*, NIE, is defined as

$$\text{NIE} = \mathbb{E}\left[Y\left(1, M(1)\right) - Y\left(1, M(0)\right)\right], \tag{4}$$

i.e., the mean difference in the outcome when the mediator is allowed to vary as it would naturally if all individuals were exposed and the outcome when the mediator is allowed to vary as it would naturally if all individuals were unexposed, disabling the direct effect of the exposure on the outcome by keeping the exposure fixed at $Z = 1$ for all individuals

in the population. In our example the NIE corresponds to the effect on death or disability of changing stroke severity from its level if all individuals were cohabitant to its level if all individuals lived alone, disabling the direct effect of cohabitation status on survival by fixing it at cohabitant for all individuals in the population.

The controlled direct effect is of interest for policy evaluation while the natural direct and indirect effects are of interest when describing and evaluating the causal mechanisms at work (Pearl, 2001). The focus of Papers III and V in the thesis is on the natural direct and indirect effects, and these will be the focus of the remainder of this chapter. The theory for controlled direct effects is similar but less complicated than that for the natural direct and indirect effects as weaker assumptions are required for identification.

In order to estimate these effects from observed data we need to make assumptions about unmeasured confounding. There are different versions of these assumptions in the literature, for a review see e.g. Ten Have and Joffe (2012). Here we present the *sequential ignorability* assumption formulated by Imai et al. (Imai, Keele and Yamamoto, 2010; Imai, Keele and Tingley, 2010). This assumption states that conditional on a set of observed pre-exposure covariates $\mathbf{X}$ there is no unobserved confounding of the exposure-mediator and exposure-outcome relationship, $\{Y(z',m), M(z)\} \perp\!\!\!\perp Z|\mathbf{X} = \boldsymbol{x}$. This corresponds to the absence of $\mathbf{U}_1$ and $\mathbf{U}_3$ in Figure 3. It also states that given $\mathbf{X}$ and the observed exposure Z there is no confounding of the mediator-outcome relationship, $Y(z',m) \perp\!\!\!\perp M(z)|Z = z, \mathbf{X} = \boldsymbol{x}$, i.e. the absence of $\mathbf{U}_2$ in Figure 3. Sequential ignorability also includes a positivity assumption that states that all individuals must have a positive probability of being exposed or unexposed given the observed covariates and that there must be a positive probability of all potential mediator values given the observed exposure and covariates.

If these assumptions are fulfilled the potential outcomes in (3)-(4) can be rewritten using conditional expectations and probabilities (densities) and the natural direct and indirect effects conditional on the covariates
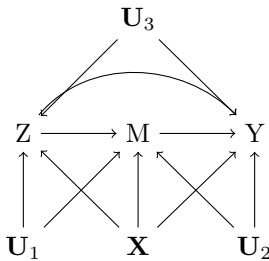
Figure 3: A directed acyclic graph with an exposure Z, a mediator M, an outcome Y, the set of observed confounders $\mathbf{X}$, and the unobserved confounders $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$.

are identified by

$$
\begin{aligned}
\text{NDE}(\boldsymbol{x}) = \sum_m \big\{ &\mathbb{E}\left(Y|Z=1, M=m, \mathbf{X}=\boldsymbol{x}\right) - \\
&\mathbb{E}\left(Y|Z=0, M=m, \mathbf{X}=\boldsymbol{x}\right) \big\} \times \mathbb{P}\left(M=m|Z=0, \mathbf{X}=\boldsymbol{x}\right),
\end{aligned} \quad (5)
$$

$$
\begin{aligned}
\text{NIE}(\boldsymbol{x}) = \sum_m &\mathbb{E}\left(Y|Z=1, M=m, \mathbf{X}=\boldsymbol{x}\right) \times \\
&\big\{ \mathbb{P}\left(M=m|Z=1, \mathbf{X}=\boldsymbol{x}\right) - \mathbb{P}\left(M=m|Z=0, \mathbf{X}=\boldsymbol{x}\right) \big\}.
\end{aligned} \quad (6)
$$

The marginal NDE and NIE are obtained by averaging these over $\boldsymbol{x}$. The version of this result given by Imai et al. (Imai, Keele and Yamamoto, 2010; Imai, Keele and Tingley, 2010) is called "non-parametric identification" since the quantities involved can be estimated nonparametrically based on sample averages and empirical distributions within strata defined by exposure, mediator, and covariates. However, in practice this quickly becomes cumbersome when the mediator is continuous or there is a large number of confounders to adjust for, the so-called "curse of dimensionality".

Traditional approaches for estimating mediation effects, e.g. the "product method" popularized by Baron and Kenny (1986), have relied on specifying parametric linear regression models for the outcome conditional on the mediator and observed covariates and the mediator conditional on the observed covariates. With the introduction of counterfactuals a more general parametric approach has been developed, allowing for non-linearities such as exposure-mediator interactions and binary outcomes and mediators (VanderWeele and Vansteelandt, 2009,

2010; Valeri and VanderWeele, 2013).

Since these methods rely on correct model specification semiparametric estimators have been suggested, e.g. multiply robust methods (van der Laan and Petersen, 2008; Tchetgen Tchetgen and Shpitser, 2012; Zheng and van der Laan, 2012). Ten Have and Joffe (2012) provide a review of different estimation methods that can be employed to estimate direct and indirect effects. The R (R Core Team, 2015) package `mediation` (Tingley et al., 2014, 2015) implements simulation algorithms introduced by Imai, Keele and Tingley (2010) that allows estimation of the natural direct and indirect effects based on parametric and/or semi-parametric models. It also offers non-parametric estimation for discrete mediators based on different research designs.

4.1   Sensitivity analysis

Sequential ignorability cannot be verified from observed data and thus a mediation analysis should be accompanied by a sensitivity analysis that investigates how sensitive the results are to violations of this assumption. The focus in the literature has been on sensitivity to unobserved mediator-outcome confounding ($\mathbf{U}_2$ in Figure 3). The motivation for this has been that confounding involving the exposure could be eliminated through randomization or by adjusting for a rich enough set of pre-exposure covariates. Confounding between the mediator and outcome is more difficult to account for, since randomization of both the exposure and mediator is difficult in practice and, without additional assumptions (Robins and Greenland, 1992; Avin et al., 2005; Petersen et al., 2006; De Stavola et al., 2015), identification builds on adjustment for pre-exposure covariates, i.e. covariates that either temporally precede the exposure or can be otherwise guaranteed to be unaffected by this. It is easier to argue that adjustment for pre-exposure covariates is sufficient for handling confounding involving the exposure than to argue that this adjustment is sufficient to render the mediator-outcome relation unconfounded, since we cannot rule out the possibility that there are variables that confound the mediator-outcome relation that we have not observed prior to exposure. It is important to note, however, that in an observational study confounding of any kind is difficult to rule out.

A sensitivity analysis is often performed by specifying a bias factor. The idea is that if there is unobserved mediator-outcome confounding

only adjusting for **X** will lead to biased estimators. If we can find expressions for this bias and use them to correct the estimates, we get an idea of how the direct and indirect effects would be affected by unobserved confounding. The bias factor is based on specification of some *sensitivity parameters*, certain quantities pertaining to the relationship between the unobserved confounder and the mediator and outcome (VanderWeele, 2010; Hafeman, 2011; le Cessie, 2016).

Imai, Keele and Yamamoto (2010) suggested a sensitivity analysis method using the correlation between the error terms in the parametric mediator and outcome models induced by unobserved mediator-outcome confounding as the sensitivity parameter. Rather than specifying a bias factor this correlation is incorporated into the expressions for the direct and indirect effects. The `mediation` package implements this approach for continuous mediators and outcomes and for situations where either the mediator or the outcome is binary. Figure 4 is a plot from the `mediation` package showing the NIE under unobserved mediator-outcome confounding with correlations ranging from -0.9 to 0.9.

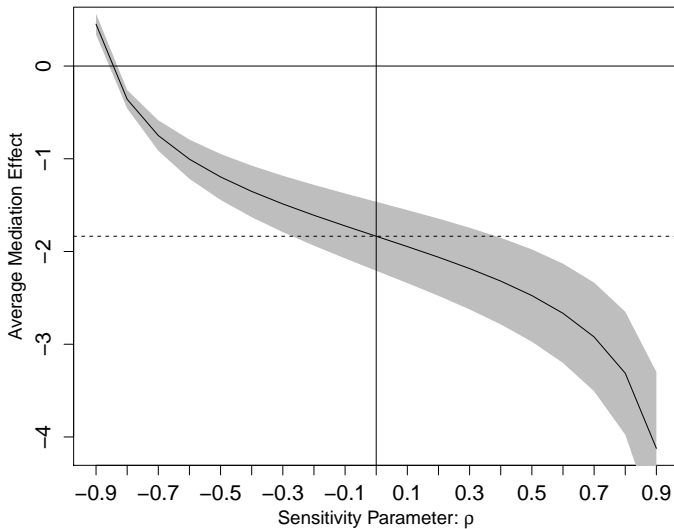

Figure 4: A sensitivity plot from the R package `mediation` for the NIE with the correlation ρ ranging from -0.9 to 0.9. The shaded area represents 95% CIs. The dashed line marks the effect under the assumption of no unobserved confounding.

Aside from showing the results from a sensitivity analysis over a range of sensitivity parameters it is often of interest to note the value(s) of the sensitivity parameter(s) that are required to eliminate the effect altogether. Alternatively, the sensitivity parameter(s) for which to give results may be determined by calculating the effect of not adjusting for the most important observed confounder to see the effect on the results of omitting an equally important confounder.

An alternative to sensitivity analysis is to derive bounds for the effects. These bounds do not use sensitivity parameters but rather find the minimum and maximum effects under some assumptions, e.g. monotonicity of the relationships between exposure, mediator and outcome (Cai et al., 2008; Sjölander, 2009; VanderWeele, 2011). A disadvantage is that these bounds are often quite wide, rendering them uninformative.

5 Truncation and censoring

Suppose that we have the following linear regression model

$$Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{7}$$

where ε_i are independent and identically distributed (i.i.d.) random variables with mean 0 and standard deviation σ_ε. The standard methods for estimating the unknown regression parameters $\boldsymbol{\beta}$ in (7) are ordinary least squares (OLS) and ML.

Now suppose that the outcome is unavailable if it is below (or above) a certain value. An outcome is *truncated* if it is not recorded at all below (or above) a certain value, t_i. An outcome is *censored* if it is not directly observed but recorded as e.g. being "smaller than" or "larger than" t_i. Take as an example the time from stroke onset to a patient's arrival to the hospital. For patients who wake up exhibiting stroke symptoms the last known symptom free time (i.e. when the patient went to sleep) is recorded in Riksstroke since the exact time of stroke onset is unknown. This is then used to calculate the time to arrival. This is an example of censoring: we know that the time from onset to arrival is less than the recorded value. A similar example of truncation would be a situation where, if a patient has exhibited stroke symptoms for a longer period of time, say ten hours, they do not make it to the hospital alive or at all and are thus never registered in Riksstroke. That is, times from stroke onset to hospital arrival over ten hours would be completely unobserved.

Data may be truncated/censored from the *left*, i.e. missing/incomplete below t_i, from the *right*, i.e. missing/incomplete above t_i, or a combination of the two. In the Riksstroke examples above time from onset to arrival was left censored and right truncated, respectively. Note that right truncated/censored data is easily transformed to left truncated/censored data by multiplying the outcome variable by -1, and vice versa.

The OLS estimator will be biased in the presence of truncation and/or censoring since the mean of ε_i will no longer be 0. For example, when the outcome is left truncated only $Y_i > t_i$ is observed, i.e. we only observe $\varepsilon_i > t_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$.

The standard ML estimator needs to be adjusted to take into account the truncation/censoring. The likelihood given the observed data under left truncation is given by

$$L\left(\boldsymbol{\beta},\boldsymbol{\alpha}|y,\boldsymbol{x}\right) = \prod_{i=1}^{n} \frac{f(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|\boldsymbol{\alpha})}{1 - F\left(t_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|\boldsymbol{\alpha}\right)}, \tag{8}$$

where $\boldsymbol{\alpha}$ is the vector of parameters from the distribution of the error term, $f(\cdot)$ is the pdf and $F(\cdot)$ the cdf of the error term. That is, the likelihood is made up of the density of the error term weighted by the probability of not being truncated, i.e. that $\varepsilon_i > t_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$. When the outcome is left censored $\max(Y_i, t_i)$ is observed and the likelihood given the observed data is given by

$$L\left(\boldsymbol{\beta},\boldsymbol{\alpha}|y,\boldsymbol{x}\right) = \prod_{i=1}^{n} f(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|\boldsymbol{\alpha})^{\mathbb{I}(y_i > t_i)} F\left(t_i - \boldsymbol{x}_i^T\boldsymbol{\beta}|\boldsymbol{\alpha}\right)^{(1 - \mathbb{I}(y_i > t_i))}, \tag{9}$$

where $\mathbb{I}(y_i > t_i)$ is an indicator that takes the value 1 if an observation is uncensored and 0 if it is censored. Maximizing (8)-(9) relies on the specification of the density of the error term. In the classic tobit model for censored outcomes (Tobin, 1958) and its equivalent for truncated outcomes (Hausman and Wise, 1977) a normality assumption is used.

The ML estimators are, however, sensitive to misspecification of the error term density (Vijverberg, 1987; Davidson and MacKinnon, 1993, p. 536) and semi-parametric estimators which require fewer assumptions on the error term have been developed. Among these are the symmetrically censored least squares (SCLS) estimator (Powell, 1986) for censoring and the symmetrically trimmed least squares (STLS) (Powell,

1986) and quadratic mode estimator (QME) (Lee, 1993; Laitila, 2001) for truncation. These only require the assumption that the error term is symmetric, although the QME has been shown to have good asymptotic properties for the slope parameters also under asymmetrically distributed error terms (Laitila, 2001). A drawback to imposing fewer assumptions on the error term is a difficulty in deriving finite sample properties of the semi-parametric estimators, meaning that inference is based on large sample theory. In addition, estimation of the asymptotic covariance matrix is often complicated so more computationally intensive bootstrap techniques may be necessary for estimation (Lee, 1993; Karlsson, 2004).

Paper IV of the thesis introduces a software implementation of three semi-parametric estimators of truncated linear regression models, while Paper V presents estimation and sensitivity analysis based on ML estimation for direct and indirect effects when the outcome is truncated or censored.

6 Summary of papers

6.1 Paper I

There is evidence that low socioeconomic status (low education and income level) is associated with increased stroke mortality (Cox et al., 2006; Addo et al., 2012) but the findings on the association between socioeconomic status and survival after stroke have been less consistent. In Paper I, *Socioeconomic disparities in stroke case fatality: Observations from Riks-Stroke, the Swedish stroke register*, we explore the association between four socioeconomic variables (education, income, country of birth and cohabitation) and survival after stroke. To investigate changes in survival (in)equality three time periods were investigated; the acute phase (0-7 days after stroke), the sub-acute phase (8-28 days), and the late phase (29 days to 1 year).

Analyses were based on Riksstroke data on 62 497 18-74 year old patients with onset of first stroke during the years 2001-2009. A logistic regression estimated using GEE was used to analyze survival (case fatality) by socioeconomic status adjusting for patient characteristics and taking into account within hospital correlation.

We found that socioeconomic status had a limited effect on acute

phase survival, indicating that there are only minor disparities in the acute stroke treatment. Differences related to income and cohabitation were however present already in the sub-acute phase, showing higher case fatality for low income patients compared to high income patients and for patients living alone compared to cohabiting. These gaps then expanded in the late phase. The association between education and case fatality was not present until 29 days-1 year after stroke, where patients with only primary school education had lower survival compared to patients with university education.

The study showed that even in a country such as Sweden with limited income inequity there may still exist large socioeconomic differences in long-term survival. This indicates that people who are socially underprivileged should constitute a target group for intensified stroke care and support during and after hospital stay.

6.2   Paper II

In this paper, entitled *The importance of integrating clinical relevance and statistical significance in the assessment of quality of care - illustrated using the Swedish stroke register*, we aim to support the complicated decision process involved in benchmarking hospital performance by illustrating the properties of a benchmarking method that integrates considerations of both clinical relevance and level of statistical significance.

The performance measure used was case-mix adjusted (directly standardized) risk of death or dependency in activities of daily living within 3 months after stroke. The decision rule used labeled a hospital as having outlying (poor) performance if its case-mix adjusted risk exceeded a benchmark value with a specified statistical confidence level. The benchmark was expressed relative to the population risk and should reflect a clinically relevant deviation. A simulation study based on Riksstroke patient data from 2008-2009 was performed, where hospitals were classified using the decision rule with varying degrees of statistical evidence and different benchmark values to investigate the effect on the sensitivity, specificity and positive and negative predictive values of the method.

The results of the simulation study showed that the widely used setting, comparing 95% confidence intervals to the national average, resulted in low sensitivity and high specificity whereas using a lower

statistical confidence level improved sensitivity with a relatively smaller loss of specificity. A lower confidence level also led to a more even balance between the positive and negative predictive values, whereas the standard approach led to a high PPV but a lower NPV. Variations due to different benchmark values were smaller, especially for sensitivity. This indicated that, for the given application, the choice of a clinically relevant benchmark could be driven by clinical factors without major concerns about sufficiently reliable evidence.

An Excel based tool (available at www.riksstroke.org/software/) allowing the user to insert relevant hospital risks and select a benchmark to obtain the theoretical sensitivities and specificities that result from different values of statistical confidence was also developed.

6.3   Paper III

In this paper, *Sensitivity analysis for unobserved confounding of direct and indirect effects using uncertainty intervals*, we propose a sensitivity analysis method for natural direct and indirect effects when the exposure, mediator and outcome are all binary. Our proposed method is able to assess sensitivity to not only mediator-outcome confounding but also confounding involving the exposure.

The method uses probit models to model the mediator, outcome, and exposure assignment mechanism and the sensitivity parameters consist of the correlations between the error terms of the mediator, outcome and exposure assignment models. These correlations are incorporated into the estimation of the model parameters which are then used to estimate the natural direct and indirect effects.

We take the sampling variability into account through the construction of uncertainty intervals, the union of all confidence intervals of the natural direct and indirect effects over a range of plausible correlation values.

To illustrate the method we applied it to a mediation study based on Riksstroke data, investigating the effect of living alone on the probability of death or being dependent in activities of daily living (ADL) 3 months after stroke, with stroke severity (level of consciousness) upon arrival to hospital as mediator. The results showed a larger sensitivity of the natural direct effect to unobserved exposure-mediator confounding than to unobserved mediator-outcome or exposure-outcome confounding. The

natural direct effect was quite sensitive to unobserved exposure-outcome confounding, but less sensitive to unobserved exposure-mediator and mediator-outcome confounding.

6.4   Paper IV

The majority of software available for the estimation and analysis of truncated regression models is based on maximum likelihood, assuming normally distributed error terms. This paper, entitled *truncSP: an R package for estimation of semi-parametric truncated linear regression models* presents an implementation of three estimators of semi-parametric truncated linear regression models in the R package truncSP (Lindmark and Karlsson, 2014). The estimators implemented are the previously mentioned STLS and QME estimators, as well as the left truncated (LT) estimator developed by Karlsson (2006) as a generalization of the QME not derived based on a symmetry assumption on the error term. The package also provides functions for the analysis of the estimated models, including bootstrap estimation of standard errors and confidence intervals. An empirical example based on air pollution data is used to illustrate the functions in the package.

6.5   Paper V

This paper, *Uncertainty intervals for unobserved confounding of direct and indirect effects with extensions to censoring and truncation*, builds on Paper III, extending the proposed sensitivity analysis method to include cases with continuous mediators and outcomes. We also investigate the case when the estimation is complicated by truncation or censoring of the continuous outcome variable, suggesting estimation methods and a sensitivity analysis for unobserved confounding based on maximum likelihood.

7   Final remarks and further research

This thesis contributes to areas within register based research, with a focus on health care. Specifically, contributions are made to the areas of hospital performance benchmarking, mediation analysis and regression when the outcome variable is limited. Through the development

of publicly available software, i.e. an Excel tool that helps guide decisions for benchmarking hospital performance and an R package that implements semi-parametric estimators for truncated linear regression models, more sophisticated statistical methods have been made available to researchers in practice.

In Paper I we found evidence that there are socioeconomic disparities in survival after stroke. It would be of interest to further investigate these disparities through a mediation study, using e.g. secondary preventative measures or life style factors as intermediates.

The sensitivity analysis method for direct and indirect effects proposed in Papers III and V is based on parametric regression models. It would therefore be relevant to investigate how sensitive the results are to violations of the distributional assumptions imposed. Future research should also include a generalization of the method to semi-parametric mediation analysis.

Paper V touched upon mediation analysis in the presence of censoring and truncation, using a maximum likelihood based approach for estimation and sensitivity analysis. Given the known limitations of ML estimation for censored/truncated regression models alternative estimation methods, such as semi-parametric estimators, should be investigated.

References

Addo, J., Ayerbe, L., Mohan, K. M., Crichton, S., Sheldenkar, A., Chen, R., Wolfe, C. D. and McKevitt, C. (2012). Socioeconomic status and stroke: An updated review, *Stroke* **43**(4): 1186–1191.

Asplund, K., Hulter Åsberg, K., Appelros, P., Bjarne, D., Eriksson, M., Johansson, Å., Jonsson, F., Norrving, B., Stegmayr, B., Terént, A., Wallin, S. and Wester, P.-O. (2011). The Riks-Stroke story: Building a sustainable national register for quality assessment of stroke care, *International Journal of Stroke* **6**(2): 99–108.

Austin, P. C., Alter, D. A. and Tu, J. V. (2003). The use of fixed- and random-effects models for classifying hospitals as mortality outliers: A Monte Carlo assessment, *Medical Decision Making* **23**(6): 526–539.

Avin, C., Shpitser, I. and Pearl, J. (2005). Identifiability of path-specific effects, *IJCAI-05, Proceedings of the nineteenth international joint*

conference on artificial intelligence, Edinburgh, Scotland, UK **July 30-August 5**: 357–363.

Baron, R. M. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology* **51**(6): 1173–1182.

Cai, Z., Kuroki, M., Pearl, J. and Tian, J. (2008). Bounds on direct effects in the presence of confounded intermediate variables, *Biometrics* **64**(3): 695–701.

Cox, A. M., McKevitt, C., Rudd, A. G. and Wolfe, C. D. A. (2006). Socioeconomic status and stroke, *The Lancet Neurology* **5**(2): 181 – 188.

Davidson, R. and MacKinnon, J. (1993). *Estimation and inference in econometrics*, New York: Oxford university press.

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B. and Micali, N. (2015). Mediation analysis with intermediate confounding: Structural equation modeling viewed through the causal inference lens, *American Journal of Epidemiology* **181**(1): 64–80.

DeLong, E. R., Peterson, E. D., DeLong, D. M., Muhlbaier, L. H., Hackett, S. and Mark, D. B. (1997). Comparing risk-adjustment methods for provider profiling, *Statistics in Medicine* **16**(23): 2645–2664.

Diggle, P., Heagerty, P., Liang, K. and Zeger, S. (2002). *Analysis of Longitudinal Data*, 2 edn, Oxford: Oxford University Press.

Firth, D. (1993). Bias reduction of maximum likelihood estimates, *Biometrika* **80**(1): 27–38.

Goetghebeur, E., Van Rossem, R., Baert, K., Vanhoutte, K., Boterberg, T., Demetter, P., De Ridder, M., Harrington, D., Peeters, M., Storme, G., Verhulst, J., Vlayen, J., Vrijens, F., Vansteelandt, S. and Ceelen, W. (2011). Quality insurance of rectal cancer - phase 3: Statistical methods to benchmark centers on a set of quality indicators, Good Clinical Practice (GCP). KCE Report 161C, D/2011/10.273/40.

Hafeman, D. (2011). Confounding of indirect effects: A sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome, *American Journal of Epidemiology* **174**(6): 710–717.

Hausman, J. A. and Wise, D. A. (1977). Social experimentation, truncated distributions, and efficient estimation, *Econometrica* **45**(4): 919–938.

Imai, K., Keele, L. and Tingley, D. (2010). A general approach to causal mediation analysis, *Psychological Methods* **15**(4): 309–334.

Imai, K., Keele, L. and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects, *Statistical Science* **25**(1): 51–71.

Kalbfleisch, J. D. and Wolfe, R. A. (2013). On monitoring outcomes of medical providers, *Statistics in Biosciences* **5**(2): 286–302.

Karlsson, M. (2004). Finite sample properties of the QME, *Communication in Statistics - Simulation and Computation* **5**: 567–583.

Karlsson, M. (2006). Estimators of regression parameters for truncated and censored data, *Metrika* **63**: 329–341.

Laitila, T. (2001). Properties of the QME under asymmetrically distributed disturbances, *Statistics and Probability Letters* **52**(4): 347–352.

Lauritzen, S. L. (1996). *Graphical models*, number 17 in *Oxford statistical science series*, Oxford: Claredon Press.

le Cessie, S. (2016). Bias formulas for estimating direct and indirect effects when unmeasured confounding is present, *Epidemiology* **27**(1): 125–132.

Lee, M. (1993). Quadratic mode regression, *Journal of Econometrics* **57**(1): 1–19.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**(1): 13–22.

Lindmark, A. and Karlsson, M. (2014). `truncSP`: Semi-parametric estimators of truncated regression models. R package version 1.2.
URL: *http://cran.R-project.org/package=truncSP*

Neyman, J. (1923). On the application of probability theory to agricultural experiments, essay on principles, *Roczniki nauk Rolczych* **X**: 1–51. In Polish. English translation by D.M. Dabrowska and T.P. Speed in *Statistical Science* **5**, 465-472.

Nicholl, J., Jacques, R. M. and Campbell, M. J. (2013). Direct risk standardisation: A new method for comparing casemix adjusted event rates using complex models, *BMC Medical Research Methodology* **13**(1): 1–9.

Normand, S.-L. T., Glickman, M. E. and Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: Issues and applications, *Journal of the American Statistical Association* **92**(439): 803–814.

Nyström, F. and Engfeldt, P. (2014). *Läkemedelsboken. Kapitel Hypertoni*, Läkemedelsverket.
URL: *http://www.lakemedelsboken.se/kapitel/hjarta-karl/hypertoni*

Pearl, J. (2001). Direct and indirect effects, *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 411–420.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. and Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis, *Journal of Clinical Epidemiology* **49**(12): 1373 – 1379.

Petersen, M. L., Sinisi, S. E. and van der Laan, M. J. (2006). Estimation of direct causal effects, *Epidemiology* **17**(3): 276–284.

Powell, J. (1986). Symmetrically trimmed least squares estimation for tobit models, *Econometrica* **54**(6): 1435–1460.

R Core Team (2015). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org*

Riksstroke (2014a). *Ett år efter stroke*, Umeå: Riksstroke.
  URL: *http://www.riksstroke.org/sve/forskning-statistik-och-verksamhetsutveckling/ovriga-rapporter/*

Riksstroke (2014b). *Riksstroke Årsrapport 2014*, Umeå: Riksstroke.
  URL: *http://www.riksstroke.org/sve/forskning-statistik-och-verksamhetsutveckling/forskning/arsrapporter/*

Riksstroke (2014c). *Riksstroke Årsrapport 2014 - 3-månadersuppföljning efter stroke*, Umeå: Riksstroke.
  URL: *http://www.riksstroke.org/sve/forskning-statistik-och-verksamhetsutveckling/forskning/arsrapporter/*

Riksstroke (2016). Information om stroke. Viewed on 19 September 2016.
  URL: *http://www.riksstroke.org/sve/patient-och-narstaende/stroke/*

Roalfe, A. K., Holder, R. L. and Wilson, S. (2008). Standardisation of rates using logistic regression: A comparison with the direct method, *BMC Health Services Research* **8**(1): 1–7.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects, *Epidemiology* **3**(2): 143–155.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology* **66**(5): 688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate, *Journal of Educational Statistics* **2**(1): 1–26.

Shahian, D. M. and Normand, S.-L. T. (2008). Comparison of "risk-adjusted" hospital outcomes, *Circulation* **117**(15): 1955–1963.

Sjölander, A. (2009). Bounds on natural direct effects in the presence of confounded intermediate variables, *Statistics in Medicine* **28**(4): 558–571.

Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance, *Statistics in Medicine* **24**(8): 1185–202.

Statistics Sweden (2016). SCB:s historia. Viewed on 19 September 2016.
  URL: *http://www.scb.se/sv_/Om-SCB/Verksamhet/SCBs-historia/*

Tchetgen Tchetgen, E. J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis, *Annals of Statistics* **40**(3): 1816–1845.

Ten Have, T. R. and Joffe, M. M. (2012). A review of causal estimation of effects in mediation analyses, *Statistical Methods in Medical Research* **21**(1): 77–107.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. and Imai, K. (2014). `mediation`: R package for causal mediation analysis, *Journal of Statistical Software* **59**(5): 1–38.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. and Imai, K. (2015). `mediation`: R package for causal mediation analysis. R package version 4.4.5.
URL: *http://CRAN.R-project.org/package=mediation*

Tobin, J. (1958). Estimation of relationships for limited dependent variables, *Econometrica* **26**(1): 24–36.

Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros, *Psychological Methods* **18**(2): 137–150.

van der Laan, M. J. and Petersen, M. L. (2008). Direct effect models, *International Journal of Biostatistics* **4**(1): Article 23.

van Dishoeck, A. M., Looman, C. W., van der Wilden-van Lier, E. C., Mackenbach, J. P. and Steyerberg, E. W. (2011). Displaying random variation in comparing hospital performance, *BMJ Quality & Safety* **20**(8): 651–7.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects, *Epidemiology* **21**(4): 540–551.

VanderWeele, T. J. (2011). Controlled direct and mediated effects: Definition, identification and bounds, *Scandinavian Journal of Statistics* **38**(3): 551–563.

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*, first edn, New York: Oxford University Press.

VanderWeele, T. J. and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition, *Statistics And Its Interface* **2**(4): 457–468.

VanderWeele, T. J. and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome, *American Journal of Epidemiology* **172**(12): 1339–1348.

Varewyck, M., Goetghebeur, E., Eriksson, M. and Vansteelandt, S. (2014). On shrinkage and model extrapolation in the evaluation of clinical center performance, *Biostatistics* **15**(4): 651–664.

Veillard, J., Champagne, F., Klazinga, N., Kazandjian, V., Arah, O. A. and Guisset, A.-L. (2005). A performance assessment framework for hospitals: the WHO regional office for Europe PATH project, *International Journal for Quality in Health Care* **17**(6): 487–496.

Vijverberg, W. P. M. (1987). Non-normality as distributional misspecification in single-equation limited dependent variable models, *Oxford Bulletin of Economics and Statistics* **49**(4): 417–430.

Wester, P. and Norberg, M. (2014). *Läkemedelsboken. Kapitel Cerebrovaskulära sjukdomar*, Läkemedelsverket.
URL: *http://www.lakemedelsboken.se/kapitel/hjarta-karl/cerebrovaskulara_sjukdomar*

World Health Organization (2014). Fact sheet 310: The top 10 causes of death. Last updated May 2014. Viewed on 19 September 2016.
URL: *http://www.who.int/mediacentre/factsheets/fs310/en/*

World Health Organization Regional Office for Europe (2016). Performance assessment tool for quality improvement in hospitals. Viewed on 19 September 2016.
URL: *http://www.pathqualityproject.eu/*

Zheng, W. and van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct and indirect effects, *International Journal of Biostatistics* **8**(1): Article 3.