Controlling for unobserved confounders in observational studies using large health care databases by means of instrumental variables in time-to-event analysis

Dem Fachbereich 03 Mathematik und Informatik der
Universität Bremen
zur Erlangung des akademischen Grades eines
Dr. rer. nat.

eingereichte Dissertation

von
Frau Dipl.-Math. Bianca Kollhorst
aus
Neustadt a. Rbg.

Datum der Einreichung: 2017/01/16

Erstgutachter: Prof. Dr. Iris Pigeot-Kübler
Zweitgutachter: Prof. Michal Abrahamowicz, PhD

Tag der mündlichen Prüfung: 2017/03/21

Danksagung

Auf diesem Wege möchte ich mich bei allen bedanken, die mich in den letzten Jahren bei meine Dissertation unterstützt haben. Der größte Dank gebührt meiner Doktormutter, Iris Pigeot, die immer ein offenes Ohr und Zeit für mich hatte, mich in die richtige Richtung geschubst hat, wenn ich mal falsch abgebogen bin und mich motiviert hat, wenn es dringend nötig war. Vielen Dank dafür!

Furthermore, I would like to offer my sincerest gratitude to my co-supervisor Michal Abrahamowicz who supported me with his patience and knowlegde and who helped me to define the topic of this dissertation.

Weiterhin gilt mein herzlicher Dank Vanessa Didelez, deren Ideen und Anregungen den Verlauf dieser Arbeit geprägt haben und die mir als Diskussionspartnerin stets mit wertvollen Ratschlägen zur Seite stand.

Ein großes Dankeschön geht auch an die Mitglieder meiner Fachgruppe, die mir in der letzten Zeit den Rücken frei gehalten haben, und alle Hochs und Tiefs mit sehr viel Verständnis, aufmunternden Worten und einer Engelsgeduld begleiteten. Insbesondere möchte ich Dirk, Frauke, Norman, Inga und Scholli dafür danken, dass sie immer ein offenes Ohr und Zeit für meine Probleme hatten.

Großer Dank gebührt auch Tike, Ricarda und Schütti für gelegentliche Ablenkung und den hin und wieder benötigten „Tritt in den Hintern“.

Nicht zuletzt gilt mein Dank meiner Familie, die immer an mich geglaubt und nie gezweifelt hat, dass ich das schon schaffen werde. Zum Schluss möchte ich Michael Dörks danken, dass er mich durchgehend unterstützt und motiviert hat und jederzeit bereit war, meine schlechte Laune zu ertragen und mit Kaffee, Fischstäbchen oder Fußball zu bekämpfen.

Abstract

Since 2011, the United States Food and Drug Administration is authorized to require post-authorization safety surveillance (PASS) studies to monitor recently approved drugs. Such PASS studies are a useful instrument to assess the effectiveness and safety of a drug after approval, since randomized controlled trials, the gold standard during the drug approval process, cannot provide all necessary information about drug reactions as they are limited by several factors. First, they often have a small sample size, second, they are additionally restricted to specific patient populations, that means they are not representative for the population of subsequent users, and, third, they are conducted in a controlled environment that does not reflect routine clinical practice. Observational studies in free-living populations are therefore necessary to evaluate the safety and effectiveness of drugs after drug marketing. Large health care databases are frequently used for this purpose, which, however, also suffer from certain limitations that may lead to biased results.

Statistical methods for controlling bias due to confounding include stratification, regression adjustment and propensity score approaches. None of these methods is able to remove this type of bias, unless all confounders are recorded in the data. A common problem of analyses based on administrative databases is that confounding variables such as smoking status or body mass index are often not recorded in health care claims databases, so that exposure effects are inconsistently estimated. Under certain conditions, instrumental variables can reduce or eliminate confounding bias in observational studies, so that IV estimators can consistently estimate treatment effects even in the absence of information on important confounders. Instrumental variable methods are well established for continuous outcomes using linear regression models, where two-stage least squares are typically used in observational studies. However, in time-to-event analysis no such common instrumental variable method exists which may be due to a number of complications that result from censoring and survivorship bias. Even if the proportional hazards model, the most popular model in time-to-event analysis, is used, two-stage estimators to account for instrumental variables are only justified for rare events. The aim of this thesis

is therefore to explore two-stage instrumental variable estimation for time-to-event outcomes in large health care databases if the assumption of rare events does not hold true.

The thesis consists of five chapters. Chapter 2 describes the German Pharmacoepidemiological Research Database (GePaRD) as an example for a large health care database. Considering the advantages and drawbacks of this database, studies of adverse drug effects pose a number of methodological challenges to be addressed by the research design. Three studies illustrating these challenges are presented in this chapter and the concept of bias and confounding is briefly introduced (Arfè et al., 2016; Kollhorst et al., 2015; Schröder et al., 2017). The chapter concludes with a description of methods to deal with measured confounding which are applied in a study on the comparative risk of death of antidepressants in older people with depression based on GePaRD presented in a forthcoming paper. Subsequently, Chapter 3 provides details on instrumental variable methods to reduce unmeasured confounding. The core conditions for a valid instrumental variable are outlined and the validity of the physician's preference as an instrumental variable in GePaRD is investigated (Kollhorst et al., 2016). Moreover, details on the two-stage predictor substitution method and the control function approach for instrumental variable estimation using linear models and the two-stage residual inclusion method for nonlinear models are given. The first two approaches are considered again in Chapter 4, where the focus is on the two-stage predictor substitution method and the control function approach in time-to-event analysis. Due to the non-collapsibility of the hazard ratio, instrumental variable methods such as the two-stage predictor substitution method and the control function approach cannot be readily applied to the Cox model. Therefore, two situations are presented where the use of these methods is justified in the context of proportional hazards models: first, there is no causal effect of the exposure on the outcome; second, the outcome is rare in the sense that the cumulative incidence remains low over the follow-up period. Finally, a simulation study that investigates the performance of these estimators for non-rare events or a non-null causal effect is conducted and results of the simulation study are transferred to an observational study based on GePaRD. Since the results of the simulation study

and the study comparing mortality risks between elderly new users of conventional and atypical antipsychotics have not been published yet, more details are provided. Finally, the thesis concludes with an overall discussion of results and a suggestion for future research and perspectives regarding the problem of unmeasured confounding in large health care databases.

Zusammenfassung

Um kürzlich zugelassene Medikamente überwachen zu können, ist die amerikanische Zulassungsbehörde „Food and Drug Administration“ seit 2011 autorisiert, sogenannte „Post-authorization safety surveillance“ (PASS) Studien zu fordern. PASS Studien eignen sich insbesondere, um die Wirksamkeit und Sicherheit eines Arzneimittels nach der Zulassung zu bewerten, da randomisierte kontrollierte Studien, der Goldstandard in der Zulassung, nicht alle nötigen Informationen über schwerwiegende Arzneimittelwirkungen liefern können. Die Anzahl der einbezogenen Personen ist häufig zu gering, um seltene schwere Arzneimittelwirkungen erkennen zu können, und die Patientenauswahl ist zusätzlich selektiv und daher nicht repräsentativ für die Gruppe der tatsächlichen Nutzer nach der Zulassung. Das Setting dieser Studien ist zudem sehr stark kontrolliert und entspricht nicht dem ärztlichen Versorgungsalltag. Um die Wirksamkeit und Sicherheit eines Arzneimittels nach Marktzulassung bewerten zu können, sind daher Beobachtungsstudien basierend auf Populationen, die den realen Versorgungsalltag widerspiegeln, erforderlich. Hierfür werden häufig große Gesundheitsdatenbanken eingesetzt, die allerdings auch bestimmten Limitationen unterworfen sind, die wiederum zu verzerrten Ergebnissen führen können.

Statistische Methoden wie Stratifizierung, Adjustierung oder Propensity Score Methoden können für Verzerrungen durch Konfounding kontrollieren, wenn alle Konfounder in den Daten enthalten sind. Da in administrativen Gesundheitsdatenbanken häufig wichtige Informationen wie zum Beispiel zum Rauchstatus oder dem Body Mass Index fehlen, können Expositionseffekte nicht konsistent geschätzt werden. Unter bestimmten Voraussetzungen können allerdings instrumentelle Variablen den Konfounding Bias reduzieren oder sogar beseitigen, so dass eine konsistente Schätzung der Behandlungseffekte möglich ist, selbst wenn wichtige Konfounder nicht beobachtet werden können. Für stetige Endpunkte und lineare Modelle sind etablierte Methoden zur instrumentellen Variablen Analyse vorhanden. Typischerweise werden in dieser Situation zweistufige Kleinste-Quadrate Schätzer verwendet. Durch eine Vielzahl von Problemen, die durch zensierte Daten und den Surviorship Bias entstehen, existieren

in der Überlebenszeitanalyse jedoch keine etablierten Methoden. Für das bekannteste Modell in der Überlebenszeitanalyse, das Cox-Modell, sind jedoch zweistufige Methoden im Falle eines seltenen Endpunkts anwendbar. Das Ziel dieser Arbeit ist die Untersuchung zweistufiger instrumenteller Variablen Methoden für die Überlebenszeitanalyse in großen Gesundheitsdatenbanken, wenn die Annahme eines seltenen Endpunkts nicht erfüllt ist.

Die Arbeit besteht aus insgesamt fünf Kapiteln. Als ein Beispiel für eine große Gesundheitsdatenbank beschreibt Kapitel 2 die Deutsche Pharmakoepidemiologische Forschungsdatenbank (GePaRD). Sowohl die Vor- als auch die Nachteile dieser Datenbank führen zu großen Herausforderungen bei der Durchführung von Arzneimittelrisikostudien, die bei der Wahl eines geeigneten Studiendesigns berücksichtigt werden müssen. Diese Herausforderungen werden in diesem Kapitel am Beispiel dreier Studien (Arfè u. a., 2016; Kollhorst u. a., 2015; Schröder u. a., 2017) erläutert und das Konzept von Bias und Konfounding wird kurz eingeführt. Das Kapitel schließt mit einer Beschreibung von Methoden, die zur Kontrolle von beobachteten Konfoundern eingesetzt werden können. Diese Methoden wurden in einer Studie zum vergleichenden Mortalitätsrisiko unter Behandlung verschiedener Antidepressiva bei älteren Menschen mit Depression basierend auf GePaRD eingesetzt, die demnächst publiziert wird. Kapitel 3 beschäftigt sich anschließend mit Methoden zur instrumentellen Variablen Analyse, um ungemessenes Konfounding zu reduzieren. Hierfür werden zunächst die Annahmen für eine valide instrumentelle Variable dargestellt und die Validität der Verschreibungspräferenz des Arztes als ein Beispiel für eine instrumentelle Variable in GePaRD untersucht (Kollhorst u. a., 2016). Außerdem werden die Two-stage Predictor Substitution Methode und der Control Function Ansatz für lineare Modelle und die Two-stage Residual Inclusion Methode für nichtlineare Modelle dargestellt. Mit Fokus auf die Überlebenszeitanalyse werden die beiden ersten Ansätze für lineare Modelle im darauffolgenden Kapitel 4 erneut betrachtet. Bedingt durch die Nicht-Kollabierbarkeit des Hazard Ratios sind beide Methoden in der Überlebenszeitanalyse nicht ohne Weiteres anwendbar. Daher werden zwei Ausnahmen dargestellt, in denen die Methoden konsistente Schätzer liefern: erstens, wenn kein kausaler Effekt der Exposition auf den Endpunkt vorliegt

ist, und zweitens, wenn der Endpunkt selten auftritt, so dass eine niedrige kumulative Inzidenz während des Follow-ups beobachtet werden kann. Schließlich wird die Performance der Schätzer mittels einer Simulationsstudie untersucht, insbesondere in Situationen, in denen der Endpunkt nicht selten ist und/oder ein kausaler Effekt angenommen wird, und die Methoden werden in einer Studie basierend auf GePaRD angewendet. Da sowohl die Simulationsstudie als auch die Studie zum Mortalitätsrisiko älterer neuer Antipsychotika-Nutzer bisher nicht publiziert wurden, werden beide detailliert dargestellt. In Kapitel 5 werden die Resultate zusammengefasst und diskutiert und es wird ein Ausblick auf zukünftige Forschung für das Problem des ungemessenen Konfoundings in großen Gesundheitsdatenbanken gegeben.

Contents

List of Figures

List of Tables

List of Abbreviations

2SPS	Two-stage predictor substitution
2SRI	Two-stage residual inclusion
AD	Antidepressant
AP	Antipsychotic
CF	Control function
CI	Confidence interval
COX-2	Cyclooxygenase-2
GePaRD	German Pharmacoepidemiological Research Database
HdPS	High-dimensional propensity score
HR	Hazard ratio
IV	Instrumental variable
MI	Myocardial infarction
NSAID	Non-steroidal anti-inflammatory drug
OAD	Oral antidiabetic drugs
PASS	Post-authorization safety surveillance
RCT	Randomized clinical trial

Chapter 1

Introduction

Since 2011, the United States Food and Drug Administration is authorized to require post-authorization safety surveillance studies to monitor recently approved drugs. Post-marketing studies are necessary, as the process that a drug has to pass before it is released to the market is not sufficient to learn about its potential risks and safety concerns. Prior to marketing, new drugs have to undergo three phases of clinical trials that are conducted in humans. A phase III study aims to prove the efficacy and safety of a new drug and mostly comprises several randomized controlled trials (RCTs) in a large patient population. However, RCTs are limited in their ability to provide all information about drug safety (Garbe and Suissa, 2014). Although usually thousands of patients are included in RCTs, an even larger sample size is required to identify rare adverse drug reactions. In addition, RCTs are restricted to a specific patient population that is not representative for the population of subsequent users after drug marketing. Especially, vulnerable groups such as children, older people, pregnant women and patients with comorbid conditions are typically excluded from RCTs. Furthermore, RCTs are conducted under strictly controlled conditions, for instance, under a fixed treatment regime that does not represent the routine clinical practice where individual treatment adaptations are required and use of the drug for an unapproved indication is common. RCTs are characterized by a short follow-up that does not allow to detect adverse drug reactions that only develop after a long induction period or in course

of a cumulative treatment.

Due to these reasons, phase III studies cannot provide sufficient information about drug safety, as these limitations increase the likelihood that certain adverse drug reactions will only be observed after drug approval. Observational studies in free-living populations are therefore necessary to evaluate the safety of drugs after approval. For a prospective monitoring of drug safety, reporting systems based on spontaneous reports of adverse events and large health care databases are frequently used (Hennessy, 2006). Health care databases can generally be subdivided into administrative databases and medical records databases. Medical records databases are derived from the electronic clinical records maintained by physician practices, usually general practitioners, whereas administrative databases are based on claims data of health insurance providers or state-funded health systems that are primarily collected for reimbursement purposes. Both types of databases offer several important advantages (Garbe and Suissa, 2014). Usually, they include a large number of patients that allow to study rare adverse drug reactions. In comparison to field studies, epidemiological studies based on these databases can be conducted in a cost efficient way and in a reasonable time frame, as the data collection is a by-product of health care delivery. As specific information about diagnoses and drug dispensations is collected, data are not affected by recall bias, especially if long-term memory is required. Due to the on-going nature of data collection, these databases also have the potential for a long follow-up that provides the possibility to investigate adverse drug reactions that only develop after a long period of drug intake or as a long-term effect. If the data collection is not restricted to a specific population, it can be used to detect drug effects in vulnerable groups such as older people or pregnant women. As no informed consent is necessary for the data collection, epidemiological studies based on health care databases are less prone to selection bias. Despite these major advantages, there are also limitations that researchers face. Since the data have been initially collected for administrative purposes, some important information is not available. Depending on the type of database, information on over-the-counter drugs, on the prescribed daily dose or the intended treatment duration, inpatient drug use, lifestyle factors such as smoking, socio-economic

status, or on laboratory values is lacking.

Since treatment is not randomly assigned in observational studies using large health care databases, a major concern is bias due to confounding. Confounding bias is present if the exposure groups differ before exposure administration, which may also affect the outcome, so that the association between the exposure and the outcome may be influenced by a third variable, a so-called confounder. Statistical methods to control for confounding include stratification, regression adjustment and propensity score approaches, but none of these methods is able to remove this type of bias unless all confounder are recorded in the data. A common problem of analyses based on health care databases is that confounding variables are often not recorded, so that exposure effects are inconsistently estimated. Unmeasured confounding can be controlled by study design such as case-only designs, by a validation study that additionally collects data on the unmeasured confounder in a subset of the study and incorporates this information in the analysis by using external adjustment, propensity score calibration, two-stage sampling and multiple imputation or at the analysis stage of a study (Uddin et al., 2016).

Under certain conditions so-called instrumental variables can reduce or eliminate unmeasured confounding at the analysis stage of a study. In observational studies, this method exploits random variation in the exposure assignment to define a variable that influences exposure but does not have an independent effect on the outcome, so that using this variable instead of the exposure is equivalent to pseudorandomizing the patients to alternative exposures. However, instrumental variable analysis can reduce bias in effect estimates due to unmeasured confounding, only if a valid instrument can be identified (Ionescu-Ittu et al., 2009; Martens et al., 2006). An observable variable is a valid instrument provided that all of the three following assumptions are met. First, the instrument is associated with the exposure. Second, the instrument is independent of unobserved confounders and third, conditionally on unmeasured confounders and exposure, the IV and the outcome are independent, which implies that the association between the instrument and the outcome is fully mediated by the observed exposure (Didelez et al., 2010).

Instrumental variable methods are well established for continuous outcomes using linear regression models, where two-stage least squares are typically used in observational studies (Angrist and Imbens, 1995). Here, the effect estimator is given as ratio of two ordinary least squares estimators obtained from the regressions of instrument on exposure and instrument on outcome. However, in time-to-event analysis no such common instrumental variable method exists which may be due to a number of complications that result from censoring and survivorship bias. Even if the proportional hazards model, the most popular model in time-to-event analysis, is used, two-stage estimators to account for instrumental variables are only justified in the context of rare events (Tchetgen Tchetgen et al., 2015). The aim of this thesis is therefore to explore two-stage instrumental variable estimation for time-to-event outcomes in large health care databases outside the context of rare events.

The thesis consists of five chapters based on five manuscripts that are reprinted in the Appendix. Chapter 2 describes the German Pharmacoepidemiological Research Database (GePaRD) as an example for a large health care database. Considering the advantages and drawbacks of this database, studies of adverse drug effects pose a number of methodological challenges to be addressed by the research design. Three studies illustrating these challenges are presented in this chapter and the concept of bias and confounding is briefly introduced (Arfè et al., 2016; Kollhorst et al., 2015; Schröder et al., 2017). The chapter concludes with a description of methods to deal with measured confounding which are applied in a study on the comparative risk of death of antidepressants in older people with depression based on GePaRD presented in a forthcoming paper. Subsequently, Chapter 3 provides details on instrumental variable methods to reduce or eliminate unmeasured confounding. The core conditions for a valid instrumental variable are outlined and the validity of the physician's preference as an instrumental variable in GePaRD is investigated (Kollhorst et al., 2016). Moreover, details on the two-stage predictor substitution method and the control function approach for instrumental variable estimation using linear models and the two-stage residual inclusion method for nonlinear models are given. The first two approaches are considered again in Chapter 4, where the focus is on the two-stage predictor substitution method

and the control function approach in time-to-event analysis. The chapter starts with an outline of these methods using additive hazards models. Due to the non-collapsibility of the hazard ratio, instrumental variable methods such as the two-stage predictor substitution method and the control function approach cannot be readily applied to the Cox model. Therefore, two situations are presented where the use of these methods is justified in the context of proportional hazards models: first, there is no causal effect of the exposure on the outcome; second, the outcome is rare in the sense that the cumulative incidence remains low over the follow-up period. Finally, a simulation study that investigates the performance of these estimators in situations outside the context of rare events or a null causal effect is conducted and results of the simulation study are transferred to an observational study based on GePaRD. Since the results of the simulation study and the study comparing mortality risks between elderly new users of conventional and atypical antipsychotics have not been published yet, more details are provided. Finally, the thesis concludes with an overall discussion of results and a suggestion for future research and perspectives.

Chapter 2

Large health care databases for pharmacoepidemiological research

Randomized clinical trials (RCTs) are used to assess the efficacy and safety of a medical intervention in a patient population. However, RCTs cannot provide all necessary information about drug reactions as they are limited by several factors. First, they often have a small sample size, second, they are additionally restricted to specific patient populations, that means that e.g. frail elders or children are excluded, and, third, they are conducted in a controlled environment that does not represent routine clinical practice. Observational studies in free-living populations are therefore necessary to evaluate the safety and effectiveness of drugs after drug marketing. Large health care databases are frequently used for this purpose (Hennessy, 2006).

The chapter is organized as follows: initially, the German Pharmacoepidemiological Research Database as an example for a large administrative health care database will be described. Then, challenges in the conduct and analysis of observational studies based on this database will be discussed. Commonly used study designs and the concept of bias and confounding will be briefly introduced. The chapter will conclude with a description of some methods to deal with confounding.

2.1 The German Pharmacoepidemiological Research Database

The German Pharmacoepidemiological Research Database (GePaRD) was established and is maintained by the Leibniz-Institute for Prevention Research and Epidemiology - BIPS. GePaRD comprises claims data from four statutory health insurance providers of about 20 million insurance members currently covering the years 2004 to 2014 and is continuously expanding. In Germany, content and structure of claims data are regulated by different articles of the Social Code Book V, so that the data can roughly be subdivided into four blocks: basic information, hospital admissions, outpatient physician visits and outpatient prescriptions presented in Figure 2.1. For each insurance member, GePaRD contains demographic information as well as information on insurance periods. Hospital data comprise information on the dates of hospitalization, diagnoses, reasons for admission and discharge as well as diagnostic and therapeutic procedures. Claims of outpatient physician visits include outpatient treatments and information on the treating physician. Furthermore, procedures and diagnoses along with the diagnosis certainty are available. All diagnoses are coded according to the German Modification of the International Statistical Classification of Diseases, whereas medical procedures are coded by the OPS classification system for surgeries and medical procedures (Operationen- und Prozedurenschlüssel). Prescription data are available for all reimbursed outpatient prescriptions and include the date of prescription and drug dispensation at the pharmacy, amount of substance prescribed, and information on the prescribing physician such as an identifier and the physician's specialty. Prescription data can be linked via the central pharmaceutical reference number to information on the anatomical-therapeutic-chemical code, the defined daily dose, packaging size, strength, formulation and generic and trade name. All four blocks are linked by a pseudonymous patient identifier. More detailed information can be found in Pigeot and Ahrens (2008).

Administrative claims databases and GePaRD in particular have several advantages. Since claims data are routinely collected, they reflect daily practice in clinical care. Specific information about diagnoses, dispensed drugs

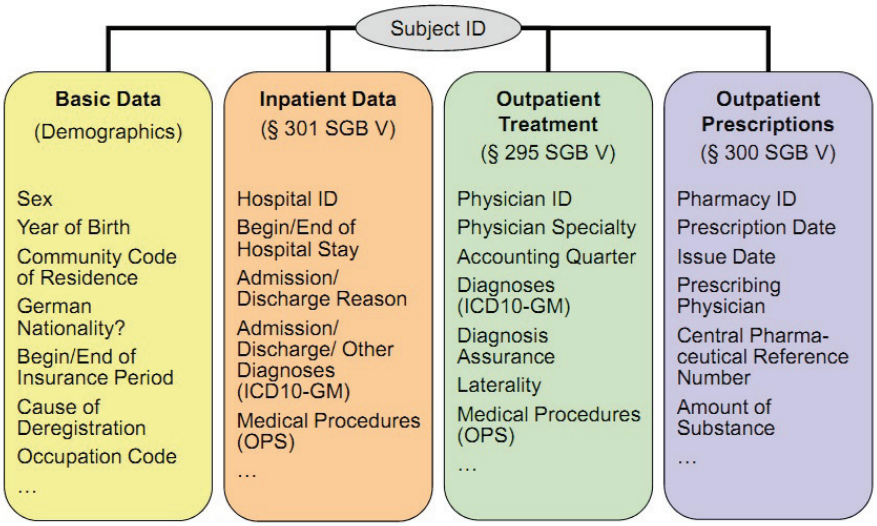

Figure 2.1: Data structure of GePaRD

and outpatient services are available, so that data are not affected by recall bias. No additional time and money are necessary to collect these data, so that quick access to a large nationwide population over a long time period is available. Due to its size, it further provides the possibility to study rare adverse events. But these advantages are counterbalanced by several limitations. As the data are mainly collected for reimbursement purposes, information on lifestyle factors, social status or clinical parameters are not available. Further limitations are the lack of information on prescribed daily dose and intended treatment duration, so that both need to be estimated, the lack of information on over-the-counter medication and inpatient drug treatment, and the lack of an exact date for ambulatory diagnoses and procedures, since these are only reimbursed on a quarterly basis.

2.2 Methodological challenges

In observational studies, advantages and limitations of GePaRD need be carefully considered when selecting an appropriate study design. This section will introduce the concept of three study designs that are commonly applied

in studies based on GePaRD, namely cross-sectional, case-control and cohort studies. Nevertheless, as not all sources of bias and confounding can be avoided by the study design, the section will conclude with a short overview of methods to deal with confounding.

2.2.1 Study designs in GePaRD

A longitudinal database such as GePaRD allows for various study designs which have to be carefully selected such that the resulting dataset is the most appropriate one to answer a specific research question. In the following, the most common study designs are illustrated by real-data examples and their potential and limitations are briefly stressed.

Cross-sectional studies

Schröder et al. (2017) compared the prevalence of antipsychotic drugs in children and adolescents based on claims data extracted from GePaRD (Appendix A). In children and adolescents, many antipsychotic drugs are often prescribed off-label, that means, that they are used for an unapproved indication, for an unapproved age group or for contraindications. The study was set up to investigate the utilization of antipsychotic drugs, especially the share of off-label prescriptions, in a pediatric population. An annual cross-sectional design for the years 2004 to 2011 was used to sample children and adolescents aged 0-17 years which were continuously observed for at least one year, unless they were born or died in the respective year. This cross-sectional study focused on describing annual prevalences of antipsychotics and proportions of off-label prescriptions in this specific study population. From 2006 to 2011, the prevalence of antipsychotic prescriptions increased from 2.03/1,000 individuals (95% CI: 1.97-2.09) to 2.61/1,000 individuals (95% CI: 2.54-2.68). The proportion of antipsychotics that were prescribed off-label increased between 2004 and 2006 from 61% to 69%, varied between 68.1% and 69.5% in the years 2007 and 2009, and decreased afterwards to 62%. The major strength of this study based on GePaRD is the size of the study population of about 2 million children and

adolescents in each year that provided the possibility to also investigate drug utilization stratified by sex, age group, drug class, substance and specialty of the prescribing physician. Using GePaRD, the cross-sectional study could be conducted with little time and effort. However, such a study design is not appropriate to investigate subsequent adverse events that may be caused by the off-label use of antipsychotics, as cross-sectional studies only cover a snapshot of time, so that usually exposure and outcome are assessed concurrently. Other study designs are needed to further examine adverse events associated with antipsychotics that were prescribed off-label. More details on results and discussion can be found in Schröder et al. (2017) (Appendix A).

Case-control studies

Arfè et al. (2016) (Appendix B) conducted a case-control study on the risk of heart failure associated with the use of non-steroidal anti-inflammatory drugs (NSAIDs) in five health care databases across four European countries. The case-control study was nested in a cohort of new users of NSAIDs that were at least 18 years old, continuously observed for at least one year in the respective database and had no diagnosis of cancer. Patients admitted to the hospital for heart failure were defined as cases. In each database, controls were selected by risk-set sampling from the underlying cohort and were matched on age, sex and the year of cohort entry which was defined as the first date of a NSAID prescription. Each control may become later a case and controls may be sampled more than once at various time-points. According to Breslow (1982), a case-control ratio of 1:4 is advisable to gain maximum efficiency, but in order to maintain efficiency even in subgroups of the original case-control set, up to 100 controls for each case were selected. To account for the matched design, multivariable conditional logistic regression was used to estimate the effect of current exposure to 27 individual NSAIDs. Current use of any NSAID was associated with an 19% increased risk of heart failure compared to past use of any NSAID. An increased risk was also found for nine individual NSAIDs. The major strength of this study was the use of five different data sources resulting from different populations and health care systems in four European countries. Nevertheless, the study also suffered from

some limitations. First, not all NSAID exposure periods could be captured as these drugs are also available as over-the counter medication. Second, dose-response analyses could only be conducted for two databases, as information on prescribed daily dose was not available in all databases. Third, due to a restrictive data privacy concept, only anonymized case-control data from each database were available that were provided in a common data format. Therefore, the underlying cohort was not available and incidence rates of the outcome could not be calculated. Details on results and their discussion can be found in Arfè et al. (2016) (Appendix B).

Cohort studies

In contrast to cross-sectional and case-control studies, cohort studies are characterized by the assessment of exposure and the subsequent surveillance for the outcome of interest, so that a clear temporal sequence of exposure and outcome is given (Miller et al., 2014). In a study based on claims data extracted from GePaRD, Kollhorst et al. (2015) (Appendix C) investigated the risk of myocardial infarction in a historical cohort of patients with type 2 diabetes treated with basal insulin. Time until occurrence of myocardial infarction (MI) was defined as the outcome. Myocardial infarction is a major complication in diabetic patients, mostly caused by insufficient glycaemic control. With an incidence rate of 13.5/1,000 person-years, the outcome was not rare, so that a cohort design was chosen. Only insurants with type 2 diabetes who were at least 18 years old and pretreated with oral antidiabetic drugs (OADs) were included in the cohort if they were additionally prescribed a basal insulin (Figure 2.2). Patients were assigned to one of three basal insulin exposure groups if they were solely treated with the respective drug during follow-up. The first prescription of an basal insulin marked the start of follow-up (cohort entry). Patients were then followed until the occurrence of myocardial infarction or administrative censoring. Outcome status for patients who could not be observed due to censoring remained unclear, so that e.g. a logistic regression model was not appropriate. To cope with this problem, data were analyzed using a proportional hazards model (see Chapter 4 for details). Adjusted hazard ratios showed no statistically significant difference between long-acting ana-

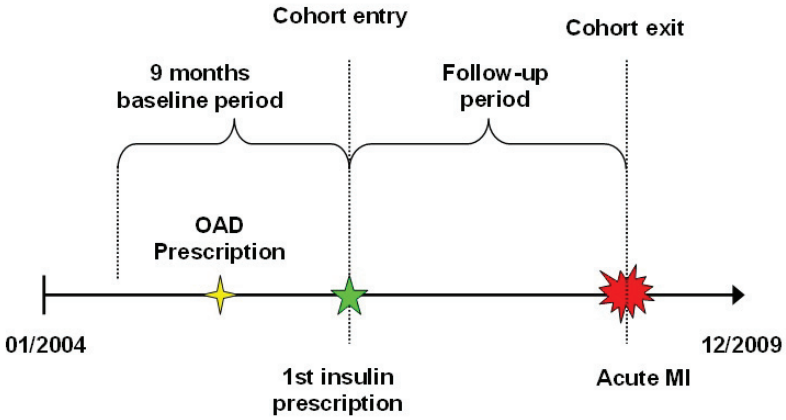

Figure 2.2: Study cohort

logue and neutral protamine Hagedorn insulin. In contrast, a 27% increased risk of acute MI for premixed insulin compared to analogue insulin was found that was no longer increased after propensity score matching. Antidiabetic treatment as prescribed in daily practice could be analyzed in a large sample of patients with type 2 diabetes which is the major strength of this study based on GePaRD. All conducted analyses were adjusted for multiple confounders, although some important confounders such as socio-economic status or duration of diabetes could not be taken into account as they are not available in the database. A limitation of this study is its relatively short follow-up of 5 years that precluded the assessment of long-term effects. For details on results and their discussion see Kollhorst et al. (2015) (Appendix C).

2.2.2 Bias and confounding

A biased estimator of the parameter of primary interest may occur either due to systematic or random errors. The validity (lack of systematic error) of a study can be distinguished in two types: internal and external validity. While internal validity corresponds to accurate measurement of the effect of an exposure on an outcome based on the study population, external validity of a study means that its results can be generalized to individuals outside the specific study population (Rothman and Greenland, 1998). Various types of

bias may be a threat to internal validity of an observational study where the most important ones may be classified as selection bias, information bias and bias due to confounding.

Selection bias stems from procedures to select individuals and from factors that influence the willingness to participate and may result in groups that are not comparable (Grimes and Schulz, 2002). In cohort studies, selection bias can be avoided by not selecting individuals for the comparison group who may be more or less likely to experience the outcome as the group with the exposure of interest, whereas in case-control studies, selections bias may occur if the selection of controls is based on the exposure or the selected controls are not representative for the source population (Section 2.2.1).

Information bias results from errors in the obtained measurements such as exposure, outcome or covariates and may lead to differential misclassification, if information between groups is assessed differently, and non-differential misclassification (Grimes and Schulz, 2002). Immortal time bias is a special case of differential misclassification bias that refers to a period in the follow-up time in which the outcome cannot occur (Suissa, 2007). This bias is introduced if the classification of exposure status is based on information that is assessed after start of follow-up rather than before follow-up. In order to be classified as an exposed patient, patients have to survive until the information can be obtained, so that the patients are immortal in this time-period. Consequently, this time is misclassified as exposed when in fact it is non-exposed time.

As the exposure in observational studies is not randomly allocated, confounding bias can impact the internal validity of an observational study. A confounder is a variable for which three conditions are necessary, but not sufficient (Rothman and Greenland, 1998). First, the confounder must be a predictor of the disease that is regarded as the outcome of interest. The confounder must at least be a marker for the cause of disease, but not necessarily the cause of the disease. Second, a confounder must be associated with the exposure in the source population. In a cohort study, this condition can be directly evaluated from the data if the confounder is assessed at cohort entry. In a case-control study, the confounder should be associated with the exposure

among the controls provided that the control group is free of selection bias. Third, a confounder must not be affected by the exposure or the outcome. Particularly, it must not be an intermediate in the causal path between exposure and outcome.

2.2.3 Methods to cope with confounding

Methods to cope with confounding can be already foreseen in the study design, or in the analysis, or even in both (Pearce and Greenland, 2014). In the design stage of a study, confounding can be avoided by randomization, restriction and matching, whereas stratification and regression adjustment can be used at the analysis stage. When using secondary data such as large health care databases, randomization is not feasible as the data are already collected. Randomization provides the possibility to balance the confounding factors across the levels of exposure, but may fail to control for all confounders. Restricting the study population to specific values of a confounding variable, e.g. female patients, prohibits the confounder from varying and thus, prevents confounding (Pearce and Greenland, 2014). To avoid bias that results from including prevalent users, pharmacoepidemiological studies based on health care databases commonly restrict their study populations to incident users and use a so-called new-user design. The inclusion of prevalent users may result in bias that is caused by, first, an underascertainment of events that occur in an early stage of the therapy and, second, the inability to control for risk factors that are altered by the exposure (Ray, 2003). Matching is another option to control for confounding at the design stage. In cohort studies, matching on potential confounding factors can be used to balance the exposure across strata of the matching variables and therefore, control for confounding by these factors. In case-control studies, matching does facilitate the control for confounding, but does not prevent confounding.

At the analysis stage, stratification can be used to investigate the association between exposure and outcome within different strata of the confounder. Usually, stratification is only an option to control for very few confounders simultaneously, as a cross-classification by many variables may result in empty

strata which in turn are uninformative. Confounding can also be controlled by adjusting for many confounders simultaneously in a regression model such as a logistic or Cox model (see Chapter 4). Nevertheless, regression models are also limited in their ability to adjust for multiple confounders, as the number of covariates that can be used for adjustment depends on the number of events per independent variable (Concato et al., 1995; Peduzzi et al., 1996). A way out of this dilemma would be to use propensity scores that do not have this limitation as the score is a scalar summary of the covariate information (D'Agostino, 1998). The propensity score of an individual is defined as the conditional probability of being treated, given the observed covariates, and is commonly estimated from the data using a logistic regression model. The propensity score balances the covariates across the levels of exposure and thus reduces confounding. In principle, four different propensity score methods can be used: matching, stratification, inverse probability of treatment weighting and adjustment (Austin, 2011).

In a cohort study not yet published, Kollhorst et al. investigated the comparative risk of death of 13 individual antidepressants in older people with depression. Antidepressants are mainly used to treat depression in older people, but are also prescribed for the treatment of other conditions e.g. neuropathic pain, anxiety or sleeping disorders. Due to the different co-indications, patients may have a higher propensity of treatment with some specific agents compared to others. To limit this confounding by indication at the design stage, the cohort was restricted to new users of antidepressants with a diagnosis of depression. Due to the inclusion of patients being older or at least 65 years old, additional confounding by frailty was suspected. Therefore, at the analysis stage, all analyses were adjusted for multiple confounders. Furthermore, a sensitivity analysis using a high-dimensional propensity score (HdPS) was conducted to assess the impact of unmeasured confounding (Guertin et al., 2016; Schneeweiss et al., 2009). The HdPS was calculated depending on a set of up to 500 empirically selected confounding variables derived from in- and outpatient diagnoses, inpatient operations and procedures, outpatient services and dispensations. The primary analysis was then adjusted by quintiles of the HdPS after 5% trimming. Further, a sensitivity analysis was conducted

by excluding cancer patients to assess the extent and direction of confounding by indication that may be caused by the common use of several drugs for the treatment of pain. However, although various approaches to deal with confounding by indication were applied, a possible effect on the results could not be ruled out. For results and a more detailed discussion, see the forthcoming paper presented in Appendix D.

Chapter 3

Instrumental variables methods to control for confounding

Statistical methods for controlling bias due to confounding include stratification, regression adjustment and propensity score approaches as described in Section 2.2.2. None of these methods is able to remove this type of bias, unless all confounders are recorded in the data. A common problem of analyses based on administrative databases is that confounding variables such as smoking status or body mass index are often not recorded in health care claims databases (Section 2.1), so that exposure effects are inconsistently estimated (Walker, 1996). Under certain conditions instrumental variables (IVs) can reduce or eliminate confounding bias in observational studies, so that IV estimators can consistently estimate treatment effects even in the absence of information on important confounders (Ionescu-Ittu et al., 2009).

The chapter is organized as follows: after motivating the need for instrumental variables, the core conditions that characterize an instrumental variable are outlined. Moreover, two-stage estimators based on the predictor substitution method, the control function approach and the residual inclusion method are introduced. This introduction is mainly based on Wooldridge (2010) and Terza et al. (2008).

3.1 Motivation

To motivate the need of instrumental variables, consider a simple linear model

$$\begin{aligned}
&Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} + \beta_p A + \gamma U + \varepsilon_0, \qquad (3.1)\\
&E(\varepsilon_0 | X_1, \ldots, X_{p-1}, A, U) = 0,\\
&Cov(X_i, X_j) = 0,\ i \neq j,\ i, j = 1, \ldots, p-1,\\
&Cov(X_j, U) = 0,\ Cov(X_j, A) = 0,\ j = 1, \ldots, p-1,\\
&Cov(X_j, \varepsilon_0) = 0,\ j = 1, \ldots, p-1,\ Cov(A, \varepsilon_0) = 0,\ Cov(U, \varepsilon_0) = 0,
\end{aligned}$$

where $Y, X_1, \ldots, X_{p-1}, A$ are observed random variables, U is an unobservable random variable, ε_0 is the random error and $\beta_0, \ldots, \beta_p$ and γ are the parameters to be estimated. As, both, U and ε_0 cannot be observed, they can be combined into one error term $\varepsilon_Y = \gamma U + \varepsilon_0$ and (3.1) can be rewritten as

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} + \beta_p A + \varepsilon_Y. \qquad (3.2)$$

As an intercept β_0 is included in (3.1), it can be assumed without loss of generality $E(U) = 0$. Furthermore, as $X_1, \ldots, X_{p-1}, A, U$ are independent of ε_0, it follows that $E(\varepsilon_0) = 0$ and therefore $E(\varepsilon_Y) = 0$. Assume that, in an observational study, A is the exposure of interest and U an unmeasured confounder and both are correlated. Hence, $Cov(A, \varepsilon_Y) \neq 0$ as U is included in ε_Y. However, the key assumption to consistently estimate the β_j's using ordinary least squares is that the error term has mean zero and is not correlated with any of the X_j's or A (Wooldridge, 2010). In the presence of unmeasured confounding, this assumption is not fulfilled and an alternative approach is needed to obtain consistent estimators of any of the β_j's in (3.1).

3.2 Core conditions

Instrumental variables methods may yield consistent estimates of exposure effects in the presence of unmeasured confounding. The idea is to find an observable variable that is associated with the exposure, but is independent of unmeasured confounders, and has no direct effect on the outcome except

through its association with the exposure (Angrist et al., 1996). The following three core conditions characterize an instrumental variable Z in an observational study. Let Z be a random variable, A the exposure and Y the outcome of interest. An unobserved variable that captures the confounding between A and Y is denoted by U, while an observed confounder of the association between A and Y is denoted by C. Let us further denote the statistical independence of B and C given D by $B \perp\!\!\!\perp C \,|\, D$. Then, Z is an instrumental variable if the following three conditions are fulfilled (Didelez et al., 2010):

1. $Z \not\perp\!\!\!\perp A \,|\, C$, that is, the IV must not be independent of the exposure, conditional on the observed confounder,
2. $Z \perp\!\!\!\perp U \,|\, C$, that is, the IV must be independent of unobserved confounders, conditional on the observed confounder, (*Independence Assumption*),
3. $Z \perp\!\!\!\perp Y \,|\, (A, U)$, that is, conditionally on the exposure and unobserved confounders, the IV and the outcome must be independent (*Exclusion Restriction*).

In general, the first assumption can be verified based on the data at hand, i.e. for a specific study question by a statistical test. As the confounder U is unobserved by definition, neither the *Independence Assumption* nor the *Exclusion Restriction* are empirically verifiable, but the plausibility of both can be explored on the basis of subject matter or background knowledge. Directed

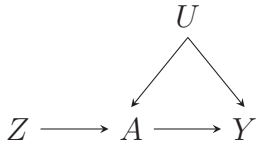

Figure 3.1: The core conditions

acyclic graphs can be used to represent the core conditions (Figure 3.1). A directed graph consists of a set of nodes connected by edges with a direction assigned. A directed acyclic graph is a directed graph in which no node has a directed sequence of nodes back to itself (Greenland et al., 1999).

One example of an instrumental variable Z is the randomized treatment assignment in an randomized clinical trial (Greenland, 2000). All three conditions are fulfilled due to the design of a clinical trial. The exposure A is defined as the received treatment that is affected, but not completely determined by the randomization. The *Independence Assumption* is met, because the assigned treatment is randomly allocated and therefore it is assumed that unobserved confounders are equally distributed between exposure groups. The *Exclusion Restriction* also holds as double-blinding ensures that the assignment to treatment groups does not depend on the outcome.

Another example for an instrumental variable in a pharmacoepidemiological context is the physician's preference in a study on the effects of selective cyclooxygenase-2 (COX-2) inhibitors compared to a non-selective non-steroidal anti-inflammatory drug (NSAID) on gastrointestinal complications (see Appendix E). In this study, it is assumed that physicians substantially differ in their preference for prescribing one of the therapeutic alternatives. To estimate the physician's true treatment preference, three different definitions were used. The physician's current preference for COX-2 inhibitors compared to non-selective NSAIDs was defined by using, first, the most recent prescription written by this physician (binary instrument), second, the proportion of previous patients (continuous instrument), and third, a set of indicator variables for the physician's seven prior prescriptions. For instance, over-the-counter aspirin use was assumed to be a strong unmeasured confounder. Kollhorst et al. (2016) evaluated the three core conditions for the physician's preference for a COX-2 inhibitor as an instrumental variable in the full cohort and in three sub-cohorts based on either patient or physician characteristics. First, three measures were calculated to assess the strength of the association between the IVs and the actual treatment: the partial F-statistic for the adjusted IV effect, the squared partial correlation r^2, and the estimated effect of the IV on the probability of treatment (Rassen et al., 2009a,b). The authors showed that, in general, the physician's preference based on the proportion of previous patients was a stronger predictor of the actual treatment than the other two instruments. Second, in order to evaluate the plausibility of the *Independence Assumption*, the authors examined whether the measured confounders between

exposure groups are balanced. To ensure the comparability of the covariate balance between the binary and the continuous IV, the partial F-statistic instead of the difference between the mean covariate values in the two treatment groups was chosen, as it applies to both analyses. If the observed confounders can be demonstrated to be equally distributed between exposure groups, it is thought to be likely that unmeasured confounders are balanced. Compared to the actual treatment groups, the IVs reduced high imbalances among the patients' risk factors, but it was discussed that the estimation of the IVs was influenced by the physician's true preference and by the types of patients seen by each physician. Therefore, it could not be excluded that there is a case-mix of patients between physicians with different specialty (Swanson et al., 2015), which would violate the *Independence Assumption*. Third, the association between the IVs and gastroprotective agents that are prescribed at the same day as the NSAID prescription was examined to explore the *Exclusion Restriction*. The authors argued that in general the physician's preference for a COX-2 inhibitor is not related to the occurrence of gastrointestinal complications, but it is possible that COX-2 prescriber are more likely to co-prescribe protective agents. The results indicated that there might be a violation of the *Exclusion Restriction* in some sub-cohorts. Finally, the IV risk estimates of gastrointestinal complications using the two-stage predictor substitution method (see 3.3 for details) had a highly inflated variance and differed from the results obtained by RCTs and from two other observational studies. The differences to the observational studies may partly be explained by the study period, the individual drugs that were studied, and the study populations, whereas the moderate strength of the association between IV and treatment may provide a possible explanation for the differing results when compared to the RCTs. In conclusion, the proportion of all previous patients is a potential IV, but the study also shows that IVs that are valid in one health care system may not be directly applicable to others. Further details on the results and their discussion can be found in Kollhorst et al. (2016) (Appendix E).

3.3 Two-stage predictor substitution method

As already mentioned, a possible solution to the problem of inconsistent estimation of treatment effects caused by unmeasured confounding presented in 3.1 is to find an instrumental variable Z that satisfies the three conditions described in Section 3.2. More precisely, the relationship between the endogenous variable A and the IV Z is given by

$$A = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_{p-1} X_{p-1} + \alpha_p Z + \varepsilon_1 \tag{3.3}$$

with $E(\varepsilon_1) = 0$ and ε_1 is uncorrelated with $X_1, \ldots, X_{p-1}, A$. First, it is assumed that $Cov(A, Z) \neq 0$, so that the IV and the exposure variable A are not independent. Second, Z is supposed to be independent of the error term in (3.2), i.e. $Cov(Z, \varepsilon_Y) = 0$. Third, the IV must be independent of Y, that means, Z is not included in (3.2).
The idea of the two-stage predictor substitution (2SPS) method is to insert (3.3) in (3.1), which leads to

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \ldots + \beta_p(\alpha_0 + \alpha_1 X_1 + \ldots + \alpha_p Z + \varepsilon_1) + \gamma U + \varepsilon_0 \\ &= (\beta_0 + \beta_p \alpha_0) + (\beta_1 + \beta_p \alpha_1) X_1 + \ldots + \beta_p \alpha_p Z + \beta_p \varepsilon_1 + \gamma U + \varepsilon_0 \\ &= \delta_0 + \delta_1 X_1 + \ldots + \delta_{p-1} X_{p-1} + \delta_p Z + \varepsilon_{2SPS} \end{aligned} \tag{3.4}$$

with $\varepsilon_{2SPS} = \beta_p \varepsilon_1 + \gamma U + \varepsilon_0$. By definition, ε_{2SPS} and Z as well as ε_{2SPS} and $X_1, \ldots, X_{p-1}$ are not correlated. Thus, the parameters δ_j, $j = 1, \ldots, p$, can now be consistently estimated using ordinary least squares. However, the parameters can only be consistently estimated if the model in (3.1) is correctly specified and the model in (3.3) corresponds to the correct projection.

3.4 Control function approach

An alternative to the two-stage predictor substitution method is the control function approach. The basic idea of this approach is that there exits a control function, which is a function of residuals of the model predicting exposure such as (3.3), that corrects for unmeasured confounding (Heckman and Robb, 1985).

The control function approach using survival models will be further described in Chapter 4. Here, a brief example to illustrate the idea will be given using linear models. Let us consider again the following two models that describe, first, the relationship between outcome Y and exposure A and, second, between instrumental variable Z and exposure A given by

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} + \beta_p A + \varepsilon_Y \tag{3.5}$$

$$A = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_{p-1} X_{p-1} + \alpha_p Z + \varepsilon_A \tag{3.6}$$

with

$$Cov(X_j, \varepsilon_Y) = 0,\ j = 1, \ldots, p-1, \tag{3.7}$$

$$Cov(X_j, \varepsilon_A) = 0,\ j = 1, \ldots, p-1,\ Cov(A, \varepsilon_A) = 0. \tag{3.8}$$

In this case, ε_Y contains an unmeasured confounder U that is associated with A, so that ε_A is correlated with ε_Y. Consequently, the linear projection of ε_Y on ε_A is given as

$$\varepsilon_Y = \rho \varepsilon_A + \varepsilon_0, \tag{3.9}$$

where $\rho = Cov(\varepsilon_A, \varepsilon_Y)/E(\varepsilon_A^2)$ and by definition, $Cov(\varepsilon_A, \varepsilon_0) = 0$. Because of (3.7) and (3.8), $Cov(X_j, \epsilon) = 0, j = 1, \ldots, p-1$, and $Cov(A, \varepsilon_0) = 0$. Plugging (3.9) in (3.5) gives

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} + \beta_p A + \rho \varepsilon_A + \varepsilon_0. \tag{3.10}$$

As ε_0 is uncorrelated with $X_1, \ldots, X_{p-1}, A$, and ε_A, all parameters in (3.10) can be estimated consistently.

Estimation is carried out in two steps. From the first stage model (3.6), parameters are estimated by ordinary least squares. For the estimation of the β_j's and ρ in the second stage model (3.10), ε_A is replaced with

$$\widehat{\varepsilon}_A = A - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X_1 + \ldots + \widehat{\alpha}_{p-1} X_{p-1} + \widehat{\alpha}_p Z).$$

Substitution and rearrangement in (3.10) gives

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1} + \beta_p A + \tau$$

with

$$\tau = \rho((\alpha_0 - \widehat{\alpha}_0) + (\alpha_1 - \widehat{\alpha}_1) X_1 + \ldots + (\alpha_{p-1} - \widehat{\alpha}_{p-1}) X_{p-1} + (\alpha_p - \widehat{\alpha}_p) Z) + \epsilon,$$

which depends on the sampling error in $\widehat{\alpha}_0, \widehat{\alpha}_1, \ldots \widehat{\alpha}_p$. The inclusion of the residuals control for the unmeasured confounding and, in the fully linear case, the estimates obtained from the control function approach and from the two-stage predictor substitution method are identical (Wooldridge, 2010).

3.5 Two-stage residual inclusion method

Let Y denote the random outcome variable, let $\boldsymbol{V} = (1, X_1, \ldots, X_{p-2}, A, U)$ denote an $n \times (p+1)$-dimensional matrix that contains $p-2$ observable exogenous random variables $X_1, \ldots, X_{p-2}$ and one observable endogenous random variable A and let U denote an unobservable random variable that mimics the unobserved confounder. Let $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T \in \mathbb{R}^{p+1}$ be a vector of unknown parameters. The linear predictor η is expressed as a linear combination of the unknown parameters $\boldsymbol{\beta}$ with $\eta = \boldsymbol{\beta}^T \boldsymbol{V}$. Here, we consider a nonlinear regression model which is given by

$$Y = f(\eta) + \varepsilon_Y, \tag{3.11}$$

where $f(\eta)$ is a twice continuously differentiable nonlinear function in $\boldsymbol{\beta}$ and ε_Y is the random error defined as $\varepsilon_Y = Y - f(\eta)$. Furthermore, the conditional expectation for Y given $\boldsymbol{V}$ is linked to η by the function $f(\cdot)$ such that

$$E(Y|\boldsymbol{V}) = f(\eta). \tag{3.12}$$

Therefore, it follows from (3.12) that $E(\varepsilon_Y|\boldsymbol{V}) = 0$.
Let $\boldsymbol{W} = (1, X_1, \ldots, X_{p-2}, Z)$ denote an $n \times p$-dimensional matrix and $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T \in \mathbb{R}^p$ a vector of unknown parameters. The parameter α_{p-1} is the coefficient of IV Z. The relationship between the exposure A and the instrumental variable Z is described by

$$A = g(\zeta) + U \tag{3.13}$$

Again, the function $g(\zeta)$ is a twice continuously differentiable nonlinear function in $\boldsymbol{\alpha}$.

Terza et al. (2008) proposed the two-stage residual inclusion (2SRI) method to correct for unmeasured confounding in nonlinear models. The method proposes to estimate the unknown parameter vector $\boldsymbol{\beta}$ in (3.11) in two stages. Let

y_i denote the realizations of Y, z_i and a_i the realizations of the instrument Z and the exposure A, respectively, and $\boldsymbol{x_i} = (1, x_{i1}, \ldots, x_{ip-2}), i = 1, \ldots, n$, the ith realization of $\boldsymbol{X} = (1, X_1, \ldots, X_{p-2})$. Note that $Cov(A, U) \neq 0$ and that the ith realization u_i of U cannot be observed. In the first stage, consistent estimates $\widehat{\boldsymbol{\alpha}}$ of the unknown parameter vector $\boldsymbol{\alpha}$ in (3.13) are obtained using e.g. maximum likelihood estimation (McCullagh and Nelder, 1989). Next, the predicted values $\widehat{a}_i$ of a_i are computed using the ith realization of $\boldsymbol{W}$, denoted as $\boldsymbol{w_i}$, as follows:

$$\widehat{a}_i = g(\widehat{\boldsymbol{\alpha}}^T \boldsymbol{w_i}),\ i = 1, \ldots, n. \tag{3.14}$$

Furthermore, the residuals of (3.13) are defined as

$$\widehat{\varepsilon}_{a_i} = a_i - \widehat{a}_i,\ i = 1, \ldots, n. \tag{3.15}$$

In the second stage, the unobserved confounder u_i in (3.11) is replaced by $\widehat{\varepsilon}_{a_i}$. Note that the actually observed values of a_i are maintained and only the values of the unobserved confounder u_i are replaced by the residuals. Thus, all explanatory variables are exogenous. Consequently, the second stage model is given by

$$Y = f(\boldsymbol{\theta}^T \boldsymbol{S}) + \varepsilon_{2SRI} \tag{3.16}$$

with $\boldsymbol{S} = (1, X_1, \ldots, X_{p-2}, A, \widehat{\varepsilon}_A)$ and $E(\varepsilon_{2SRI}|\boldsymbol{S}) = 0$. Due to the IV assumptions, ε_{2SRI} and $X_1, \ldots, X_{p-2}, A$ as well as ε_{2SRI} and $\widehat{\varepsilon}_A$ are not correlated.

The 2SRI method is a special case of the control function approach described in Section 3.4 with the key assumption in (3.9) restated as $\varepsilon_Y = \rho\varepsilon_A$. Essentially, Terza et al. (2008) assumed that the relation between ε_Y and ε_A is deterministic by setting the random error ϵ to 0. As this assumption rules out that there exits any other cause of the exposure A, that is not directly related to the outcome Y, such as e.g. the instrumental variable Z, it may be unrealistic in observational studies (Tchetgen Tchetgen et al., 2015). The next paragraph shows that consistent estimates for $\boldsymbol{\theta}$ can be obtained, however, it cannot be shown that $\boldsymbol{\beta} = \boldsymbol{\theta}$.

To obtain estimators of the unknown parameter $\boldsymbol{\theta}$, it is assumed that for some $\boldsymbol{\theta_0} \in \Theta$ with $\Theta \subset \mathbb{R}^{p+1}$ it holds that $E(Y|\boldsymbol{S}) = f(\boldsymbol{\theta_0}^T \boldsymbol{S})$. Let $q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}) = f(\theta_0 + \theta_1 X_1 + \ldots + \theta_{p-2} X_{p-2} + \theta_{p-1} A + \theta_p (A - g(\boldsymbol{\alpha^*}^T \boldsymbol{W})))$.

Hence, the value $\boldsymbol{\theta_0}$ of $\boldsymbol{\theta}$ that minimizes the expected squared error between Y and $q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*})$ has to be determined, that is, $\boldsymbol{\theta_0}$ has to fulfill

$$\min_{\boldsymbol{\theta} \in \Theta} E[(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))^2]. \tag{3.17}$$

It is assumed that $\widehat{\boldsymbol{\alpha}}$ converges in probability to $\boldsymbol{\alpha^*}$ for $n \to \infty$, denoted by $\widehat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha^*}$. Wooldridge (2010) showed that $\boldsymbol{\theta_0}$ uniquely solves (3.17) (see Appendix F).

Large sample properties

To prove consistency of $\widehat{\boldsymbol{\theta}}$, the following assumptions have to be fulfilled (Newey and McFadden, 1994):

i.) $E[(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))^2]$ is uniquely minimized by $\boldsymbol{\theta_0}$,

ii.) Θ is compact,

iii.) $(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))$ is continuous on Θ and

iv.) $n^{-1} \sum_{i=1}^{n} (Y_i - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))^2$ uniformly converges in probability to its expected value, that is, $\forall \epsilon > 0$:

$$\lim_{n \to \infty} Pr(\max_{\boldsymbol{\theta} \in \Theta} |\frac{1}{n} \sum_{i=1}^{n} (Y_i - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))^2 - E[(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha^*}))^2]| > \epsilon) = 0.$$

Newey and McFadden (1994) have shown that under the above assumptions it holds that $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta_0}$, where the assumptions are not really crucial: Assumption i.) is shown in Appendix F and Assumption iii.) is fulfilled by definition. The parameter space Θ can be defined as a closed and bounded set, large enough to cover $\boldsymbol{\theta_0}$. Continuity iii.) and compactness ii.) are also necessary to establish uniform convergence. Furthermore, the uniform weak law of large numbers implies that the sample average uniformly converges in probability to its expected value, that means, that Assumption iv.) is fulfilled whenever the weak law of large numbers holds which is formally shown by Bierens (2004).

To prove asymptotic normality, Newey and McFadden (1994) showed that, under certain regularity conditions, the first step estimator affects the asymptotic variance related to the second step if and only if inconsistency in the first step estimation leads to inconsistent estimates in the second step. The key condition is that the score evaluated at $\boldsymbol{\theta_0}$ has expected value zero. Assume that

$\boldsymbol{\theta_0} \in \mathring{\Theta}$, the interior of Θ. Since $q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha})$ is continuously differentiable, the score vector that contains the first-order partial derivatives of $(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha}))$ is given as

$$s(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha})^T = \nabla_{\boldsymbol{\alpha}}(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha})), \tag{3.18}$$

where $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\alpha}$. Two cases need to be distinguished. First, if $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta_0}$, but $\widehat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha_*} \neq \boldsymbol{\alpha_0}$, then $E(\nabla_{\boldsymbol{\alpha}}(Y - q(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}))) = 0$. Second, if there exist an $\boldsymbol{\alpha_*}$ with $\widehat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha_*}$ such that $\widehat{\boldsymbol{\theta}}$ does not converge in probability to $\boldsymbol{\theta_0}$, then $E(\nabla_{\boldsymbol{\alpha}}(Y - q(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}))) \neq 0$. Let $\boldsymbol{H}(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha})$ denote the Hessian matrix of the second-order partial derivatives of $(Y - q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\alpha}))$. Wooldridge (2010) has shown that, in the first case, $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})$ is asymptotically normally distributed with mean 0 and variance

$$\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}, \tag{3.19}$$

where $\boldsymbol{A}_0 \equiv E(\boldsymbol{H}(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}))$ and $\boldsymbol{B}_0 \equiv E(s(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*})s(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*})^T)$ $= Var(s(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}))$. In the second case, the asymptotic variance must be adjusted to account for the first step estimator. Let $r(\boldsymbol{\alpha})$ denote the score vector that contains the first-order partial derivatives of $(A - g(\boldsymbol{\alpha}))$ with respect to $\boldsymbol{\alpha}$ with $E(r(\boldsymbol{\alpha_*})) = 0$. Furthermore, $g(\boldsymbol{\theta_0}, \boldsymbol{\alpha_*})$ is defined as

$$g(\boldsymbol{\theta_0}, \boldsymbol{\alpha_*}) = s(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}) + E(\nabla_{\boldsymbol{\alpha}}(Y - q(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*})))r(\boldsymbol{\alpha_*}). \tag{3.20}$$

Newey and McFadden (1994) showed that $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})$ is asymptotically normally distributed with mean 0 and variance

$$\boldsymbol{A}_0^{-1}\boldsymbol{D}_0\boldsymbol{A}_0^{-1}, \tag{3.21}$$

where $\boldsymbol{D}_0 \equiv E(g(\boldsymbol{X}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*})g(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*})^T) = Var(g(\boldsymbol{S}, \boldsymbol{\theta_0}, \boldsymbol{\alpha_*}))$.
Since no closed-form expression of the estimator $\widehat{\boldsymbol{\theta}}$ can be derived, the Newton-Raphson or the Gauss-Newton method have to be applied to find a solution to the minimization problem. For further details see Seber and Wild (1989), Wooldridge (2010) and Fahrmeir et al. (2007).

Chapter 4

Instrumental variable estimation in time-to-event analysis

Time-to-event analysis focuses on research questions where the time from a well-defined starting point until the occurrence of an event is of interest. IV methods such as the two-stage predictor substitution method and the control function approach described in Chapter 3 are well-established for continuous outcomes using linear regression models. However, these methods cannot be readily applied in time-to-event analysis due to complications that result from censoring and the fact that the assumptions for a valid instrumental variable that may be fulfilled at the start of follow-up might be violated within the risk sets comprising all patients who survived up to a certain time point. Consistent estimation of the marginal causal effect using an instrumental variable with the two-stage predictor substitution method or the control function approach can be achieved in the additive hazards model (Aalen, 1980, 1989). Nevertheless, for the control function approach, very specific assumptions on the structure of the unobserved confounding are necessary, similar to assumptions made for linear models (Section 3.4). Although the Cox proportional hazards model (Cox, 1972) is perhaps the most popular model in time-to-event analysis, analogous results on consistent estimation of either the marginal or the conditional effect can only be achieved in the context of rare outcomes, but interestingly the control function approach requires the same special assump-

tions on the unobserved confounding (Tchetgen Tchetgen et al., 2015).
The fact that estimation of the marginal effect is only justified for rare events can be explained by the non-collapsibility of the hazard ratio. Unlike the hazard ratio, the hazard difference, the exposure coefficient in the Aalen additive hazards model, is a collapsible effect measure and, therefore, the hazard difference measures a causal effect, so that both methods can be applied in the additive hazards model. Here, non-collapsibility means that the conditional and the marginal hazard ratio are unequal if the analysis is adjusted for a factor regardless of whether it is associated with the exposure, hence, it is not a confounder. Since the marginal effect is similar to what we would typically measure in a randomized controlled trial, effects conditional on an unmeasured confounder are not preferred as they are difficult to interpret and they are not what we would obtain from a randomized trial. However, in case of no causal effect of the exposure on the outcome, the marginal and the conditional parameter are zero and in this sense, the hazard ratio is collapsible (Martinussen and Vansteenlandt, 2013). Furthermore, if the outcome is rare, the marginal and the conditional hazard ratio are approximately equal (VanderWeele, 2011). Similar findings are also known for the logistic IV model (Didelez et al., 2010; Robins and Rotnitzky, 2004).
The chapter is organized as follows: first, the problem of non-collapsibility of the hazard ratio will be described and it will be shown that the hazard ratio is collapsible under the null hypothesis of no causal effect. Second, IV methods under the Aalen additive hazards model are introduced, mainly based on Tchetgen Tchetgen et al. (2015). Then, IV methods under the Cox proportional hazards model are presented and it is outlined that only in the situation of a rare outcome and of a null causal effect, a marginal effect can be consistently estimated. This latter case is important as a test of the null hypothesis of no causal exposure effect is then guaranteed to be valid even if the estimator is inconsistent in case of a non-null effect. Finally, a simulation study based on a hypothetical cohort study in a health care database is presented. In this simulation study, the accuracy of estimators derived for the Cox IV models is examined in situations outside the context of rare events or a null causal effect. Furthermore, the robustness of the control function approach regarding the violation of the assumptions for the unmeasured confounder is investigated.

4.1 Notation

Suppose that n individuals are put on study at time 0. Let $\tilde{T}$ be a positive random variable representing the times to event of n individuals, each of which can only be observed in a fixed time interval $[0, C]$ for certain censoring times C. Then $T = \min(\tilde{T}, C)$ denotes the length of follow-up. Let A denote the exposure, Z the instrumental variable and U the unobserved confounder of the effect of A on T. Furthermore, it is assumed that censoring is independent of A and T given Z. Let us also assume that the IV Z fulfills the three core conditions discussed in Section 3.2. Furthermore, $h(t|A, U, Z)$ denotes the hazard function at time t, given A, U and Z, whereas $\tilde{h}(t|Z)$ and $\bar{h}(t|A, Z)$ will denote the hazard function at time t if it depends on Z and (A, Z) only, respectively. For the sake of simplicity, methods will be presented without taking account of measured covariates to focus on the key ideas, but all methods can easily be extended to situations where measured confounders are included.

As described in Chapter 3, the two-stage predictor substitution method as well as the control function approach involve fitting two consecutive regression models. In the following, the first stage model is referred to as the exposure-IV model and the second stage model is called the outcome model.

4.1.1 Outcome models

Throughout this chapter, two models for the outcome T are considered, the Aalen additive hazards model and the Cox proportional hazards model. In the IV context, the former model assumes

$$h(t|A, U, Z) = b_0(t) + b_A(t)A + b_U(U, t), \tag{4.1}$$

where the functions $b_0(\cdot), b_A(\cdot), b_U(\cdot, \cdot)$ are unrestricted. The function $b_U(\cdot, \cdot)$ describes the association between the unobserved confounder and the outcome and is allowed to remain unrestricted at each time point t and across time points. The hazards difference for the effect of A on T is denoted by $b_A(t)$ and when integrating out U, $b_A(t)$ is still the same parameter. Since the

hazards difference is collapsible over U, it is interpretable as a marginal causal parameter (Martinussen and Vansteenlandt, 2013).

For an IV situation, the Cox proportional hazards model assumes

$$h(t|A, U, Z) = h_0(t) \exp(\theta_A A + b_U(U, t)), \tag{4.2}$$

where $h_0(t)$ is the baseline hazard function and θ_A denotes the log-hazard ratio for the effect of A on T given U. When U is integrated out, θ_A does not describe the marginal $A - T$ relation, even if A is independent of U. Since the hazard ratio is non-collapsible, the parameter θ_A can only be interpreted conditional on U.

Both models explicitly assume that the outcome T and the IV Z are conditionally independent given U and A, as the left-hand side of model (4.1) and model (4.2) condition on Z, although the right-hand side does not include Z. Moreover, both assume no interaction between A and U.

4.1.2 Exposure-IV models

For the exposure-IV models, it is distinguished between a continuous and a binary exposure, as this has important implications for the residual term.
First, when the exposure A is continuous, the linear model for A is

$$A = c_0 + c_Z Z + \varepsilon_A, \tag{4.3}$$

where ε_A is a mean zero random error with $Cov(U, \varepsilon_A|Z) \neq 0$.
Second, when the exposure A is binary, the following logistic regression model is assumed

$$\text{logit}\, Pr(A = 1|Z) = c_0 + c_Z Z. \tag{4.4}$$

For both exposure-IV models, M is defined as $M = E(A|Z)$. Then we have in (4.3) that $M = c_0 + c_Z Z$ and in (4.4) that $M = \text{expit}(c_0 + c_Z Z)$. Note that while (4.3) models the value of A and, hence, includes an explicit residual ε_A, this is not the case for (4.3). As it is assumed that the core conditions for Z are fulfilled, it holds that $c_Z \neq 0$.

4.2 Non-collapsibility in the Cox proportional hazards model

In contrast to the hazard difference in the Aalen additive hazards model, the exposure effect θ_A in (4.2) is not equal to the marginal causal exposure effect $\tau(t)$ due to the non-collapsibility of the hazard ratio. This situation is uncomfortable as we would come up with a parameter that can only be interpreted conditional on U and this conditional parameter does not equal the effect we would measure in a randomized controlled trial, the marginal causal effect. Under the null hypothesis of no causal effect, both, the marginal and the conditional parameter are zero and in this sense, the hazard ratio can be shown to be collapsible.

Let us consider the Cox model given in (4.2). Martinussen and Vansteenlandt (2013) showed that the marginal exposure effect $\tau(t)$ for the proportional hazards model is then given by

$$\tau(t) = \theta_A + \log\left(\frac{g(A=1,\theta_A,b_U(U,t),h_0(t))}{g(A=0,\theta_A,b_U(U,t),h_0(t))}\right),$$

where

$$g(A=a,\theta_A,b_U(U,t),h_0(t)) = \frac{E(S_1(A=a,\theta_A,b_U(U,t),h_0(t)))}{E(S_0(A=a,\theta_A,b_U(U,t),h_0(t)))}.$$

and

$$S_0(A=a,\theta_A,b_U(U,t),h_0(t)) = \exp(-\int_0^t [h_0(s)\exp(\theta_A a + b_U(U,s))])\,\mathrm{d}s)$$

$$S_1(A=a,\theta_A,b_U(U,t),h_0(t)) = \exp(-\int_0^t [h_0(s)\exp(\theta_A a + b_U(U,s))])\,\mathrm{d}s)$$
$$\exp(b_U(U,t)).$$

To obtain equality of the conditional and the marginal effect, $\tau(t) = \theta_A$, the following equation needs to be fulfilled:

$$\frac{g(A=1,\theta_A,b_U(U,t),h_0(t))}{g(A=0,\theta_A,b_U(U,t),h_0(t))} = 1.$$

Therefore, we need:

$$g(A=1,\theta_A,b_U(U,t),h_0(t)) = g(A=0,\theta_A,b_U(U,t),h_0(t)) \tag{4.5}$$

$$\iff \frac{E(S_1(A=1,\theta_A,b_U(U,t),h_0(t)))}{E(S_0(A=1,\theta_A,b_U(U,t),h_0(t)))} = \frac{E(S_1(A=0,\theta_A,b_U(U,t),h_0(t)))}{E(S_0(A=0,\theta_A,b_U(U,t),h_0(t)))}$$

$$\iff \frac{E(\exp(-\int_0^t [h_0(s)\exp(\theta_A A + b_U(U,s))])\,\mathrm{d}s)\exp(b_U(U,t)))}{E(\exp(-\int_0^t [h_0(s)\exp(\theta_A A + b_U(U,s))])\,\mathrm{d}s))}$$

$$= \frac{E(\exp(-\int_0^t [h_0(s)\exp(b_U(U,s))])\,\mathrm{d}s)\exp(b_U(U,t)))}{E(\exp(-\int_0^t [h_0(s)\exp(b_U(U,s))])\,\mathrm{d}s))}.$$

The last equality is fulfilled if and only if $\theta_A = 0$, that means, under the null hypothesis of no conditional causal effect of the exposure A on the outcome T, the marginal and the conditional effect are equal and the hazard ratio measures a causal effect. As $\tau(t)$ is a smooth function of θ_A, we have that when θ_A is close to zero, $\tau(t)$ must also be close to zero. Informally, we can say that the collapsibility holds near the null.

The fact that the hazard ratio is collapsible in the situation of no causal effect can now be exploited to obtain a marginal hazard ratio derived from IV methods such as the two-stage predictor substitution method and the control function approach.

4.3 Instrumental variable estimation under an additive hazards model

Before investigating the two-stage predictor substitution (2SPS) method and the control function (CF) approach for the proportional hazards model, these methods are first discussed for the additive hazards model, since very specific assumptions about the structure of the unmeasured confounding for the control function approach are also needed for IV estimation using the Cox model. These assumptions are subtly different for continuous and binary exposure.
Basically, a function $\tilde{h}(t|Z)$ of M for the 2SPS method and a function $\bar{h}(t|A,Z)$

of A and ε_A for the CF approach, respectively, are needed which include the target parameter $b_A(t)$ in a way such that consistent estimators can be derived. In the following sections, $\tilde{h}(t|Z)$ and $\bar{h}(t|A,Z)$ will be therefore considered under different assumptions.

4.3.1 Two-stage predictor substitution method

Again, as already described in Section 3.3, the idea of the the 2SPS method is to insert the first stage model (4.3) into the second stage model (4.1). Under the IV assumptions, Tchetgen Tchetgen et al. (2015) showed that when integrating out A and U, we obtain

$$\tilde{h}(t|Z) = \tilde{b_0}(t) + b_A(t)M. \tag{4.6}$$

The proof by Tchetgen Tchetgen et al. (2015) basically involves relegating the unmeasured confounder U and ε_A into the baseline hazard function, such that $\tilde{b_0}$ is the modified baseline hazard function. $b_A(t)$, the marginal causal effect of A on T in (4.1) can now be consistently estimated.

As described in Section 3.3, two steps are needed to estimate $b_A(t)$. For this purpose, M is first consistently estimated by $\hat{M}$. Second, M is replaced by $\hat{M}$ in the second stage model (4.6).

4.3.2 Control function approach

As already outlined in Section 3.4, the idea of the control function approach is that the residual ε_A in the exposure model captures some of the variation of the hazard function due to unobserved confounders. For this approach, a submodel of (4.1) that further specifies the influence of U on the hazard is assumed that will be called the confounder model. This model depends on the scale of the exposure, hence, we consider binary and continuous exposures, separately.

Confounder models

For a continuous exposure A, the confounder model describes the association between the unmeasured confounder U and the random error ε_A in the linear exposure model (4.3) and is given by

$$b_U(U,t) = \rho_0(t)\varepsilon_A + \epsilon(t), \tag{4.7}$$

where $\rho_0(t)$ is an unknown and unrestricted function with $\rho_0(t) \neq 0$. To allow for a non-deterministic association between U and ε_A, $\epsilon(t)$ is a random error which need not have mean zero and for which $\epsilon(t) \perp\!\!\!\perp (\varepsilon_A, Z)$ holds.

For a binary exposure A described by the logistic exposure model (4.4), the residual ε_A would be determined by $\varepsilon_A = A - \text{expit}(c_0 + c_Z Z)$. Assuming (4.7) for a binary exposure would violate the IV assumption that Z is independent of U. To maintain the independence of Z and U, the following confounder model is assumed

$$b_U(U,t) = E(b_U(U,t)|A,Z) + \epsilon(t), \tag{4.8}$$

where $\epsilon(t)$ is an independent mean zero error. In case that $b_U(U,t) = b^*(t)U$ is linear in U, this assumption results in a location shift model for the density of U conditional on A and Z, so that U and (A,Z) are associated only on the mean scale.

Control function approach for a continuous exposure

Assuming a continuous exposure and that the association between U and ε_A is given by the confounder model (4.7), we obtain by integrating out U

$$\bar{h}(t|A,Z) = \bar{b}_0(t) + b_A(t)A + \rho_0(t)\varepsilon_A. \tag{4.9}$$

In deriving (4.9), the confounder model (4.7) is useful as it allows the random error $\epsilon(t)$ to be relegated to the baseline hazard function, so that $\bar{b}_0(t)$ is the modified baseline hazard function, while retaining $b_A(t)$ as coefficient fo A (Tchetgen Tchetgen et al., 2015). Under assumption (4.7), the marginal causal effect of A on T, $b_A(t)$, can be consistently estimated.

Control function approach for a binary exposure

Assuming (4.8) for binary exposure and that Z is binary, Tchetgen Tchetgen et al. (2015) showed that we obtain

$$\bar{h}(t|A,Z) = \tilde{b}_0(t) + b_A(t)A + (\rho_0(t) + \rho_1(t)Z)\varepsilon_A, \tag{4.10}$$

where $\tilde{b}_0(t)$ is the baseline hazard function. For a continuous instrument Z, $E(b_U(U,t)|A,Z)$ needs to be linear in Z to consistently estimate the causal effect $b_A(t)$.

As described in Section 3.4, estimation of both models using the control function approach is performed in two steps and analogously for continuous and binary exposures. First, the residual $\widehat{\varepsilon}_A = A - \hat{M}$ is obtained from the respective first stage exposure model. The error terms ε_A in (4.9) and (4.10) are then replaced by their estimates $\widehat{\varepsilon}_A$, respectively.

4.4 Instrumental variable estimation under a proportional hazards model

Due to the non-collapsibility of the Cox model, analogous results on consistent estimation can only be achieved if there is no causal effect of the exposure on the outcome, since then the marginal and conditional effect are equal, or if the outcome is rare, since the marginal effect and the conditional effect are then approximately equal. Interestingly, the control function approach requires the same special assumptions on the unobserved confounding as under the additive hazards model.

4.4.1 Estimation under the rare-outcome condition

Let us consider an outcome that is rare over the follow-up period such that only very few events can be observed. Then, the conditional survival function

is near unity, that means,

$$S(t|A,U,Z) \approx 1 \tag{4.11}$$

for all t during follow-up. The conditional hazard function of model (4.2) is defined as

$$h(t|A,U,Z) = \frac{f(t|A,U,Z)}{S(t|A,U,Z)},$$

where $f(t|A,U,Z)$ denotes the density of T, given A,U,Z. Under the rare-outcome condition (4.11), it follows:

$$f(t|A,U,Z) = h(t|A,U,Z)S(t|A,U,Z) \approx h(t|A,U,Z). \tag{4.12}$$

Likewise, it follows from (4.11):

$$E(f(t|A,U,Z)|Z) = f(t|Z) = \tilde{h}(t|Z)S(t|Z) \approx \tilde{h}(t|Z). \tag{4.13}$$

Replacing $f(t|A,U,Z)$ by $h(t|A,U,Z)$ according to (4.12) in (4.13) yields:

$$\tilde{h}(t|Z) \approx E(f(t|A,U,Z)|Z) \approx E(h(t|A,U,Z)|Z).$$

Two-stage predictor substitution method for a continuous exposure

Let us assume a continuous exposure and model (4.3). According to Tchetgen Tchetgen et al. (2015), the proportional hazards model (4.2) can then be rewritten by replacing the exposure A by $M + \varepsilon_A$:

$$\begin{aligned}\tilde{h}(t|Z) &\approx E(h(t|A,U,Z)|Z) \\ &= E(h_0(t)\exp(\theta_A A + b_U(U,t))|Z) \\ &= E(h_0(t)\exp(\theta_A M + \theta_A\varepsilon_A + b_U(U,t))|Z).\end{aligned}$$

Using $M = c_0 + c_Z Z$, $U \perp\!\!\!\perp Z$ and $\varepsilon_A \perp\!\!\!\perp Z$, it follows

$$\begin{aligned}\tilde{h}(t|Z) &\approx E[h_0(t)\exp(\theta_A M) * \exp(\theta_A\varepsilon_A + b_U(U,t))] \\ &= h_0(t)\exp(\theta_A M) * E[\exp(\theta_A\varepsilon_A + b_U(U,t))] \\ &= h_0^*(t)\exp(\theta_A M),\end{aligned} \tag{4.14}$$

where $h_0^*(t) = h_0(t)E(\exp(\theta_A\varepsilon_A + b_U(U,t)))$ is the modified baseline hazard function. If $S(t|A,U,Z)$ is close to one, (4.14) gives a justification for using the 2SPS method with a proportional hazards model, as the marginal parameter θ_A remains the target parameter in a Cox regression of T on $\hat{M}$.

Control function approach for a binary exposure

Using a similar argument as in (4.12) and (4.13), it can be shown that:

$$\bar{h}(t|A,Z) \approx E(f(t|A,U,Z)|A,Z) \approx E(h(t|A,U,Z)|A,Z).$$

Let us now assume a binary exposure A and a binary IV Z. The proportional hazards model can then be rewritten as

$$\begin{aligned}\bar{h}(t|A,Z) &\approx E[h(t|A,U,Z)|A,Z] \\ &= E[h_0(t)\exp(\theta_A A + b_U(U,t))|A,Z] \\ &= E[h_0(t)\exp(\theta_A A)\exp(b_U(U,t))|A,Z] \\ &= h_0(t)\exp(\theta_A A)E[\exp(b_U(U,t))|A,Z].\end{aligned}$$

Using the confounder model (4.8), it follows:

$$\begin{aligned}\bar{h}(t|A,Z) &\approx h_0(t)\exp(\theta_A A)E[\exp(E(b_U(U,t)|A,Z) + \epsilon(t))|A,Z] \\ &= h_0(t)\exp(\theta_A A)E[\exp(E(b_U(U,t)|A,Z))\exp(\epsilon(t))|A,Z].\end{aligned}$$

Since (A,Z) and $\epsilon(t)$ are assumed to be independent, this can be rewritten as:

$$\bar{h}(t|A,Z) \approx h_0(t)\exp(\theta_A A)\exp(E(b_U(U,t)|A,Z))E(\exp(\epsilon(t))).$$

Tchetgen Tchetgen et al. (2015) showed that

$$\begin{aligned}E(b_U(U,t)|A,Z) =&(\rho_0(t) + \rho_1(t)Z)(A - Pr(A=1|Z)) + E(b_U(U,t)) \\ &\text{where} \\ \rho_0(t) =&E(b_U(U,t)|A=1,Z=0) - E(b_U(U,t)|A=0,Z=0) \\ \rho_1(t) =&E(b_U(U,t)|A=1,Z=1) - E(b_U(U,t)|A=0,Z=1) \\ &- E(b_U(U,t)|A=1,Z=0) + E(b_U(U,t)|A=0,Z=0),\end{aligned}$$

which yields

$$\begin{aligned}\bar{h}(t|A,Z) &\approx h_0(t)\exp(\theta_A A + (\rho_0(t) + \rho_1(t)Z)\varepsilon_A)\exp(E(b_U(U,t)))E(\exp(\epsilon(t))) \\ &= h_0^{**}(t)\exp(\theta_A A + (\rho_0(t) + \rho_1(t)Z)\varepsilon_A)\end{aligned}$$

with $h_0^{**}(t) = h_0(t)\exp(E(b_U(U,t)))E(\exp(\epsilon(t)))$. If $\rho_0(t)$ and $\rho_1(t)$ are assumed to be constant, i.e. $\rho_0(t) = \rho_0$ and $\rho_1(t) = \rho_1$, the standard proportional

hazards model is obtained and θ_A, the approximated marginal and conditional causal effect, can be consistently estimated.

In both cases, θ_A is then estimated by a proportional hazards regression of T on A replacing ε_A by $\widehat{\varepsilon}_A$ which is derived from the first stage model. In case of a binary exposure, an interaction between the instrument and the first stage residual is additionally included.

4.4.2 Estimation under the assumption of no causal effect

Section 4.2 shows that under the null hypothesis of no causal effect, the conditional and the marginal parameter are zero and in this sense, the hazard ratio is collapsible. Under the IV assumptions and under the null hypothesis of no exposure effect ($\theta_A = 0$), the outcome T and the instrument Z are independent and, therefore, T and $M = c_0 + c_Z Z$ are also independent. Per definition, U and Z are also independent, so that the association between T and Z is not confounded by U. Employing the 2SPS method, the proportional hazards model of T on $\hat{M}$ consistently estimates the null exposure effect when the null hypothesis of no causal effect is true and in this case, the marginal and the conditional effect coincide (Section 4.2). Therefore, the 2SPS estimator is robust under the null hypothesis. In contrast, when using the control function approach, the exposure A remains in the model, so that a similar argument can not be used for the control function approach.

4.4.3 Simulation study

Objective

This simulation study aims to assess the accuracy of IV estimators obtained from the two-stage predictor substitution method and the control function approach under the proportional hazards model for binary exposure, which is the typical situation in pharmacoepidemiological studies. Due to the non-collapsibility of the hazard ratio, both IV methods may only be used under a

proportional hazards model in situations with a rare outcome (Section 4.4.1). In addition, the 2SPS estimator consistently estimates the exposure effect under the null hypothesis of no casual effect (Section 4.4.2). Therefore, this simulation study aims to also investigate the performance of both methods in situations where these assumptions do not hold. First, scenarios with common outcomes and with a non-null causal effect are examined, since the 2SPS estimator is likely to have small bias when the true exposure effect is close to but not equal to zero. Since censoring introduces a loss of information that could lead to inefficient estimation of the exposure effect, scenarios with additional censoring are compared to scenarios with no censoring (administrative censoring only). Second, the models for the unmeasured confounder (Section 4.3.2) are a necessary prerequisite for the control function approach to yield a collapsible hazard ratio. However, these assumptions are not empirically verifiable and are not expected to hold in practice. Therefore, the unmeasured confounder is generated irrespectively of these additional assumptions, but it is simulated based on settings that are typically expected in pharmacoepidemiological studies. Third, the robustness of both methods regarding the misspecification of the first stage model by fitting a linear model for the binary exposure is examined. Additionally, it is investigated if the prevalence of the binary exposure is related to the impact of misspecification on the accuracy of estimators. Last, it is examined if the performance of both methods is poor for weak instruments, since weak instruments lead to decreased efficiency and biased estimators (Martens et al., 2006).

General design and assumptions

A hypothetical cohort study embedded in a claims database is simulated to explore the association between a binary fixed-in-time exposure (drug A vs. B) and time to an adverse event as outcome. Patients included in the cohort are followed for up to 365 days. For simplicity, it is assumed that both the exposure and the outcome are affected by only three patient's characteristics that serve as potential confounders. One of the three confounders is assumed to be not recorded in the study database and, thus, represents the unmeasured confounder. As outlined in the section above,

i.) the size of the effect of exposure on the outcome,
ii.) the probability of survival at the end of follow-up,
iii.) the type of censoring,
iv.) the prevalence of drug A, and
v.) the strength of the association between the exposure and the IV

will be varied across the simulated scenarios.

A common example for an IV in pharmacoepidemiological studies is the physician's preference that can be defined in different ways, for example, as a continuous IV based on the proportion of all previous prescriptions of the same physician (Brookhart et al., 2006; Kollhorst et al., 2016). In order to mimic a realistic IV used in pharmacoepidemiological studies, a continuous physician's prescribing preference-based IV is constructed, which is then used to account for unmeasured confounding at the stage of data analyses. As the proportion of all previous prescriptions of the same physician ranges between 0 and 1, the continuous instrument Z is generated from a truncated normal distribution. The distribution of the confounding variables and their association with the exposure and the outcome are chosen based on situations typically observed in pharmacoepidemiological studies (Table 4.1). All three confounders are generated independently, which implies that they are not correlated. Furthermore, it is assumed that all IV assumptions are fulfilled including no associations between the IV and any of the three confounders.

The binary exposure A is simulated from a Bernoulli distribution with the probability of receiving drug A, conditional on the IV and the measured and unmeasured confounders and is defined by the following multivariable logistic model:

$$\text{logit}\, Pr(A = 1|Z, X_1, X_2, U) = 0.0 + \beta_Z Z + \beta_{X_1} X_1 + \beta_{X_2} X_2 + 0.4U, \quad (4.15)$$

where β_Z denotes the effect of the continuous instrumental variable on the exposure. To assess the impact of the strength of the association between the instrument and the exposure, the values of β_Z are varied across scenarios to simulate a strong (partial $r^2 \approx 0.04$) and a weak association (partial $r^2 \approx 0.003$ to 0.004) between the exposure and the instrument, respectively (Table 4.1). Furthermore, the values of β_{X_1} and β_{X_2} are varied to obtain a

prevalence of drug A of 60% and 80%, respectively, to examine the impact of misspecification of the first stage model if a linear model is used to fit the binary exposure. To further evaluate the robustness of the control function approach regarding the functional form of the unmeasured confounder, a sensitivity analysis by adding a mean random error ϵ with $\epsilon \sim \mathcal{N}(0, 0.2)$ to the linear predictor of model (4.15) is conducted, so that the unmeasured confounder is given as $\tilde{U} = U + \epsilon$. The sensitivity analysis is only conducted in scenarios with a strong instrument, a prevalence of drug A of 60% and with only administrative censoring where IV estimators are expected to be less biased than in all other scenarios.

The outcome is defined as the time to an event T and is generated, conditional on the exposure and the measured and unmeasured confounders, but not on the IV. Specifically, event times are assumed to arise from an exponential distribution with a hazard defined by the following Cox proportional hazards model:

$$h(\tilde{t}|A, X_1, X_2, U) = h_0(\tilde{t}) \exp(\theta_A A + \log(0.8)X_1 + \log(1.1)X_2 + \log(0.4)U), \tag{4.16}$$

where θ_A denotes the effect of exposure on the outcome and $h_0(\tilde{t})$ denotes the baseline hazard that is varied to simulate, first, a survival probability of $S(t) \approx 0.97$ to 0.98 at the end of follow-up ($t = 365$), and, second, a survival probability of $S(t) \approx 0.55$ at the end of follow-up. Censoring times C are simulated, first, with only administrative censoring at the end of follow-up ($C = 365$ days after cohort entry), and second, with additional censoring before the end of follow-up. Censoring times are then assumed to arise from an exponential distribution with a varying censoring rate λ to obtain a survival probability of 0.97 to 0.98 and 0.55, respectively. The observed survival time T_i of each patient i is then defined as $T_i = \min(\tilde{T}_i, C_i)$ and the status is determined as event ($\delta_i = 1$) if $\tilde{T}_i < C_i$ or censored ($\delta_i = 0$) if $\tilde{T}_i \geq C_i$.

Model (4.16) implies that event times are generated from a conditional model such that the exposure effect θ_A is specified as being conditional on X_1, X_2, and U. Due to the non-collapsibility of the hazard ratio, conditional and marginal hazard ratios are generally not equal. To obtain a marginal estimator of the causal effect that will then be compared with the IV estimators,

the following approach is used. First, a very large cohort of $500,000$ patients is generated using the same methods and assumptions as described above with the only exception that the exposure $A \sim \mathcal{B}(1, 0.5)$ is simulated independently of any covariates, so that X_1, X_2 and U are not confounders. Second, event times are generated based on A, X_1, X_2 and U as defined above with administrative censoring only. The marginal exposure effect τ_A is then obtained by fitting a Cox model that adjusts for X_1 and X_2.

The details of the simulation design, including the distributions of the measured and unmeasured variables, their associations, and alternative values of the relevant parameters, varied across different scenarios, are summarized in Table 4.1.

Table 4.1: Simulation design

	Scenario
Instrumental variable Z	$Z \sim \mathcal{N}_{0.01,0.99}(0.5, 0.2)$
Confounders $X_1, X_2, U, \tilde{U}$	
Measured	$X_1 \sim \mathcal{B}(1, 0.25)$, $X_2 \sim \mathcal{N}(0.5, 1/12)$
Unmeasured	
Main simulation	$U \sim \mathcal{B}(1, 0.3)$
Sensitivity analysis	$\tilde{U} = U + \epsilon$ with $\epsilon \sim \mathcal{N}(0, 0.2)$
Exposure A	$A \sim \mathcal{B}(1, p)$
	with $p = Pr(A = 1\|Z, X_1, X_2, U)$
Strength of the association between Z and A	
Strong	Partial $r^2 \approx 0.04$
Weak	Partial $r^2 \approx 0.004$ to 0.003
Effect β_Z of IV on exposure	
Prevalence 60%	$\beta_Z = 2, 10$ (weak/strong)
Prevalence 80%	$\beta_Z = 2.2, 16$ (weak/strong)
Event times $\tilde{T}$	$\tilde{T} \sim Exp(1/h(\tilde{t}\|A, X_1, X_2, U))$
	with $h(\tilde{t}\|A, X_1, X_2, U)$
	as defined in (4.16)
Censoring times C	
Administrative censoring	$C = 365$
Additional censoring	$C \sim Exp(1/\lambda)$ with
	$\lambda \in [0,00007; 0,0008; 0,0055; 0,007]$
Effect θ_A of exposure on outcome	$\theta_A = \log(1), \log(2)$
Survival probability $S(t)$	
Rare outcome	$S(t) \approx 0.97$ to 0.98
Common outcome	$S(t) \approx 0.55$

Methods

For each scenario, 1,000 independent random samples with a sample size of $30,000$ patients are generated and results are summarized. For each simulated

sample, in each scenario, the adjusted hazard ratio (HR) for the exposure effect is estimated separately using seven multivariable proportional hazards models that are all adjusted for X_1 and X_2:

i.) Oracle HR=exp($\hat{\theta}_A$): exposure effect estimated by the conventional Cox model that additionally adjusts for U,
ii.) Marginal HR=exp($\hat{\tau}_A$): exposure effect estimated by the marginal model (unadjusted for U),
iii.) Cox HR=exp($\hat{\theta}_A$): exposure effect estimated by the conventional Cox model unadjusted for U,
iv.) Lin. 2SPS HR=exp($\hat{\theta}_A$): exposure effect estimated by the 2SPS method with a first stage linear model (Section 4.4),
v.) Log. 2SPS HR=exp($\hat{\theta}_A$): exposure effect estimated by the 2SPS method with a first stage logistic model (Section 4.4),
vi.) Lin. CF HR=exp($\hat{\theta}_A$): exposure effect estimated by the CF approach based on a first stage linear model (Section 4.4),
vii.) Log. CF HR=exp($\hat{\theta}_A$): exposure effect estimated by the CF approach based on a first stage logistic model (Section 4.4).

As a measure of the strength of the association between the IV and the exposure, the squared partial Spearman correlation coefficient r^2 is calculated, based on the first stage linear model (Rassen et al., 2009a,b). For each simulated scenario, the exposure effect estimators of the seven models are compared in terms of mean hazard ratios and empirical standard deviation. In addition, estimators based on the 2SPS method and the CF approach are compared in a real-life study on the difference in mortality risks between users of conventional and atypical antipsychotics using the physician's preference as an instrumental variable in GePaRD.

Results of the simulation study

Rare outcome

Table 4.2 compares the mean hazards ratios for the seven proportional hazards models. As expected, the estimators of the conventional Cox model are

Table 4.2: Mean hazard ratios for the exposure effect for a rare outcome

Partial r^2	Drug prevalence	Censoring	Oracle HR	Marginal HR	Cox HR	Lin. 2SPS HR	Log. 2SPS HR	Lin. CF HR	Log. CF HR
0.041	60%	W/o	1.00	1.00	0.87	1.04	1.04	1.04	1.04
		With	1.00	1.00	0.87	1.06	1.06	1.06	1.06
0.039	80%	W/o	1.01	1.00	0.87	1.06	1.06	1.07	1.07
		With	1.00	1.00	0.87	1.10	1.10	1.10	1.11
0.004	60%	W/o	1.00	1.00	0.86	1.59	1.58	1.59	1.59
		With	1.00	1.00	0.87	1.79	1.78	1.80	1.79
0.003	80%	W/o	1.01	1.00	0.86	2.61	1.92	2.64	1.93
		With	1.00	1.00	0.86	3.22	2.14	3.32	2.17
0.041	60%	W/o	2.00	2.00	1.74	2.02	2.03	2.11	2.12
		With	2.01	2.00	1.75	2.08	2.09	2.18	2.19
0.039	80%	W/o	2.02	2.00	1.75	1.99	2.01	2.19	2.20
		With	2.02	2.00	1.74	2.07	2.11	2.27	2.30
0.004	60%	W/o	2.00	2.00	1.73	4.04	4.01	4.34	4.32
		With	2.01	2.00	1.74	4.49	4.46	4.85	4.83
0.003	80%	W/o	2.02	2.00	1.74	6.27	4.32	7.30	4.87
		With	2.01	2.00	1.73	10.31	5.37	11.43	5.88

w/o: without additional censoring (administrative censoring only).

biased due to the unmeasured confounder. The exposure effect is underestimated by about 13%, when compared to the oracle and the marginal hazard ratio, respectively (Table 4.2). In contrast, the four IV models using the 2SPS method and the CF approach overestimate the exposure effect. Due to loss of information during follow-up, estimators obtained from the two-stage predictor substitution method and the control function approach are generally more biased in scenarios with additional censoring than in scenarios with administrative censoring only (Figure 4.1). In scenarios targeting no causal effect with an IV being strongly associated with the exposure (partial $r^2 \approx 0.04$), 2SPS and CF estimators are comparable. Please note that the bias is slightly more pronounced in scenarios with a prevalence of drug A of 80% (6-11%) compared with a prevalence of drug A of 60% (4-7%). In scenarios with a

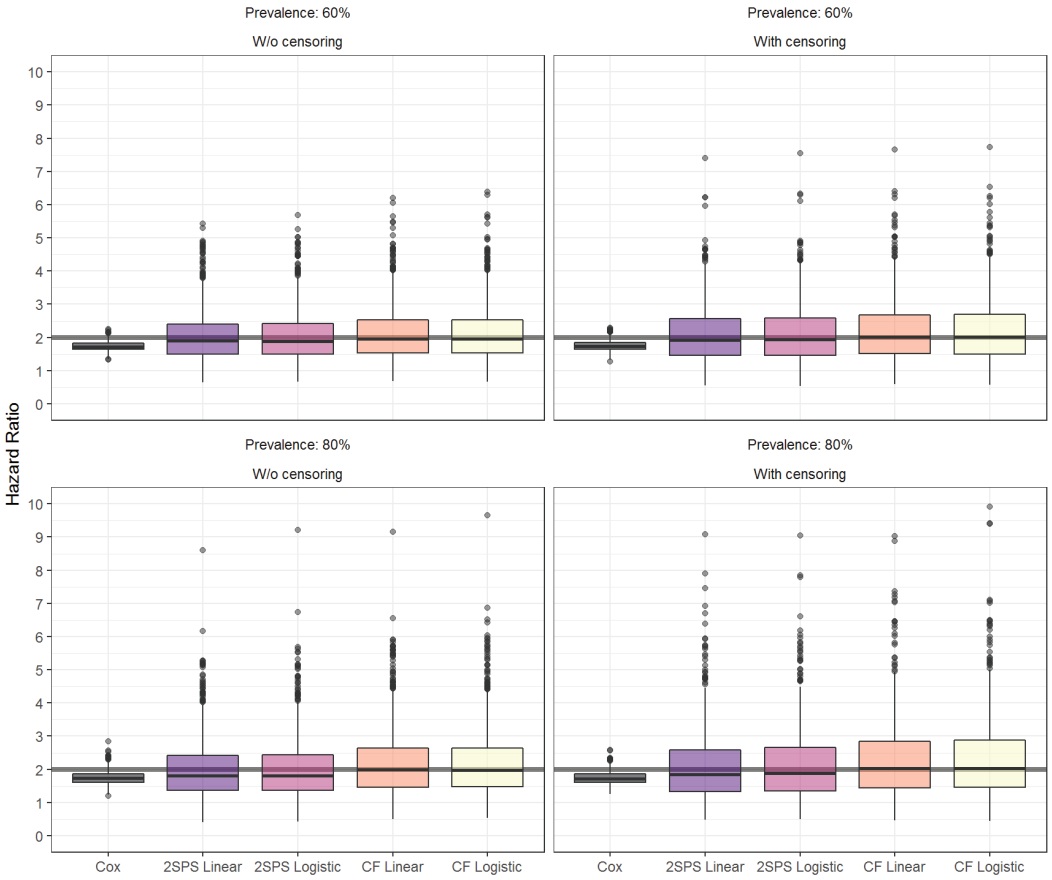

Figure 4.1: Boxplots of estimators for a strong instrument, a rare outcome and a non-null causal effect

marginal hazard ratio of 2 and a strong IV, the IV models using the 2SPS method (0.5-5.5%) outperform the models using the CF approach (5.5-15%). Bias of the CF estimators is more pronounced in situations where the exposure ratio is 80%. In general, bias of IV estimators further increases in scenarios using the weak instrument (partial $r^2 \approx 0.004$) with slightly less bias for the estimators using a first stage logistic model.

The bias reduction of the IV estimators comes at the cost of an increased variance (data not shown). The estimators of the conventional Cox model have a smaller empirical standard deviation than the estimators of the four IV models which becomes obvious from the boxplots in Figure 4.1. In general, 2SPS models lead to a smaller empirical standard deviation than the CF models. Only for the weak instrument, the 2SPS as well as the CF model with a first stage

Table 4.3: Mean hazard ratios for the exposure effect for a common outcome

Partial r^2	Drug prevalence	Censoring	Oracle HR	Marginal HR	Cox HR	Lin. 2SPS HR	Log. 2SPS HR	Lin. CF HR	Log. CF HR
0.041	60%	W/o	1.00	1.00	0.87	1.00	1.00	1.00	1.00
0.041		With	1.00	1.00	0.88	1.00	1.00	1.00	1.00
0.039	80%	W/o	1.00	1.00	0.87	1.00	1.00	1.00	1.00
0.039		With	1.00	1.00	0.87	1.01	1.01	1.01	1.01
0.004	60%	W/o	1.00	1.00	0.87	1.02	1.02	1.02	1.02
0.004		With	1.00	1.00	0.88	1.04	1.04	1.04	1.04
0.003	80%	W/o	1.00	1.00	0.87	1.05	1.03	1.06	1.04
0.003		With	1.00	1.00	0.87	1.08	1.05	1.09	1.06
0.041	60%	W/o	2.00	1.96	1.71	1.90	1.91	1.95	1.95
0.041		With	2.00	1.96	1.71	1.92	1.92	1.95	1.95
0.039	80%	W/o	2.00	1.96	1.70	1.86	1.87	1.95	1.95
0.039		With	2.00	1.96	1.70	1.89	1.90	1.97	1.98
0.004	60%	W/o	2.00	1.96	1.71	1.94	1.94	1.98	1.98
0.004		With	2.00	1.96	1.70	2.00	1.99	2.03	2.03
0.003	80%	W/o	2.00	1.96	1.70	1.97	1.93	2.05	2.01
0.003		With	2.00	1.96	1.69	2.06	1.98	2.14	2.06

w/o: without additional censoring (administrative censoring only).

logistic model yield a smaller standard deviation than the other two models. For all IV models, empirical standard deviation increases in scenarios with a weak instrument, an oracle hazard ratio of 2, additional censoring and for a prevalence of drug A of 80%.

Common outcome

Again, the estimators of the conventional Cox model are biased due to the unmeasured confounder, such that the exposure effect is underestimated by about 13%, when compared to the marginal hazard ratio, respectively (Table 4.3). In scenarios with a null causal effect, IV estimators are comparable and unbiased if a strong IV is used and are slightly biased (2-9%) for a weak IV. Due to the non-collapsibility of the hazard ratio, the oracle and the marginal

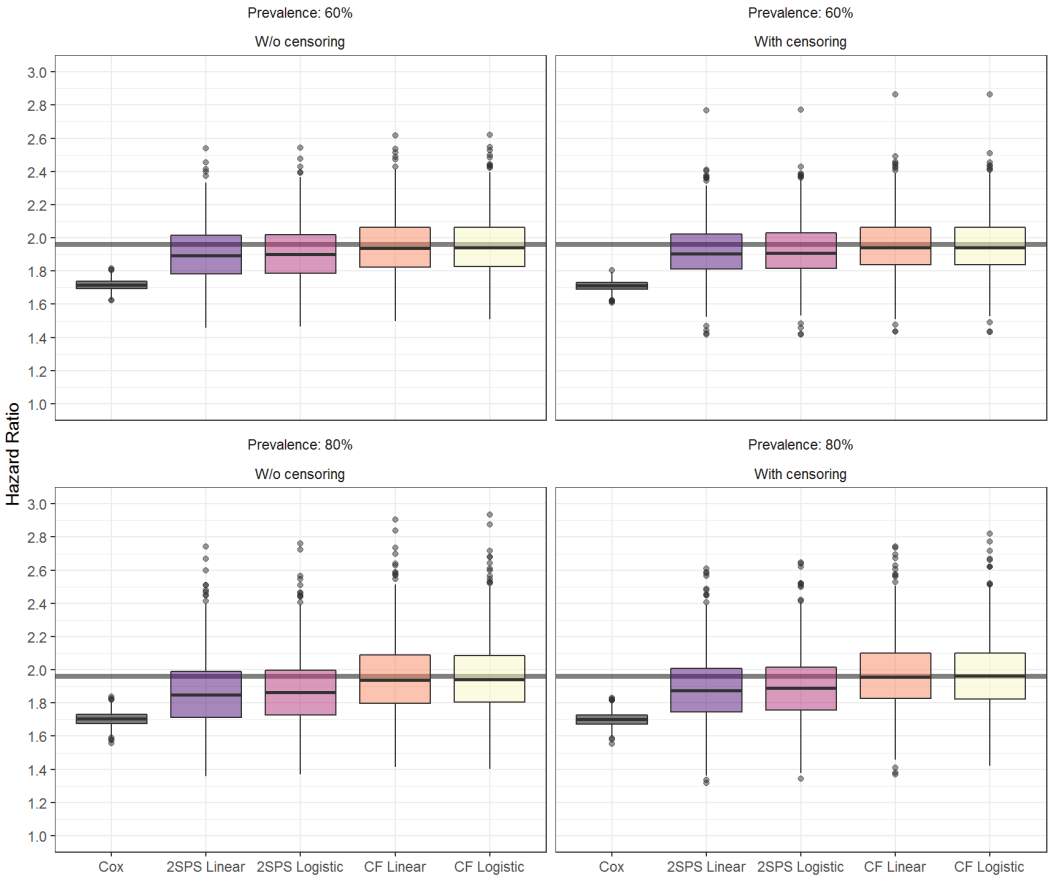

Figure 4.2: Boxplots of estimators for a strong instrument, a common outcome and a non-null causal effect

effect estimates are not equal in scenarios with a non-null causal effect (see lower part of Table 4.3). In contrast to the results for the rare outcome, the four IV models mostly underestimate the marginal effect for a strong IV. The CF estimators are less biased (0.5-1%) than the 2SPS estimators (2-5%) for a strong IV-exposure association and vice versa for a weak association, especially for a prevalence of drug A of 80%. In general, bias of IV estimators further increases in scenarios using the weak instrument (partial $r^2 \approx 0.004$) with slightly smaller bias for the estimators using a first stage logistic model (Figure 4.2).

Again, the estimators of the conventional Cox model have a smaller empirical standard deviation than the estimators of the four IV models. In contrast to the results for the rare outcome, the standard deviation of estimators obtained

from the four IV models is of comparable size which becomes obvious from the boxplots in Figure 4.2). The empirical standard deviation of IV estimators increases in scenarios with a weak instrument, an oracle hazard ratio of 2, additional censoring and for a prevalence of drug A of 80%.

Sensitivity analysis

In scenarios with a strong instrument, a prevalence of drug A of 60% and with only administrative censoring, the robustness of the control function approach regarding the functional form of the unmeasured confounder is investigated (Table 4.4). When compared to the main simulation, 2SPS (marginal HR=1: 4% vs. 5%, marginal HR=2: 1-1.5% vs. 4-4.5%) and CF estimators (marginal HR=1: 4% vs. 6%, marginal HR=2: 5.5% vs. 8.5%) are slightly more biased for the rare outcome, whereas results for the common outcome are identical to the results obtained from the main analysis.

Table 4.4: Mean hazard ratios of the exposure effect in the sensitivity analysis for selected scenarios

Partial r^2	Survival probability	Oracle HR	Marginal HR	Cox HR	Lin. 2SPS HR	Log. 2SPS HR	Lin. CF HR	Log. CF HR
0.040	0.97	1.00	1.00	0.87	1.05	1.05	1.06	1.06
0.040	0.98	2.01	2.00	1.75	2.08	2.09	2.17	2.17
0.041	0.55	1.00	1.00	0.87	1.00	1.00	1.00	1.00
0.041	0.55	2.00	1.96	1.71	1.90	1.91	1.95	1.95

Results of the observational study

To apply the 2SPS method and the CF approach to real-life data, mortality risks between elderly new users of conventional and atypical antipsychotics (APs) using the physician's preference as IV are compared. Details on the study design, definition of exposure, instruments, outcomes and confounders as well as on the results can be found in Appendix G.

During the study period (2005-2012), 231,074 new users of APs are identified with 68% of the patients starting with a conventional AP, a survival probability of 0.9 at the end of follow-up and additional censoring. The physician's preference based on the most recent prescription is used as a binary instrument. This instrument is moderately strongly associated with the exposure to antipsychotics (partial $r^2 = 0.028$). Table 4.5 compares the risk of all-cause mortality between conventional and atypical AP users, where this risk is estimated using the Cox model and the four IV models based on the 2SPS method and the CF approach for the binary IV. The conventional Cox model suggests a higher mortality risk for users of conventional APs (HR=1.25) compared to users of atypical APs. Results of the four IV models vary, but all IV models suggest a more pronounced mortality risk for users of conventional APs. The simulation study showed that 2SPS estimators outperform CF estimators in case of a rare outcome. Furthermore, an impact of a misspecified first stage model on the accuracy of estimators was only observed for weak instruments. Here, a moderately strong instrument was identified, so that the use of a first stage logistic model is recommended. Previous observational studies suggested a higher mortality risk associated with conventional than with atypical antipsychotics (Brookhart et al., 2007; Schneeweiss et al., 2007), so that a non-null causal effect is assumed. Consequently, the two-stage predictor substitution method using a first stage logistic model is here supposed to yield the most accurate estimator.

Table 4.5: Adjusted hazard ratios for all-cause mortality of conventional vs. atypical antipsychotics for the conventional Cox model and the four IV models

Model	Hazard ratio
Cox model	1.25
Lin. 2SPS model	1.70
Log. 2SPS model	1.59
Lin. CF model	1.82
Log. CF model	1.66

Discussion

In this simulation study, the performance of IV estimators obtained with the two-stage predictor substitution method and the control function approach under the proportional hazards model for binary exposure is assessed. As expected, the estimator of the conventional Cox model is biased in all scenarios due to the unmeasured confounder, whereas 2SPS and CF estimators are less biased in most scenarios. The simulation study confirms that for a rare outcome, the marginal exposure effect can be approximated by the conditional exposure effect that is estimated by the IV models. Moreover, the marginal and the conditional exposure effect coincide in case of a null causal effect. In conclusion, the following recommendations can be given based on the simulation results. If no causal effect of the exposure on the outcome is expected, 2SPS or CF estimators should be used to estimate the exposure effect in the presence of unmeasured confounding. In case of a non-null causal effect, the use of 2SPS estimators is recommended in situations with a rare outcome, whereas CF estimators should be used if a common outcome is present given that the assumptions about the unmeasured confounding are fulfilled. In summary, misspecification of the first stage model only has an impact on the results if the prevalence of the exposure exceeds 80% or the association between the instrument and the exposure is weak. In this case, results obtained from 2SPS or CF estimators should be interpreted with caution. Moreover, especially for a rare outcome, additional censoring might introduce bias due to the loss of information during the follow-up period.

It has to be noted that this recommendation may only be valid for pharmacoepidemiological studies with a similar data structure. As shown in the sensitivity analysis, both estimators are sensitive to the functional form of the unmeasured confounder. Clearly, further simulations are needed to fully assess the impact of the distribution and the functional form of the unmeasured confounder on the performance of 2SPS and CF estimators. In addition, it seems to be reasonable to further investigate the impact of a violation of the exclusion restriction and of the sample size on the bias of both estimators. Since the hazard ratio is collapsible only if the parameter θ_A is approximately

zero (Section 4.2), the performance of both estimators should be assessed in situations of a higher causal effect e.g. a hazard ratio of 5. Finally, since the uncertainty of the first stage estimation is not taken into account in the second stage estimation, standard errors of both estimators are incorrect. Therefore, the coverage of 95% bootstrap percentile confidence intervals needs to be examined.

Chapter 5

Discussion and conclusion

This thesis illustrated the importance of and the need for adequate methods to address measured and unmeasured confounders when conducting analyses based on administrative databases. The studies presented in Chapter 2 elucidated challenges that a researcher faces when conducting a study based on the German Pharmacoepidemiological Research Database (GePaRD) and how these can be tackled by an appropriate research design or an adequate analytical method. Understanding claims data as a set of proxies that indirectly describe the health of a patient, high-dimensional propensity score are useful to adjust for residual confounding due to measured covariates, especially in situation where little is known about the determinants and effects of exposures (Schneeweiss et al., 2009). However, propensity score methods lack theoretical justification and have several limitations such that the algorithm may select colliders or covariates not related to the outcome which leads to biased estimators (Joffe, 2009; Toh et al., 2011). Consequently, even if several approaches to deal with bias and confounding due to measured covariates are applied, it remains unclear if, as e.g. in the study on comparative risk of death of antidepressants in older people with depression (cf. Section 2.2.3), the analysis was sufficiently controlled for confounding due to measured covariates and if exposure effects were consistently estimated. Furthermore, a possible effect of confounding by indication on the results could not be ruled out, although various analyses were conducted to cope with this problem. Moreover, in the

study on the risk of myocardial infarction in patients with type 2 diabetes treated with basal insulin, for instance, confounding by disease severity was suspected to impact the results of the study (cf. Section 2.2.1). Since information on the indication of treatment or duration of disease are typically not collected in administrative databases, these studies are two examples where unmeasured confounding may lead to inconsistently estimated treatment effects. This underlines the need for even more sophisticated statistical methods such as instrumental variables.

A crucial point in instrumental variable analysis is the identification of an appropriate instrument that fulfills the underlying assumptions. A task that is complicated by the fact that two of the three IV assumptions are not empirically verifiable. Intuitively, the instrument should be strongly related to the exposure, particularly since weak instruments provide misleading inferences about parameter estimates and standard errors. On the one hand, if the exposure and the instrument are only weakly associated, the IV estimator is biased, when the sample size is small, and the estimator has large standard errors. On the other hand, if a strong association between exposure and instrument could be observed and, in addition, the unmeasured confounding is suspected to be strong, the Exclusion Restriction is likely to be violated (Martens et al., 2006). Even if the sample size is large, only a slight violation of the Exclusion Restriction already leads to a biased estimator. This ambiguity further complicates the search for a suitable instrumental variable. Our study on the physician's preference as an instrumental variable in GePaRD illustrated this dilemma (cf. Section 3.2). In some sub-cohorts, the proportion of all previous patients of the same physician who were prescribed COX-2 inhibitors was identified as a moderately strong instrumental variable. However, results indicated that there might be a violation of the Exclusion Restriction. In other sub-cohorts where the Exclusion Restriction was likely to hold, the instrument was only weakly associated with the exposure to COX-2 inhibitors and an incredibly large sample size would be needed to obtain unbiased IV estimates. Consequently, instrumental variable assumptions need to be carefully checked for each particular research questions and for the relevant study population and it is advised to apply several methods to explore the plausibil-

ity of the Independence Assumption and the Exclusion Restriction (Brookhart et al., 2006; Davies et al., 2013; Kollhorst et al., 2016). Our study further demonstrated that valid instruments in one study cannot be generalized to other settings, because whether a variable meets the criteria for a valid instrumental variable may differ between health care systems (cf. Section 3.2). Since in Germany physicians are limited by the health care system in their choice between competing treatment options, the estimated preference of the physician likely reflects reimbursement rules and policy decisions instead of the true physician's preference. Therefore, alternative instrumental variables such as calendar time based on the approval of a drug should be investigated (Johnston et al., 2008).

The identification of a valid instrumental variable is a prerequisite to obtain consistent estimators, but whether consistent estimation can be achieved further depends on the scale of exposure and outcome. The two-stage predictor substitution method and the control function approach are well-established for continuous outcomes using linear models (cf. Chapter 3). However, two-stage predictor substitution estimation employing linear models is not justified in the context of dichotomous exposures and outcomes, since linear models may yield predicted values outside of [0,1] and the distribution of residuals may not fulfill the assumptions of normality and homoscedascity. Nevertheless, it has been suggested that these are theoretical rather than practical problems if the assumption of linearity holds at least approximately (Angrist, 2001; Johnston et al., 2008). However, it seems to be more self-evident to use logistic models for binary exposures and outcomes, but consistent estimation can only be achieved for rare outcomes and in case of no exposure effect (Angrist, 2001; Didelez et al., 2010; Palmer et al., 2011; Robins and Rotnitzky, 2004).
A similar situation occurs when instrumental variables are to be used in survival analysis (cf. Chapter 4), which is due to the fact that hazard ratios as well as odds ratios are non-collapsible. Two-stage estimation such as the two-stage predictor substitution method and the control function approach are shown to consistently estimate the exposure effect in additive hazards models (Li et al., 2015; Tchetgen Tchetgen et al., 2015). Although the proportional hazards model is still the most popular model in time-to-event analysis and

is frequently applied in pharmacoepidemiological studies (cf. Chapter 2), the use of two-stage IV methods for this model is only justified for rare outcomes (Tchetgen Tchetgen et al., 2015).
This thesis showed that consistent estimation with the two-stage predictor substitution method can be additionally achieved in situations with a null causal effect, so that at least a test on the causal null hypothesis is valid even if the estimator is biased (cf. Chapter 4). In addition, the simulation study demonstrated that only a small bias is expected if the true exposure effect is close to but not equal to zero. Furthermore, it could be shown that the resulting control function estimator is also unbiased in case of a null causal effect, although a theoretical justification is lacking (cf. Chapter 4).
In contrast to the two-stage predictor substitution method, the control function approach requires very specific assumptions on the structure of unmeasured confounding to yield a consistent estimator when the outcome is rare. In case of a continuous outcome, assumptions are very similar to those required for the linear model. However, these assumptions cannot empirically verified, since the influence of the unmeasured confounder on the hazard is unclear. Although, in at least some situations, the control function estimator was shown to be less biased than the estimator derived by the two-stage predictor substitution (cf. Chapter 4), conclusions about the performance of this method besides situations with rare outcomes can only be drawn for this specific simulated data structure. It needs to be noted that results might not be transferable to other data situations that assume, for instance, a different distribution and functional form of the unmeasured confounder. Due to these reasons, the use of the control function approach seems to be not appealing, since two-stage methods are only useful to the extent that a valid instrumental variable exists. Nevertheless, if the analysis is not controlled for unmeasured confounding, exposure effect estimators may be highly biased (cf. Chapter 4). In proportional hazards models, application of the two-stage predictor substitution method is preferred over the control function approach, but researchers need to be aware that these estimators are only consistent in situations with rare outcomes and under the null hypothesis of no exposure effect.

This thesis focused on instrumental variable estimation using two-stage

methods, but alternative instrumental variable methods for censored survival outcomes using structural proportional hazards models are available. However, these methods are limited to either binary instruments or binary exposure variables and cannot be generalized to other situations. Furthermore, these models aim to estimate a so-called complier effect and additionally rely on a monotonicity assumption about the effect of the instrument on exposure (Cuzick et al., 2007; Loeys et al., 2005). Recently, MacKenzie et al. (2014) proposed to use instrumental variables to estimate a Cox model in situations where the unmeasured confounder is assumed to be additive on the scale of the hazard ratio and the unmeasured confounder is assumed to have zero mean. These assumptions are overly restrictive and the proposed method cannot be applied to the situation of a multiplicative effect of the unobserved confounder which would be more plausible. Despite the popular use of the proportional hazards model in observational studies based on large health care databases, survival models other than the proportional hazards model should be considered. The Aalen additive hazards model is less commonly used, but consistent estimation of the causal effect using instrumental variables can be achieved (Li et al., 2015; Tchetgen Tchetgen et al., 2015). Nevertheless, the use of Aalen models instead of proportional hazards models cannot generally be recommended (Burgess, 2015). First, there is the practical issue that the hazard function possibly takes negative values for some individuals. Second, the model assumes an additive exposure effect on the hazard, but a multiplicative effect as assumed in the Cox model is often more plausible.

This thesis showed that instrumental variable analysis is useful to address unmeasured confounding in pharmacoepidemiological studies, but its application is limited to specific settings, e.g. time-independent exposure and confounder. Large health care databases have the potential for a long follow-up that provides the possibility to investigate adverse drug reactions that only develop after cumulative drug intake. In this setting, time-varying exposure in the presence of time-dependent confounding is likely and the use of marginal structural models is advisable. However, consistent estimation of the causal effect of exposure relies here on the assumption of no unmeasured confounding (Robins et al., 2000). Instrumental variable methodology in Aalen or

Cox models is currently lacking for settings where exposure and instruments are time-varying and, hence, instrumental variable analysis are restricted to intention-to-treat analysis that assumes that exposure is time-independent. Future research is needed to cope with unmeasured confounding in case of time-varying exposure and instruments.

Instrumental variable methods are further not applicable in situations where the unmeasured confounder is suspected to be time-dependent. In time-to-event analysis, propensity calibration and imputation of the unmeasured confounders based on their relationships with exposure, measured confounders, and outcome that are estimated in a validation subsample are two possibilities to cope with time-varying unmeasured confounding (Burne and Abrahamowicz, 2016; Stürmer et al., 2005).

When using weak instruments in small samples or when the instrument variable assumptions are violated, it is possible that IV estimators are even more biased than conventional exposure effect estimators (Martens et al., 2006). Moreover, the amount and structure of confounding and its impact on results is unclear and findings from one study cannot be generalized to other settings. In these situations, it should be aimed to conduct bias sensitivity analysis to assess the potential impact of unmeasured confounding on the association between exposure and outcome and to assess how sensitive the estimated exposure effect is to unmeasured confounding and in which situations the conclusions drawn may alter (Lin et al., 2013; Schneeweiss, 2006). An alternative method to instrumental variables that may detect and reduce the impact of unmeasured confounding is the missing cause approach that relies on the assumption that the impact of unmeasured confounding may be detected by the discrepancies between the treatment a patient actually received and the treatment that a patient would be expected to receive, based on the observed data (Abrahamowicz et al., 2016).

Unmeasured confounding is a complex problem that will not be solved by a single method. In considering various methods that rely on different assumptions such as propensity calibration, instrumental variable and sensitivity analysis, it should be aimed to gain knowledge about the impact of unmeasured confounding. Even if IV assumptions are questionable to be valid, IV analysis can still be a part of the sensitivity analyses (Greenland, 2000).

This thesis demonstrated that studies of adverse drug reactions based on large health care databases face a variety of problems. Even if several approaches to deal with bias and confounding are applied, conclusions about the effect of exposure on adverse events should be cautious. Due to the complexity of the various aspects of measured and unmeasured confounding, careful planning, analysis and interpretation of studies remain a great challenge in pharmacoepidemiology.

Appendix A

Paper: Outpatient antipsychotic drug use in children and adolescents in Germany between 2004 and 2011

Contribution to the manuscript I herewith certify that I contributed to the design of the study, performed the statistical analyses, interpreted the results, and revised the manuscript critically for important intellectual content.
Due to copyright issues, the paper was removed from the published version of the thesis. Only the abstract is given here.

Outpatient antipsychotic drug use in children and adolescents in Germany between 2004 and 2011

C Schröder[1], M Dörks[2], B Kollhorst[3], T Blenk[1], RW Dittmann[4], E Garbe[1], O Riedel[1]

[1]Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[2]Department of Health Services Research, Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[3]Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[4]Paediatric Psychopharmacology, Department of Child and Adolescent Psychiatry and Psychotherapy, Medical Faculty Mannheim, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany

Abstract

Studies from different countries showed increasing use of antipsychotics in pediatric patients. However, these studies were methodologically limited and could not assess underlying diagnoses and off-label use sufficiently. This is the first study to examine antipsychotic prescriptions in a representative sample of minors over a long period, looking at changes regarding substances and drug classes, underlying diagnoses, and the rate of off-label use. Claims data of about two million pediatric subjects were used to calculate annual prevalences and incidence rates of antipsychotic prescriptions for the years 2004-2011. Analyses were stratified by sex, age, and drug type. Numbers of prescriptions, frequencies of diseases/disorders, the prescribing physicians' specialties, and the share of off-label prescriptions were examined. During the study period, the prevalence of antipsychotic prescriptions ranged between 2.0 and 2.6 per 1000 minors. Antipsychotic prescriptions in children younger than 6 years decreased from 2.42 per 1000 subjects in 2004 to 0.48 in 2011. Among antipsychotic users, 47.0 % had only one prescription and hyperkinetic disorder was, by far, the most frequent diagnosis. The annual share of off-label

prescriptions varied between 61.0 and 69.5 %. Antipsychotics were mainly prescribed to manage aggressive and impulsive behaviors in hyperkinetic disorder patients. This explains the high share of off-label prescriptions but raises concerns, since efficacy and safety of antipsychotics in this indication have not been sufficiently investigated. The decreasing antipsychotic use in younger children and the high proportion of antipsychotic users with one-time prescriptions are striking and should be further investigated in the future.

Appendix B

Paper: Non-steroidal ant-inflammatory drugs and risk of heart failure in four European countries: nested case-control study

Contribution to the manuscript I herewith certify that I performed parts of the statistical analyses, interpreted the results, and revised the manuscript critically for important intellectual content.

Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study

Andrea Arfè,[1] Lorenza Scotti,[1] Cristina Varas-Lorenzo,[2] Federica Nicotra,[1] Antonella Zambon,[1] Bianca Kollhorst,[3] Tania Schink,[3] Edeltraut Garbe,[3] Ron Herings,[4] Huub Straatman,[4] René Schade,[5] Marco Villa,[6] Silvia Lucchi,[6] Vera Valkhoff,[5] Silvana Romio,[5] Frantz Thiessard,[7] Martijn Schuemie,[5] Antoine Pariente,[7] Miriam Sturkenboom,[5] Giovanni Corrao[1] On behalf of the Safety of Non-steroidal Anti-inflammatory Drugs (SOS) Project Consortium

[1]Unit of Biostatistics, Epidemiology, and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, 20126 Milan, Italy
[2]RTI Health Solutions, Barcelona, Spain
[3]Leibniz Institute of Prevention Research and Epidemiology, Bremen, Germany
[4]PHARMO Institute, Utrecht, Netherlands
[5]Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, Netherlands
[6]Local Health Authority ASL Cremona, Cremona, Italy
[7]University of Bordeaux Segalen, Bordeaux, France
Correspondence to: G Corrao giovanni.corrao@unimib.it
Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;354:i4857
http://dx.doi.org/10.1136/bmj.i4857
Accepted: 01 September 2016

ABSTRACT

OBJECTIVES
To investigate the cardiovascular safety of non-steroidal anti-inflammatory drugs (NSAIDs) and estimate the risk of hospital admission for heart failure with use of individual NSAIDs.

DESIGN
Nested case-control study.

SETTING
Five population based healthcare databases from four European countries (the Netherlands, Italy, Germany, and the United Kingdom).

PARTICIPANTS
Adult individuals (age ≥18 years) who started NSAID treatment in 2000-10. Overall, 92 163 hospital admissions for heart failure were identified and matched with 8 246 403 controls (matched via risk set sampling according to age, sex, year of cohort entry).

MAIN OUTCOME MEASURE
Association between risk of hospital admission for heart failure and use of 27 individual NSAIDs, including 23 traditional NSAIDs and four selective COX 2 inhibitors. Associations were assessed by multivariable conditional logistic regression models. The dose-response relation between NSAID use and heart failure risk was also assessed.

RESULTS
Current use of any NSAID (use in preceding 14 days) was found to be associated with a 19% increase of risk of hospital admission for heart failure (adjusted odds ratio 1.19; 95% confidence interval 1.17 to 1.22), compared with past use of any NSAIDs (use >183 days in the past). Risk of admission for heart failure increased for seven traditional NSAIDs (diclofenac, ibuprofen, indomethacin, ketorolac, naproxen, nimesulide, and piroxicam) and two COX 2 inhibitors (etoricoxib and rofecoxib). Odds ratios ranged from 1.16 (95% confidence interval 1.07 to 1.27) for naproxen to 1.83 (1.66 to 2.02) for ketorolac. Risk of heart failure doubled for diclofenac, etoricoxib, indomethacin, piroxicam, and rofecoxib used at very high doses (≥2 defined daily dose equivalents), although some confidence intervals were wide. Even medium doses (0.9-1.2 defined daily dose equivalents) of indomethacin and etoricoxib were associated with increased risk. There was no evidence that celecoxib increased the risk of admission for heart failure at commonly used doses.

CONCLUSIONS
The risk of hospital admission for heart failure associated with current use of NSAIDs appears to vary between individual NSAIDs, and this effect is dose dependent. This risk is associated with the use of a large number of individual NSAIDs reported by this study, which could help to inform both clinicians and health regulators.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Several randomised clinical trials and observational studies have shown an association between use of non-steroidal anti-inflammatory drugs (NSAIDs) and increased risk of heart failure, but the risk and dose-response relation associated with individual NSAIDs is largely unknown

Heart failure was included as an outcome of interest in the Safety of Non-Steroidal Anti-Inflammatory Project, a multinational project funded by the European Commission under the seventh Framework Programme

WHAT THIS STUDY ADDS

Use of seven individual traditional NSAIDs (diclofenac, ibuprofen, indomethacin, ketorolac, naproxen, nimesulide, and piroxicam) and two individual COX 2 selective NSAIDs (etoricoxib and rofecoxib) is associated with and increased risk of hospital admission for heart failure

Risk of admission for heart failure is doubled for some NSAIDs used at very high doses

Estimates of the risk of heart failure associated with the use of many individual NSAIDs in this study could help to inform both clinicians and health regulators

Introduction

Non-steroidal anti-inflammatory drugs (NSAIDs) are a broad class of agents with analgesic and anti-inflammatory properties that inhibit the two recognised isoenzymes of prostaglandin G/H synthase (also known as cyclo-oxygenase (COX))—namely, COX 1 and COX 2.[1] Because the therapeutic action of these drugs is mostly mediated by inhibition of COX 2, while their gastrointestinal adverse reactions are largely due to COX 1 inhibition, NSAIDs selectively inhibiting COX 2 were developed in the 1990s to reduce the risk of gastrointestinal toxicity.[2]

Nevertheless, reports of cardiovascular adverse reactions began to emerge in 2000-03,[3 4] and subsequent placebo controlled trials showed that COX 2 inhibitors were associated with an increased risk of atherothrombotic vascular events.[5 6] However, meta-analyses of randomised trials and observational studies have since shown that the higher cardiovascular risk is not restricted to COX 2 inhibitors, but also applies to some traditional NSAIDs.[7-12]

In particular, NSAID use has been found to be associated with an increased risk of heart failure in several randomised clinical trials[11] and observational studies.[13 14] A large meta-analysis of over 600 randomised trials showed that COX 2 inhibitors and high doses of traditional NSAIDs (that is, diclofenac, ibuprofen, and naproxen) increased the risk of hospital admission for heart failure from 1.9-fold to 2.5-fold compared with placebo.[11] In the light of this evidence, current guidelines limit the use of NSAIDs in patients predisposed to heart failure, with a full contraindication for patients with diagnosed heart failure.[15]

Nevertheless, there is still limited information on the risk of heart failure associated with the use of individual NSAIDs (both COX 2 inhibitors and traditional NSAIDs) in clinical practice, and especially on their dose-response associations. Therefore, heart failure was included as an outcome of interest in the overall cardiovascular and gastrointestinal risk evaluation of individual NSAIDs within the Safety of Non-Steroidal Anti-Inflammatory (SOS) Project, a multinational project funded by the European Commission under the seventh Framework Programme. A large, common protocol, nested case-control study based on electronic healthcare databases from four European countries was carried out.

Methods

Data sources

This study was based on five electronic health databases from four European countries: the Netherlands, Italy, Germany, and the UK. Overall, these databases covered over 37 million people with different time windows of data availability between 1999 and 2010; table 1 summarises their main characteristics.

Briefly, PHARMO is a population based, medical record linkage system covering more than two million inhabitants from the Netherlands. SISR is an electronic administrative healthcare database in Italy, covering the about 10 million residents in the Lombardy region, who all receive free healthcare assistance from the Italian national health service. OSSIFF is a healthcare database covering about three million individuals who are beneficiaries of eight local health authorities in the Lombardy region. Because OSSIFF covers a subset of the population already covered by SISR, we included only the seven million beneficiaries of the Italian national health service not already included in OSSIFF in this study. GePaRD is a claims database in Germany covering about 14 million individuals enrolled in four German statutory health insurance providers. Lastly, THIN is a general practice database comprising primary care medical records from more than 10 million individuals in the UK.

Each database longitudinally recorded data on each member of its target population, including demographic data, hospital discharge diagnoses, and outpatient drug prescriptions. Data on outpatient diagnoses were also available from GePaRD. In two databases (PHARMO and THIN), the daily dose prescribed by physicians was recorded for each dispensed prescription. Further details are reported elsewhere.[17]

Harmonisation and data processing

Databases differed in several aspects, including type of collected information (that is, healthcare use, claims, and primary care data) and classification systems used for disease and medication coding (table 1). As a result, we performed data harmonisation according to a procedure developed and assessed in the European Union (EU)-ADR (exploring and understanding adverse drug reactions by integrative mining of clinical records and biomedical knowledge) Project[18] and also implemented in other EU funded projects.[19] Specifically, the Unified Medical Language system (for clinical diagnoses and conditions) and the Anatomic Therapeutic Chemical (ATC) classification system (for drug prescriptions) were mapped into the coding systems used by the individual databases. This mapping ensured that the data extraction processes targeted the same semantic concepts across all databases, thus allowing analyses to be performed under a common data model.[19]

Anonymised data were extracted locally and processed with Jerboa software (developed by Erasmus MC), providing individual level datasets in a common data format. These datasets were securely transferred into the SOS data warehouse, hosted by the University of Milano-Bicocca, to be analysed centrally and securely.[19]

Table 1 | Databases considered as data sources for the present study among individuals participating in the SOS Project

Country	Database*	Type of database	Size of covered population	Covered period	Diagnoses coding	Drugs coding
Netherlands	PHARMO (PHARMO Institute for Drug Outcomes Research)	Record linkage	2.2 million	1999-2008	ICD-9-CM	Anatomical Therapeutic Chemical classification system
Italy	SISR (Sistema Informativo Sanitario Regionale)†	Healthcare use	7.5 million	2003-08	ICD-9-CM	Anatomical Therapeutic Chemical classification system
	OSSIFF (Osservatorio Interaziendale per la Farmacoepidemiologia e la Farmacoeconomia)	Healthcare use	2.9 million	2000-08	ICD-9-CM	Anatomical Therapeutic Chemical classification system
Germany	GePaRD (German Pharmacoepidemiological Research Database)	Claims	13.7 million	2004-09	ICD-10-GM	Anatomical Therapeutic Chemical classification system
UK	THIN (The Health Improvement Network)	General practice	11.1 million	1999-2010	READ version 2	BNF/Multilex codes

ICD-9-CM=International Classification of Diseases, 9th revision, clinical modification; ICD-10-GM=International Classification of Diseases, 10th revision, German modification; READ=READ clinical classification system; BNF=British National Formulary; Multilex=Multifunctional Standardised Lexicon for European Community Languages drug terminology.
*Other databases participated in the SOS Project but did not contribute data to this study.[16]
†Because OSSIFF covers a subset of patients also covered by SISR, this database excluded the common subset of patients to avoid overlap.

Cohort selection and follow-up

Following the new users paradigm,[16] a cohort of individuals starting NSAID treatment was selected from all databases. In detail, adults (age ≥18 years) who received at least one NSAID prescription or dispensation (ATC code M01A; excluding topical NSAIDs) during 2000-10 were considered eligible to enter the cohort. The date of first recorded prescription or dispensation was defined as the date of cohort entry. We excluded participants if they:

- Did not have at least one year of uninterrupted observation before the date of cohort entry, to ensure enough time of observation for assessing baseline covariates and applying the next exclusion criteria
- Received one or more NSAIDs within the year preceding the date of cohort entry, to exclude prevalent NSAIDs users
- Received a diagnosis of malignant cancer, with the exception of non-melanoma skin cancers, to exclude patients who may have had particular contraindications
- Were admitted to hospital with a primary diagnosis of heart failure in the year before the date of cohort entry, to avoid the inclusion of events occurring before the start of NSAIDs use (note that secondary hospital or outpatient heart failure diagnoses were not considered as exclusion criteria).

Each cohort member accumulated person years of follow-up, from the date of cohort entry to the earliest date of outcome onset (date of first hospital admission with a primary diagnosis of heart failure), censoring (end of registration in the database due to death or emigration), diagnosis of malignancy (excluding non-melanoma skin cancers), or end of database specific data availability.

Cases and controls

A case-control study was nested into the cohort of new users of NSAIDs. The endpoint of interest was the first hospital admission for heart failure (that is, with heart failure as the main cause or reason of hospital admission) identified during follow-up. Heart failure is a clinical syndrome involving several pathophysiological mechanisms that, along with factors triggering circulatory decompensation, could produce heterogeneous clinical manifestations that often receive delayed diagnosis. Therefore, our endpoint definition did not include diagnostic codes for clinical heart failure in the outpatient setting and secondary hospital discharge codes for heart failure (which are likely to represent heart failure manifestations occurring during hospital admission for other causes).

Consequently, cases were all cohort members admitted for heart failure during follow-up, identified either from primary hospital discharge diagnoses (PHARMO, SISR, OSSIFF, GePaRD) or codes registered by the general practitioner (THIN). We defined the date of the first admission for heart failure identified during follow-up as the index date. Codes used to identify heart failure cases in each database are reported in the supplementary material (table S1).

We matched each case to up to 100 controls. Controls were randomly selected by risk set sampling from all cohort members whose follow-up did not end before the index date of the considered case (that is, among individuals still at risk of an admission for heart failure). Matching was performed within each database according to sex, age at cohort entry (within 1 year's difference), and date of cohort entry (within 28 days' difference).

Exposure to NSAIDs

All NSAIDs dispensations received by cohort members during follow-up were identified; this included 27 individual NSAIDs (23 traditional NSAIDs and four selective COX 2 inhibitors). For each cohort member, we directly calculated the period covered by the availability of each individual NSAID by the prescribed daily dose, if available (that is, PHARMO and THIN databases), or by dividing the total amount of drug prescribed for the defined daily dose.

We classified cohort members into the following categories of NSAID use: current, recent, and past. Current users were patients with NSAID availability at the index date or the preceding 14 days. The remaining patients were defined recent users if they had NSAID availability during the time window of 15-183 days before the index date, or past users otherwise (reference).

Covariates

We assessed several covariates for each cohort member if available in the corresponding database, including:

- History of outpatient or secondary inpatient diagnoses of heart failure, comorbidities, and lifestyle features or clinical characteristics, assessed in the 12 months before cohort entry
- Concomitant use of specific drugs, assessed in the 90 days before the index date.

Comorbidities were assessed by hospital discharge diagnoses (PHARMO, GePaRD, SISR, OSSIFF), outpatient clinical diagnoses (GePaRD), clinical electronic general practice records (THIN), and use of specific drugs. Table 2 reports the full list of covariates.

Statistical analysis

Individual level data from all databases were firstly gathered into a pooled dataset and analysed by means of a multivariable conditional logistic regression model.[20] The obtained odds ratio, with 95% confidence intervals, estimated the risk of hospital admission for heart failure associated with current use of individual NSAIDs with respect to past use of any NSAID. We also estimated the odds ratio associated with recent use of any NSAID, compared with past use of any NSAID. Given the substantial number of associations assessed in this analysis, we used the Bonferroni-Holm procedure[21] to assess the impact of uncertainty due to multiple comparisons on the results.[22] Some evidence has supported the relative cardiovascular safety of celecoxib by comparison with other NSAIDs is available in the literature.[14 23-28] Therefore, as a secondary analysis, we estimated the odds ratios measuring the association

Table 2 | Clinical features and other selected characteristics of patients admitted to hospital for heart failure and matched control patients included in the study (SOS Project). Data are No (%) of patients unless stated otherwise

	Case patients (n=92 163)	Controls (n=8 246 403)
Men	41 652 (45.2)	3 671 565 (44.5)
Age at cohort entry (years, mean (standard deviation))	77 (11)	76 (10)
Comorbidities and other characteristics*		
Acute myocardial infarction†	3063 (3.3)	81 222 (1.0)
Alcohol abuse (ATC code starting with N07BB)	1942 (2.1)	128 871 (1.6)
Asthma†	1031 (1.1)	57 079 (0.7)
Atrial fibrillation and flutter†	4606 (5.0)	110 217 (1.3)
Chronic liver disease†	1815 (2.0)	98 762 (1.2)
Chronic respiratory disease† (ATC code starting with R03)	16 190 (17.6)	870 497 (10.6)
Diabetes† (ATC code starting with A10)	17 888 (19.4)	725 320 (8.8)
Heart failure† (ATC code C07AG02)	8353 (9.1)	209 125 (2.5)
Hyperlipidaemia† (ATC code starting with C10)	18 793 (20.4)	1 160 532 (14.1)
Hypertension†	19 905 (21.6)	1 515 002 (18.4)
Iron deficiency anaemia (ATC code starting with B03A)	2159 (2.3)	83 926 (1.0)
Ischaemic heart disease†	8406 (9.1)	294 986 (3.6)
Kidney failure	1445 (1.6)	41 094 (0.5)
Obesity (ATC code starting with A08A)	4555 (4.9)	181 104 (2.2)
Osteoarthritis†	6916 (7.5)	483 721 (5.9)
Other cardiovascular disease (ATC code starting with C01B)†‡	13 055 (14.2)	463 797 (5.6)
Rheumatoid arthritis and inflammatory polyarthritis† (ATC code starting with M01C)	736 (0.8)	40 269 (0.5)
Smoking	164 (0.2)	8155 (0.1)
Stroke†	1869 (2.0)	85 109 (1.0)
Valvular disease and endocarditis†	2383 (2.6)	70 646 (0.9)
Concomitant use of other drugs§		
ACE inhibitor/angiotension II antagonists†	38 834 (42.1)	2 030 050 (24.6)
Anticoagulants†	17 589 (19.1)	442 725 (5.4)
Aspirin†	31 658 (34.4)	1 669 443 (20.2)
β blockers†	22 506 (24.4)	1 253 749 (15.2)
Calcium channel blockers†	28 911 (31.4)	1 754 965 (21.3)
Cardiac glycosides†	14 429 (15.7)	342 042 (4.1)
Cyp2C9 inducers	38	1149
Cyp2C9 inhibitors	8289 (9.0)	174 253 (2.1)
Diuretics†	48 991 (53.2)	1 536 700 (18.6)
Glucocorticoids†	8636 (9.4)	349 012 (4.2)
Nitrates†	24 029 (26.1)	717 669 (8.7)
Platelet aggregation inhibitor†	9105 (9.9)	367 716 (4.5)
Vasodilators†	1654 (1.8)	44 916 (0.5)

ATC= Anatomic Therapeutic Chemical (ATC) classification system; ACE=angiotensin converting enzyme; Cyp2C9=cytochrome P450 2C9.
*Comorbidities assessed during the 12 months before cohort entry, on the basis of inpatient diagnoses, outpatient diagnoses (German GePaRD database only), medical history (UK THIN database only), or selected drug prescriptions belonging to the indicated ATC codes (only for specific covariates).
†Available in all databases.
‡Other cardiovascular diseases include: cardiac arrhythmia or conduction disorders and arrest, cardiomyopathies, peripheral arterial diseases, arterial embolism and thrombosis, myocarditis, and pericarditis.
§Drug use assessed during the 14 days preceding the index hospital admission for heart failure.

between current use of individual NSAIDs and heart failure risk, using current use of celecoxib as reference. Among the covariates mentioned above, those available in all databases (including history of outpatient or secondary inpatient diagnoses of heart failure) entered the model. We did subgroups analyses after stratification for sex and history of heart failure diagnoses.

Because databases differed with respect to covered populations, as well as type and level of detail of available covariates, we evaluated the robustness of the pooled estimates using a meta-analytic approach by means of the following procedure. Firstly, we separately fitted a conditional logistic regression model to estimate the effect of each individual NSAID within each database. To avoid computational issues (that is, model convergence failure due to sparse data), only individual NSAIDs with at least five exposed cases were considered in the model. The covariates available for all databases were always forced to enter the model, provided they reached at least 5% prevalence among controls. Other covariates were included, provided they were significantly ($P<0.05$) associated with the outcome in a univariate analysis, and selected from a backward procedure ($P>0.10$ for removal). Secondly, we used a random effects meta-analytic model[29 30] to estimate a summary odds ratio (and 95% confidence interval) across databases for current use of each individual NSAID (provided that a point estimate was available from at least two databases), compared with past use of any NSAID. Heterogeneity between database specific odds ratios was assessed by Cochran's Q and Higgins' I^2 statistics.[31]

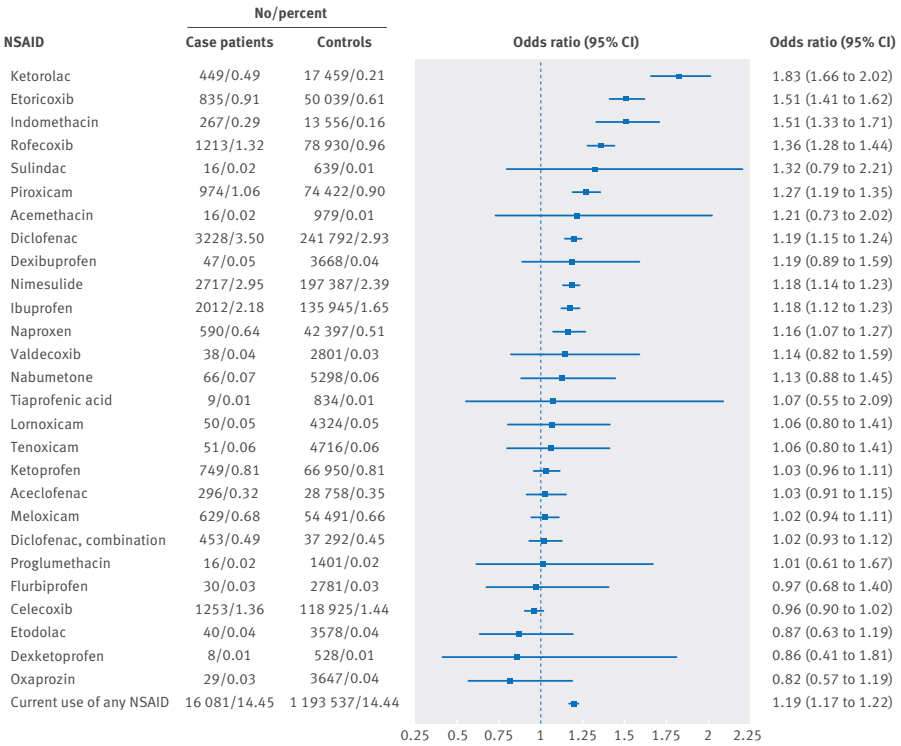

Fig 1 | **Distribution of current use of individual NSAIDs among cases and controls and pooled associations between current use of individual NSAIDs and risk of hospital admission for heart failure, with past use of any NSAID as reference. Estimates obtained by pooling individual data from all available databases. Pooled odds ratios and 95% confidence intervals estimated by fitting a conditional logistic regression model after correcting for available covariates**

Dose-response analysis

We did a dose-response analysis to assess how the risk of hospital admission for heart failure associated with current use of individual NSAIDs varied along the considered categories of prescribed daily dose. Because Italian and German databases did not record data on prescribed daily doses, we pooled individual level data from the Netherlands (PHARMO) and the UK (THIN) databases. Patients for whom the information on the prescribed daily dose was not available were excluded.

The prescribed daily dose was expressed in defined daily dose equivalents (DDD) and categorised as low (≤0.8 DDD), medium (0.9-1.2 DDD), high (1.3-1.9), or very high dose (≥2 DDD) with respect to the corresponding defined daily dose. To avoid computational issues, we considered only NSAIDs for which all the considered categories included at least one heart failure case in the analysis. Tests for trends in odds ratios were performed. We did statistical analyses using SAS software (version 9.3; SAS Institute). All tests were two sided and considered significant for P values less than 0.05.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

Study cohort

Supplementary figure S1 shows the flowchart describing the attrition of eligible NSAIDs users after exclusion criteria were applied. Among nearly 10 million new users of NSAIDs identified across all databases, 7 680 181 met the inclusion criteria and constituted the study cohort. Cohort members accumulated 24 555 063 person years of follow-up and generated 92 163 cases of heart failure admitted to hospital (incident rate, 37.5 heart failure events per 10 000 person years). Cases were matched to 8 246 403 controls.

Mean age was 77 (standard deviation 11) years and 76 (10) years among cases and controls, respectively (table 2). About 45% of both cases and controls were men. Compared with controls, cases had more comorbidities (mainly cardiovascular disease, such as acute myocardial infarction, other ischaemic heart diseases, atrial fibrillation and flutter, and valvular disease and endocarditis) and received concomitant drug treatments more often (eg, anticoagulants, cardiac

Table 3 | Risk of hospital admission for heart failure for current users of individual NSAIDs or recent users of any NSAID (versus past use of any NSAIDs), according to evidence of prior heart failure and by sex. Evidence of prior heart failure obtained from outpatient or secondary hospital diagnoses before start of NSAID treatment. P values test homogeneity of odds ratios between groups. NA=not available

	Risk of admission for heart failure (pooled odds ratio (95% CI))			Risk of admission for heart failure (pooled odds ratio (95% CI))		
	No prior heart failure (n=83 810)	Prior heart failure (n=8353)	P	Men (n=41 652)	Women (n=50 511)	P
Current use of NSAID						
Indomethacin	1.52 (1.31 to 1.77)	1.58 (0.55 to 4.51)	0.94	1.71 (1.41 to 2.07)	1.25 (1.00 to 1.57)	0.04
Sulindac	1.62 (0.90 to 2.94)	NA	NA	2.19 (0.80 to 5.97)	1.50 (0.71 to 3.16)	0.55
Diclofenac	1.21 (1.16 to 1.26)	1.14 (0.91 to 1.42)	0.61	1.21 (1.13 to 1.29)	1.19 (1.13 to 1.26)	0.70
Etodolac	0.83 (0.59 to 1.17)	NA	NA	0.92 (0.56 to 1.50)	0.77 (0.48 to 1.23)	0.61
Acemetacin	1.67 (0.83 to 3.35)	0.28 (0.03 to 2.32)	0.13	1.22 (0.52 to 2.85)	2.08 (0.84 to 5.12)	0.40
Ketorolac	1.94 (1.71 to 2.19)	5.09 (0.97 to 26.57)	0.25	1.86 (1.52 to 2.28)	1.96 (1.70 to 2.27)	0.68
Aceclofenac	1.00 (0.87 to 1.14)	0.89 (0.20 to 3.89)	0.88	1.13 (0.90 to 1.41)	0.94 (0.80 to 1.11)	0.19
Diclofenac, combinations	1.02 (0.92 to 1.14)	0.89 (0.36 to 2.16)	0.77	1.03 (0.87 to 1.22)	1.01 (0.89 to 1.16)	0.86
Piroxicam	1.31 (1.21 to 1.41)	1.90 (1.01 to 3.59)	0.25	1.34 (1.18 to 1.53)	1.31 (1.20 to 1.44)	0.78
Tenoxicam	1.03 (0.74 to 1.43)	NA	NA	0.88 (0.47 to 1.62)	1.07 (0.75 to 1.53)	0.59
Lornoxicam	1.13 (0.81 to 1.57)	2.25 (0.28 to 18.08)	0.52	1.22 (0.68 to 2.18)	1.07 (0.73 to 1.56)	0.71
Meloxicam	0.99 (0.91 to 1.09)	0.95 (0.43 to 2.07)	0.92	1.13 (0.97 to 1.31)	0.95 (0.85 to 1.06)	0.07
Ibuprofen	1.15 (1.08 to 1.21)	1.34 (1.05 to 1.70)	0.23	1.18 (1.09 to 1.29)	1.16 (1.09 to 1.25)	0.76
Naproxen	1.19 (1.08 to 1.31)	0.87 (0.32 to 2.38)	0.54	1.24 (1.08 to 1.42)	1.15 (1.01 to 1.30)	0.43
Ketoprofen	1.04 (0.95 to 1.13)	1.00 (0.50 to 2.03)	0.91	1.15 (1.00 to 1.32)	0.98 (0.88 to 1.09)	0.07
Flurbiprofen	1.08 (0.72 to 1.62)	NA	NA	1.19 (0.62 to 2.31)	0.83 (0.50 to 1.39)	0.40
Oxaprozin	0.82 (0.55 to 1.23)	0.26 (0.02 to 3.77)	0.40	0.45 (0.19 to 1.08)	0.96 (0.61 to 1.51)	0.13
Dexibuprofen	1.24 (0.89 to 1.74)	NA	NA	0.92 (0.50 to 1.67)	1.38 (0.93 to 2.03)	0.27
Celecoxib	0.95 (0.89 to 1.02)	1.05 (0.53 to 2.06)	0.77	1.01 (0.90 to 1.13)	0.94 (0.87 to 1.01)	0.30
Rofecoxib	1.34 (1.25 to 1.44)	0.91 (0.35 to 2.42)	0.43	1.35 (1.20 to 1.52)	1.37 (1.26 to 1.48)	0.84
Valdecoxib	1.04 (0.69 to 1.56)	0.47 (0.03 to 8.01)	0.58	0.95 (0.45 to 2.02)	1.09 (0.70 to 1.71)	0.76
Etoricoxib	1.55 (1.42 to 1.69)	1.35 (0.75 to 2.44)	0.65	1.80 (1.57 to 2.07)	1.45 (1.31 to 1.61)	0.01
Nabumetone	1.07 (0.81 to 1.43)	11.14 (0.67 to 184.24)	0.10	1.14 (0.73 to 1.80)	1.17 (0.83 to 1.64)	0.93
Nimesulide	1.21 (1.16 to 1.27)	1.00 (0.62 to 1.60)	0.43	1.31 (1.21 to 1.42)	1.17 (1.10 to 1.23)	0.02
Recent use of any NSAID	1.01 (0.99 to 1.03)	0.97 (0.85 to 1.11)	0.55	1.04 (1.01 to 1.07)	0.99 (0.96 to 1.01)	0.01

glycosides, nitrates, and cytochrome P450 2C9 inhibitors). We found 9.1% of cases and 2.5% of controls with a history of heart failure diagnosis, recorded as either an outpatient diagnosis or a secondary hospital diagnosis in the year before start of NSAID treatment (cohort entry).

NSAID use and heart failure risk

A total of 16 081 (17.4%) cases and 1 193 537 (14.4%) matched controls were current users of NSAIDs. Fig 1 reports the distribution of current use of individual NSAIDs among all cases and controls. Among controls, the most frequently used traditional NSAIDs were diclofenac (2.9%), nimesulide (2.4%), and ibuprofen (1.7%), while the most frequently used COX 2 inhibitors were celecoxib (1.4%), rofecoxib (1.0%), and etoricoxib (0.6%).

According to the pooled analysis, current users of any NSAID had a 20% higher risk of heart failure than past users (odds ratio 1.19; 95% confidence interval 1.17 to 1.22). Conversely, there was no evidence that recent use of any NSAID was associated with differences in heart failure risk with respect to past use (1.00; 0.99 to 1.02). We observed a statistically significantly higher risk of heart failure in association with current use of nine individual NSAIDs than with past use of any NSAIDs (fig 1). These NSAIDs were ketorolac, etoricoxib, indomethacin, rofecoxib, piroxicam, diclofenac, ibuprofen, nimesulide, and naproxen. Other less frequently used NSAIDs (eg, sulindac, acemethacin, and dexibuprofen) were also found to be associated with an increased risk of heart failure, although the 95% confidence intervals included the null value. All nine significant associations identified in this analysis were also identified as significant by the Bonferroni-Holm procedure (supplementary table S2).

Compared with current use of celecoxib, current use of other individual NSAIDs was not associated with a significant decrease in heart failure risk. Odds ratios ranged from 0.83 (95% confidence interval 0.57 to 1.20) for oxaprozin to 1.84 (1.67 to 2.04) for ketorolac (supplementary table S3).

For the nine individual NSAIDs significantly associated with heart failure risk, their association was also confirmed regardless of whether there was recorded evidence of a prior heart failure diagnosis and regardless of sex (table 3). The estimated risk of heart failure associated with current use of NSAIDs of nimesulide, etoricoxib, and indomethacin among women was lower in magnitude than among men, compared with past use of any NSAIDs.

According to meta-analytic analysis, current users of any NSAID had a 24% higher risk of heart failure risk than past users (odds ratio 1.24; 95% confidence interval 1.12 to 1.36; fig 2). In addition to the nine individual NSAIDs with significant associations with heart failure risk, we found current use of nabumetone was also associated with higher risk of heart

doi: 10.1136/bmj.i4857 | BMJ 2016;354:i4857 | the bmj

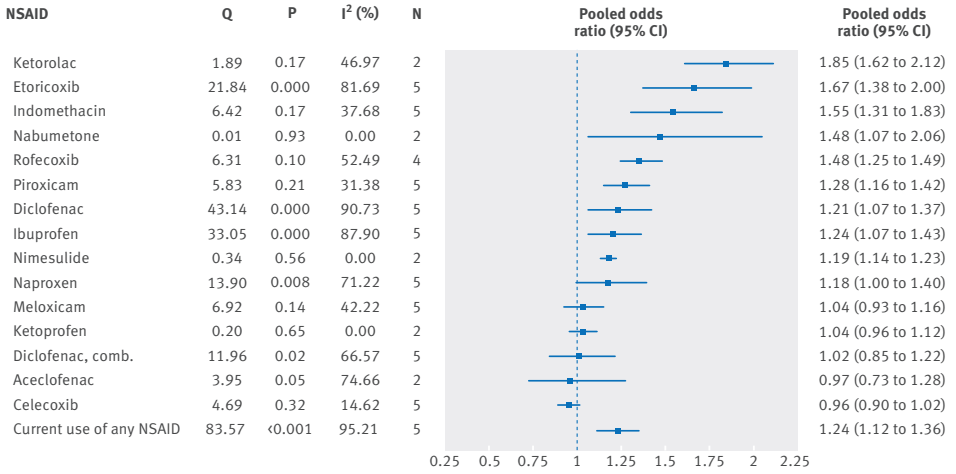

Fig 2 | **Summarised associations between current use of individual NSAIDs and risk of hospital admission for heart failure, compared with past use of any NSAID. Estimates (with 95% confidence intervals) were obtained by summarising database specific odds ratios (provided at least two were available) by use of the random effects meta-analytic approach. Heterogeneity between database specific odds ratios was assessed by Cochran's Q (and corresponding P value) and Higgins' I^2 statistics. No=number of summarised databases**

failure (fig 2). Although between database heterogeneity was relevant (I^2>70%), meta-analytic estimates of odds ratios were generally consistent with corresponding values obtained from the analysis of pooled individual level data.

Dose-response relation

Twenty (0.2%) cases and 855 (0.1%) controls from PHARMO and 753 (4.3%) cases and 61 777 (4.3%) controls from THIN were excluded because prescribed daily dose data were not recorded. The remaining 25 179 cases and 2 083 706 controls gathered from PHARMO and THIN entered the dose-response analysis.

Current users of very high doses of diclofenac, etoricoxib, indomethacin, piroxicam, and rofecoxib had more than a twofold higher risk of heart failure than past users (fig 3). The odds ratio associated with current high dose use of ibuprofen was also compatible with an increased risk of heart failure, despite the wide confidence interval. Finally, there was no evidence that celecoxib increased the risk of hospital admission heart failure at commonly used doses compared with past use of any NSAIDs. However, we cannot exclude an increase in risk when celecoxib is used at very high doses, given the wide confidence intervals obtained for this dose class.

Supplementary findings

Supplementary tables S4-S7 report the distribution of case and controls according to the considered covariates, use of individual NSAIDs, and dose categories of current NSAIDs use (in DDD equivalents and corresponding daily amount of active principle in mg), as well as the effects of individual NSAIDs on the heart failure risk.

Discussion

Principal findings

Our study, based on real world data on almost 10 million NSAIDs users from four European countries, provides evidence that current use of both COX 2 inhibitors and traditional individual NSAIDs are associated with increased risk of heart failure. Furthermore, the magnitude of the association varies between individual NSAIDs and according to the prescribed dose.

NSAIDs inhibit the isoenzymes of prostaglandin G/H synthase, COX 1 and COX 2.[1] The overall effects of this inhibition of the prostaglandin synthesis are to increase peripheral systemic resistance and reduce renal perfusion, glomerular filtration rate, and sodium excretion in susceptible individuals.[32 33] Taken together, these mechanisms could trigger clinical manifestations of heart failure, especially in susceptible patients.[23] Additionally, because the level of prostaglandin inhibition mediated by NSAIDs increases with dose,[14 34] the risk of clinical heart failure could be expected to increase along with the used NSAIDs dose.

Our study found an increased risk of hospital admission for heart failure in association with current use of several traditional NSAIDs (diclofenac, ibuprofen, indomethacin, ketorolac, naproxen, nimesulide, piroxicam, and possibly nabumetone) and two COX 2 inhibitors (etoricoxib and rofecoxib). We confirmed these findings after adjusting for multiple comparisons. Additionally, we found evidence that the increased risk of heart failure also affected patients without prior outpatient diagnosis or secondary hospital diagnosis heart failure—that is, those ideally less susceptible to heart failure decompensations. We also observed an increasing dose dependent risk of heart failure for most individual NSAIDs. Finally, indomethacin and etoricoxib

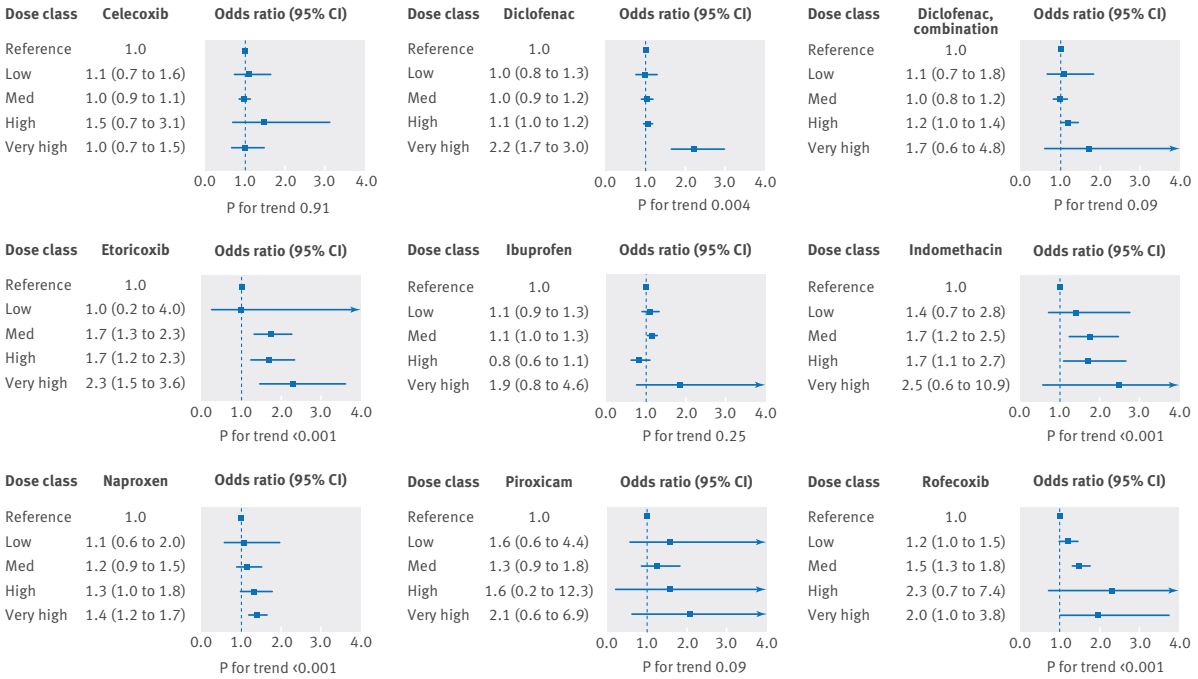

Fig 3 | **Dose-response relation between currently prescribed doses of specific NSAIDs and risk of heart failure, compared with past use of any NSAID. Pooled data were obtained from the Netherlands (PHARMO) and UK (THIN) databases for this analysis. Currently prescribed doses of each NSAID categorised as low (0.8 defined daily dose equivalents), medium (0.9-1.2), high (1.3-1.9), and very high (≥2). Odds ratios and 95% confidence intervals estimated by fitting a conditional logistic regression model after correcting for available covariates**

seemed to increase the risk of hospital admission for heart failure even if used at medium doses.

No significant differences in the magnitude of the association between use of individual NSAIDs and heart failure risk were found between patients with or without prior heart failure (for all NSAIDs) and between the sexes (with a few exceptions of NSAIDs). However, power of our analysis could have been too low to detect significant differences between the considered subgroups.

Our study did not find that celecoxib, the most widely prescribed selective COX 2 inhibitor, increases the risk of hospital admission for heart failure. Lack of statistical power is unlikely explain such lack of evidence, because our main analysis had 80% power to detect significant odd ratios as low as 1.08 for the current use of celecoxib.[35] Celecoxib also did not show an increased heart failure risk when used at the highest doses, although power of our analysis was low for this dose class (about 30% power to detect significant odds ratios of 2.00). Furthermore, our study found little evidence that celecoxib is associated with a greater risk of heart failure than any of the other 26 considered individual NSAIDs.

Comparison with other studies

Our findings extend those of the meta-analysis of randomised trials,[11] which showed that the risk of hospital admission for heart failure was roughly doubled by all studied NSAID regimens compared with placebo. Similarly, a meta-analysis of six trials did not show differences in heart failure risk between traditional NSAIDs and COX 2 inhibitors.[13] Estimates provided by the few published observational studies on the NSAID heart failure association are compatible with an increased risk of heart failure associated with naproxen, ibuprofen, ketoprofen, piroxicam, indomethacin, and rofecoxib, but not for celecoxib.[14 23-27] Our results also accord with the body of evidence supporting the relative cardiovascular safety of low to medium doses of celecoxib for treatment of arthritis compared with all other COX 2 inhibitors.[28]

Taken together, our findings support the hypothesis that selective and non-selective COX 2 inhibitors increase the risk of heart failure, but that the magnitude of this effect varies between individual drugs and according to the dose used.[32] The effect of individual NSAIDs could depend on a complex interaction of pharmacological properties, including duration and extent of platelet inhibition, extent of blood pressure increase, and properties possibly unique to the molecule.[28]

Strengths and limitations of study

Our findings, which focused only on prescription NSAIDs, might apply to NSAIDs obtained over the counter as well. Although over-the-counter NSAIDs are probably typically used at lower doses, by younger people, and for shorter durations than prescribed NSAIDs,

doi: 10.1136/bmj.i4857 | *BMJ* 2016;354:i4857 | the bmj

they are sometimes available at the same doses than those prescribed[36] and may be inappropriately overused.[37] Therefore, our findings could have large scale consequences in public health and further research needs to assess the safety of over-the-counter NSAIDs under the conditions they are typically used.

The present study, conducted as part of the EU funded SOS Project, is based on data from large and unselected populations and obtained by combining different healthcare databases together. The same approach was considered in several other EU funded projects addressing various issues on drug safety, such as the arrhythmogenic risk of drugs (ARITMO project), safety of vaccines (VAESCO project), and detection of adverse drug reactions (EU ADR project).[18 19] The use of five different data sources from the SOS Project should be considered a strength of this study because it allowed us to compare the risk of heart failure associated with many individual NSAIDs as used in different populations and healthcare systems from four EU countries.

Our study had some limitations. Firstly, our study might not have captured all NSAID exposure, because some of these drugs (eg, ibuprofen) are also available over the counter in all the four countries. Hence, patients classified as non-current users of NSAIDs in this study might actually have been current users of over the counter NSAIDs. Such misclassification would tend to, on average, bias estimates toward the null,[38 39] with the implication that our findings might understate the actual association between use of individual NSAIDs and heart failure risk.

Secondly, validity of outcome ascertainment might be of concern because heart failure is often associated with other cardiovascular diseases (eg, myocardial infarction), which could affect how hospital discharge codes are recorded. Nevertheless, although privacy concerns inhibited the validation of records in most participating databases, the positive predictive value for heart failure hospital admissions included in the Italian OSSIFF database was found to be 80% (95% confidence interval 66% to 90%). Additionally, high positive predictive values have been reported by other investigations based on healthcare databases for heart failure diagnosis codes at hospital discharge considered in our study.[40] In fact, the incidence of almost 37.5 heart failure cases every 10 000 person years observed in our study does not substantially differ from rates reported by available population based studies.[41] Still, even with some outcome misclassification,[42] this is expected to be non-differential—that is, independent of current use of NSAIDs—leading to a bias moving estimated associations towards the null.[43] However, non-differential misclassification (of outcome or exposure) might lead to inflated observed associations due to chance alone.[44]

Thirdly, our dose-response analysis could have been underpowered for some NSAID dose classes because only the PHARMO and THIN databases could be considered. Additionally, a portion of patients registered in these two databases had to be excluded from the dose-response analysis because they lacked the prescribed daily dose information. Although this exclusion might have led to some bias,[45] the number of excluded individuals was low and is unlikely to have had a significant effect on the results.

Fourthly, the effect of heterogeneous patient characteristics at baseline must be considered in the interpretation of our findings. Some individual NSAIDs more frequently used for different acute or chronic indications could have resulted in different patterns of use as well as in different types of populations of users.[46] To address this possibility, we adjusted pooled estimates for several demographic, therapeutic, and clinical characteristics (including osteoarthritis, rheumatoid arthritis and inflammatory polyarthritis) at baseline, measured in all the included data sources. In addition, estimates did not substantially change in the random effects meta-analytic approach, where database specific estimates were adjusted for all baseline covariates available in the considered data source. Relative risk estimates for individual NSAIDs among patients with prior outpatient or secondary hospital diagnoses of heart failure (that is, those with a contraindication for NSAID use who also should be more susceptible for acute clinical manifestations of heart failure) did not seem to differ substantially from those obtained in the overall analysis. Taken together, these results provide some protection to our findings. Nevertheless, we cannot exclude that residual differences in patient's baseline characteristics could account for some of the observed variations in relative risk estimates associated with different individual NSAIDs.

Lastly, some diseases that modify both the risk of heart failure and probability of current NSAID use might not have been fully accounted for in this study. To protect against this possibility, we adjusted all our estimates for concomitant (that is, in the current period) use of specific drugs (eg, nitrates, diuretics, or other drugs for cardiovascular diseases) as a proxy of patients' current health status. Still, residual confounding cannot be excluded. For example, gout is potentially an uncontrolled confounder of the association between current use of NSAIDs and heart failure risk in this study. This is because gout is an independent risk factor for heart failure,[47] and NSAIDs are the first pharmacological choice for treating acute gout episodes.[48] However, the following considerations further strengthen our conclusions. We assumed that gout has a 1% prevalence in our source population and that it increases heart failure risk by 1.74-fold.[47 49] With these figures, we estimated[50] that, to fully explain the observed association between naproxen and heart failure (naproxen being the NSAID with the weakest statistically significant association with heart failure in this study), acute gout episodes should have increased the odds of being treated in the current period rather than the past period by 33-fold, an implausibly high amount.

Conclusions

Our study offers further evidence that the most frequently used individual traditional NSAIDs and selective COX 2 inhibitors are associated with an increased risk of hospital admission for heart failure. Moreover,

the risk seems to vary between drugs and according to the dose. For the individual NSAIDs less frequently used, we were not able to exclude a risk of low to moderate magnitude owing to the limited numbers of exposed cases identified in this study. Because any potential increased risk could have a considerable impact on public health, the risk effect estimates provided by this study may help inform both clinical practices and regulatory activities.

We thank all members of the Safety of Non-steroidal Anti-inflammatory Drugs (SOS) Project Consortium for their collaborative efforts.

Contributors: All authors were members of the SOS Project Consortium. AA, LS, FN, AZ, BK, TS, HS, SL, and RS were involved in data management and statistical analysis. FT, AP, and MaS were responsible for the data harmonisation procedure and data preparation. CV-L, RH, MV, VV, and SR helped to interpret the findings and manage the study. MiS was the principal investigator of the SOS Project. AA, LS, FN, AZ, and GC (guarantors) drafted the manuscript. All listed authors helped revise the final manuscript.

Funding: The research leading to the results of this study received funding from the European Community's seventh Framework Programme (grant agreement no 223495—the SOS Project). The funding source did not play any role in planning and implementing this study, in interpreting its results, and in writing this paper.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: GC collaborated with the advisory boards of Novartis and Roche and participated in projects funded by GlaxoSmithKline (GSK). HS and RH are employees of the PHARMO Institute, an independent research institute that performs financially supported studies for the government and related healthcare authorities and several pharmaceutical companies in the European Union. BK and TS work in departments that occasionally perform studies for the pharmaceutical companies, including Bayer-Schering, Celgene, GSK, Mundipharma, Novartis, Purdue, Sanofi-Aventis, Sanofi-Pasteur, Stada, and Takeda. EG runs a department that occasionally performs studies for pharmaceutical industries, including Bayer, Celgene, GSK, Mundipharma, Novartis, Sanofi, Sanofi Pasteur MSD, and STADA; has been a consultant to Bayer, Nycomed, Teva, GSK, Schwabe, and Novartis (the SOS Project was not funded or cofunded by any of these companies). SL and MV, as employees of the local health authority of Cremona, have perfomed research studies sponsored by pharmaceutical companies (Pfizer Italia, GSK, and Novartis V&D) unrelated to this study. CV-L, as an employee of RTI Health Solutions, worked on projects funded by pharmaceutical companies including manufacturers of treatments for pain and inflammation; and participates in advisory boards funded by pharmaceutical companies. MaS has, since completion of this research, accepted a full time position at Janssen R&D. VV, as an employee of Erasmus MC, has conducted research for AstraZeneca. MiS is head of a unit that conducts some research for pharmaceutical companies Pfizer, Novartis, Lilly, and Altana (the SOS Project was not funded or cofunded by any of these companies). All other authors have no conflicts of interest to declare.

Ethical approval: Ethical approval was not required for this study.

Data sharing: No additional data available.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 FitzGerald GA, Patrono C. The coxibs, selective inhibitors of cyclooxygenase-2. *N Engl J Med* 2001;345:433-42. doi:10.1056/NEJM200108093450607.
2 Fries JF. Selective cyclooxygenase inhibition: promise for future NSAID therapy?*Scand J Rheumatol Suppl* 1996;102:1. doi:10.3109/03009749609097224.
3 Bombardier C, Laine L, Reicin A, et al. VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000;343:1520-8, 2, 1528. doi:10.1056/NEJM200011233432103.
4 Amer M, Bead VR, Bathon J, Blumenthal RS, Edwards DN. Use of nonsteroidal anti-inflammatory drugs in patients with cardiovascular disease: a cautionary tale. *Cardiol Rev* 2010;18:204-12. doi:10.1097/CRD.0b013e3181ce1521.
5 Bresalier RS, Sandler RS, Quan H, et al. Adenomatous Polyp Prevention on Vioxx (APPROVe) Trial Investigators. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* 2005;352:1092-102. doi:10.1056/NEJMoa050493.
6 Solomon SD, McMurray JJ, Pfeffer MA, et al. Adenoma Prevention with Celecoxib (APC) Study Investigators. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 2005;352:1071-80. doi:10.1056/NEJMoa050405.
7 Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *BMJ* 2006;332:1302-8. doi:10.1136/bmj.332.7553.1302.
8 García Rodríguez LA, Tacconelli S, Patrignani P. Role of dose potency in the prediction of risk of myocardial infarction associated with nonsteroidal anti-inflammatory drugs in the general population. *J Am Coll Cardiol* 2008;52:1628-36. doi:10.1016/j.jacc.2008.08.041.
9 McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. *PLoS Med* 2011;8:e1001098. doi:10.1371/journal.pmed.1001098.
10 Trelle S, Reichenbach S, Wandel S, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *BMJ* 2011;342:c7086. doi:10.1136/bmj.c7086.
11 Bhala N, Emberson J, Merhi A, et al. Coxib and traditional NSAID Trialists' (CNT) Collaboration. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet* 2013;382:769-79. doi:10.1016/S0140-6736(13)60900-9.
12 Fabule J, Adebajo A. Comparative evaluation of cardiovascular outcomes in patients with osteoarthritis and rheumatoid arthritis on recommended doses of nonsteroidal anti-inflammatory drugs. *Ther Adv Musculoskelet Dis* 2014;6:111-30. doi:10.1177/1759720X14541668.
13 Scott PA, Kingsley GH, Scott DL. Non-steroidal anti-inflammatory drugs and cardiac failure: meta-analyses of observational studies and randomised controlled trials. *Eur J Heart Fail* 2008;10:1102-7. doi:10.1016/j.ejheart.2008.07.013.
14 García Rodríguez LA, Hernández-Díaz S. Nonsteroidal antiinflammatory drugs as a trigger of clinical heart failure. *Epidemiology* 2003;14:240-6. doi:10.1097/01.EDE.0000034633.74133.C3.
15 McMurray JJ, Adamopoulos S, Anker SD, et al. ESC Committee for Practice Guidelines. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2012;33:1787-847. doi:10.1093/eurheartj/ehs104.
16 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158:915-20. doi:10.1093/aje/kwg231.
17 Valkhoff VE, Schade R, 't Jong GW, et al. Safety of Non-steroidal Anti-inflammatory Drugs (SOS) project. Population-based analysis of non-steroidal anti-inflammatory drug use among children in four European countries in the SOS project: what size of data platforms and which study designs do we need to assess safety issues?*BMC Pediatr* 2013;13:192. doi:10.1186/1471-2431-13-192.
18 Avillach P, Coloma PM, Gini R, et al. EU-ADR consortium. Harmonization process for the identification of medical events in eight European healthcare databases: the experience from the EU-ADR project. *J Am Med Inform Assoc* 2013;20:184-92. doi:10.1136/amiajnl-2012-000933.
19 Trifirò G, Coloma PM, Rijnbeek PR, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? *J Intern Med* 2014;275:551-61. doi:10.1111/joim.12159.
20 Hosmer DW, Lemeshow S. *Applied Logistic Regression*.2nd ed. John Wiley & Sons, 2000. doi:10.1002/0471722146.
21 Wright SP. Adjusted p-values for simultaneous inference. *Biometrics* 1992;48:1005-13. doi:10.2307/2532694.
22 Sedgwick P. Pitfalls of statistical hypothesis testing: multiple testing [correction in: *BMJ* 2014;349:g5624]. *BMJ* 2014;349:g5310. doi:10.1136/bmj.g5310.
23 Huerta C, Varas-Lorenzo C, Castellsague J, García Rodríguez LA. Non-steroidal anti-inflammatory drugs and risk of first hospital admission for heart failure in the general population. *Heart* 2006;92:1610-5. doi:10.1136/hrt.2005.082388.
24 Bernatsky S, Hudson M, Suissa S. Anti-rheumatic drug use and risk of hospitalization for congestive heart failure in rheumatoid arthritis. *Rheumatology (Oxford)* 2005;44:677-80. doi:10.1093/rheumatology/keh610.

25 Mamdani M, Juurlink DN, Lee DS, et al. Cyclo-oxygenase-2 inhibitors versus non-selective non-steroidal anti-inflammatory drugs and congestive heart failure outcomes in elderly patients: a population-based cohort study. *Lancet* 2004;363:1751-6. doi:10.1016/S0140-6736(04)16299-5.
26 Mangoni AA, Woodman RJ, Gaganis P, Gilbert AL, Knights KM. Use of non-steroidal anti-inflammatory drugs and risk of incident myocardial infarction and heart failure, and all-cause mortality in the Australian veteran community. *Br J Clin Pharmacol* 2010;69:689-700. doi:10.1111/j.1365-2125.2010.03627.x.
27 McGettigan P, Han P, Jones L, Whitaker D, Henry D. Selective COX-2 inhibitors, NSAIDs and congestive heart failure: differences between new and recurrent cases. *Br J Clin Pharmacol* 2008;65:927-34. doi:10.1111/j.1365-2125.2008.03121.x.
28 Howes LG. Selective COX-2 inhibitors, NSAIDs and cardiovascular events - is celecoxib the safest choice?*Ther Clin Risk Manag* 2007;3:831-45.
29 Cooper H, Patall EA. The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods* 2009;14:165-76. doi:10.1037/a0015565.
30 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88. doi:10.1016/0197-2456(86)90046-2.
31 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58. doi:10.1002/sim.1186.
32 Patrono C, Baigent C. Nonsteroidal anti-inflammatory drugs and the heart. *Circulation* 2014;129:907-16. doi:10.1161/CIRCULATIONAHA.113.004480.
33 Bleumink GS, Feenstra J, Sturkenboom MC, Stricker BH. Nonsteroidal anti-inflammatory drugs and heart failure. *Drugs* 2003;63:525-34. doi:10.2165/00003495-200363060-00001.
34 Bäck M, Yin L, Ingelsson E. Cyclooxygenase-2 inhibitors and cardiovascular risk in a nation-wide cohort study after the withdrawal of rofecoxib. *Eur Heart J* 2012;33:1928-33. doi:10.1093/eurheartj/ehr421.
35 Lui KJ. Estimation of sample sizes in case-control studies with multiple controls per case: dichotomous data. *Am J Epidemiol* 1988;127:1064-70.
36 Duong M, Salvo F, Pariente A, et al. Usage patterns of 'over-the-counter' vs. prescription-strength nonsteroidal anti-inflammatory drugs in France. *Br J Clin Pharmacol* 2014;77:887-95. doi:10.1111/bcp.12239.
37 Koffeman AR, Valkhoff VE, Celik S, et al. High-risk use of over-the-counter non-steroidal anti-inflammatory drugs: a population-based cross-sectional study. *Br J Gen Pract* 2014;64:e191-8. doi:10.3399/bjgp14X677815.
38 Yood MU, Campbell UB, Rothman KJ, et al. Using prescription claims data for drugs available over-the-counter (OTC). *Pharmacoepidemiol Drug Saf* 2007;16:961-8. doi:10.1002/pds.1454.
39 Delaney JA, Biggs ML, Kronmal RA, Psaty BM. Demographic, medical, and behavioral characteristics associated with over the counter non-steroidal anti-inflammatory drug use in a population-based cohort: results from the Multi-Ethnic Study of Atherosclerosis. *Pharmacoepidemiol Drug Saf* 2011;20:83-9. doi:10.1002/pds.2065.
40 Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21(Suppl 1):129-40. doi:10.1002/pds.2313.
41 Corrao G, Ghirardi A, Ibrahim B, Merlino L, Maggioni AP. Burden of new hospitalization for heart failure: a population-based investigation from Italy. *Eur J Heart Fail* 2014;16:729-36. doi:10.1002/ejhf.105.
42 Romano PS, Mark DH. Bias in the coding of hospital discharge data and its implications for quality assessment. *Med Care* 1994;32:81-90. doi:10.1097/00005650-199401000-00006.
43 Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488-95.
44 Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol* 2005;34:680-7. doi:10.1093/ije/dyi060.
45 Altman DG, Bland JM. Missing data. *BMJ* 2007;334:424. doi:10.1136/bmj.38977.682025.2C.
46 Moride Y, Ducruet T, Boivin JF, Moore N, Perreault S, Zhao S. Prescription channeling of COX-2 inhibitors and traditional nonselective nonsteroidal anti-inflammatory drugs: a population-based case-control study. *Arthritis Res Ther* 2005;7:R333-42. doi:10.1186/ar1488.
47 Krishnan E. Gout and the risk for incident heart failure and systolic dysfunction. *BMJ Open* 2012;2:e000282. doi:10.1136/bmjopen-2011-000282.
48 Kim KY, Ralph Schumacher H, Hunsche E, Wertheimer Al, Kong SX. A literature review of the epidemiology and treatment of acute gout. *Clin Ther* 2003;25:1593-617. doi:10.1016/S0149-2918(03)80158-3.
49 Smith EU, Díaz-Torné C, Perez-Ruiz F, March LM. Epidemiology of gout: an update. *Best Pract Res Clin Rheumatol* 2010;24:811-27. doi:10.1016/j.berh.2010.10.004.
50 Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006;15:291-303. doi:10.1002/pds.1200.

Supplementary appendix: Additional material

Appendix C

Paper: Comparison of basal insulin therapies with regard to the risk of acute myocardial infarction in patients with type 2 diabetes: an observational study

Contribution to the manuscript I herewith certify that I contributed to the design of the study, performed the statistical analyses, interpreted the results, and drafted the manuscript.
Due to copyright issues, the paper was removed from the published version of the thesis. Only the abstract is given here.

Comparison of basal insulin therapies with regard to the risk of acute myocardial infarction in patients with type 2 diabetes: an observational study

B Kollhorst[1], S Behr[1], D Enders[1], FW Dippel[2], K Theobald[2], E Garbe[1,3]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[2]Sanofi Aventis Deutschland GmbH, Berlin, Germany
[3]Core Scientific Area 'Health Sciences', University of Bremen, Bremen, Germany

Abstract

Aims: To assess the risk of acute myocardial infarction (AMI) in patients with type 2 diabetes mellitus treated with long-acting insulin analogues in comparison with other basal insulin therapy.

Methods: We used German insurance claims data from the years 2004-2009 to conduct a study in a retrospective cohort of patients with type 2 diabetes. Naïve insulin users were defined as patients who had an insulin-free history before the first prescription of long-acting analogue insulin, human NPH insulin or premixed insulin and who were pretreated with oral antidiabetic drugs. Adjusted hazard ratios (HRs) of AMI and corresponding 95% confidence intervals (CIs) were calculated using sex-stratified Cox models. Propensity-score-matched analyses were conducted as sensitivity analyses.

Results: We identified 21,501 new insulin users. Patients treated with premixed insulin were older than patients treated with analogue or NPH insulin (mean age 70.7 vs. 64.1 and 61.6?years, respectively) and had more comorbidities. Regarding the risk of AMI, adjusted HRs showed no statistically significant difference between NPH and analogue insulin (HR 0.94, 95% CI 0.74-1.19), but a higher risk for premixed than for analogue insulin (HR 1.27, 95% CI 1.02-1.58). Contrary to the primary analysis, the propensity-score-

matched analysis did not show an increased risk for premixed insulin.

Conclusions: In contrast to a former database study, no difference was observed for the risk of AMI between long-acting analogue and NPH insulin in this study. Neither long-acting analogue insulin nor premixed insulin appears to be associated with AMI in patients with type 2 diabetes.

Appendix D

Paper: Comparative risk of death of antidepressants in older patients with depression

Contribution to the manuscript I herewith certify that I contributed to the design of the study, performed the statistical analyses, interpreted the results, and drafted the manuscript.
This manuscript will be submitted shortly after submission of this thesis. Only the abstract is presented here.

Comparative risk of death of antidepressants in older patients with depression

B Kollhorst[1], K Jobski[1,2], J Krappweis[3], T Schink[1], E Garbe[1], N Schmedt[1,4]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[2]Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[3]Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany
[4]InGef - Institute for Applied Health Research, Berlin, Germany

Abstract

Importance Antidepressants (ADs) are frequently used in older patients with depression, but little is known about the comparative safety of individual agents.

Objective To determine the comparative risk of death of ADs in older patients with depression.

Design Retrospective population-based cohort study between 2005 and 2012.

Setting Claims data from four statutory health insurance providers from the German Pharmacoepidemiological Research Database.

Participants Patients aged $\geq$ 65 years with depression initiating treatment with ADs.

Exposure(s) Dispensations of opipramol, trimipramine, amitriptyline, doxepin, fluoxetine, paroxetine, sertraline, escitalopram, venlafaxine, duloxetine, mirtazapine, and St. John's wort compared to citalopram.

Main Outcome(s) and Measure(s) Cox models were used to estimate crude and adjusted hazard ratios (HRs) with 95% confidence intervals (CIs) of indi-

vidual ADs compared to citalopram for the main outcome all-cause mortality. Subgroup analyses by age and dementia status were conducted. In sensitivity analyses, adjustment by high-dimensional propensity score (HdPS) was used to address unmeasured confounding.

Results The cohort included 259,920 new users of ADs of whom 6.2% died during follow-up. The majority of patients entered the cohort with citalopram (21%), followed by mirtazapine (17%), and amitriptyline (16%). In the primary analysis, amitriptyline was associated with an increased risk of death (HR, 1.14; 95% CI, 1.08-1.20) relative to citalopram, whereas opipramol, trimipramine, doxepin, mirtazapine, duloxetine, venlafaxine, and St. John's wort were associated with a lower risk of death. With an exclusion of patients with a history of cancer, the risk of amitriptyline vs. citalopram was no longer significantly elevated. After adjustment by HdPS, in patients $\geq$ 80 years or with dementia, HRs for all ADs tended toward a null effect.

Conclusions and Relevance This study suggests that ADs recommended as first-line treatment in patients with depression have a similar safety profile with regard to the risk of death, especially in patients $\geq$ 80 years and with dementia. Although the slightly elevated risk of death observed for amitriptyline may be triggered by confounding, its use should be avoided except in patients with co-existing indications such as pain. The higher potential for side effects of other tricyclic ADs does not seem to affect the risk of death.

Appendix E

Paper: The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs

Contribution to the manuscript I herewith certify that I contributed to the design of the study, performed the statistical analyses, interpreted the results, and drafted the manuscript.
Due to copyright issues, the paper was removed from the published version of the thesis. Only the abstract is given here.

The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs

B Kollhorst[1], M Abrahamowicz[2,3], I Pigeot[1,4]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Department Biometry and Data Management, Achterstr. 30, Bremen 28359, Germany
[2]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal H3A 1A2, Québec, Canada
[3]Division of Clinical Epidemiology, McGill University Health Centre Royal Victoria Hospital, 687 Pine Avenue West, Montreal H3A 1A1, Québec, Canada
[4]Faculty of Mathematics and Computer Science, University of Bremen, Bibliothekstraße 1, Bremen 28359, Germany

Abstract

Objectives: To investigate whether physician's prescribing preference is a valid instrumental variable (IV) for patients' actual prescription of selective cyclooxygenase-2 (COX-2) inhibitors in the German Pharmacoepidemiological Research Database (GePaRD).

Study design and setting: We compared the effect of COX-2 inhibitors vs. traditional nonsteroidal anti-inflammatory drugs (tNSAIDs) on the risk of gastrointestinal complications using physician's preference as IV. We used different definitions of physician's preference for COX-2 inhibitors. A retrospective cohort of new users was built which was further restricted to subcohorts. We compared IV-based risk difference estimates, using a two-stage approach, to estimates from conventional multivariate models.

Results: We observed only a small proportion of COX-inhibitor users (3.2%) in our study. All instruments, in the full cohort and in the subcohorts, reduced the imbalance in most of the covariates. However, the IV treatment

effect estimates had a highly inflated variance. Compared to the most recent prescription, the proportion of previous patients was a stronger instrument and reduced the variance of the estimates.

Conclusion: The proportion of all previous patients is a potential IV for comparing COX-2 inhibitors vs. tNSAIDs in GePaRD. Our study demonstrates that valid instruments in one health care system may not be directly applicable to others.

Appendix F

Mathematical details

F.1 Proof of (3.17)

Wooldridge (2010) showed that $\boldsymbol{\theta_0}$ is the value that minimizes $E[(Y-q(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2]$. Let us first rewrite $(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2$ as follows:

$$\begin{aligned}(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2 =&(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})+q(\boldsymbol{X},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2\\ =&(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2\\ &+2(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))(q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{\alpha^*}))\\ &+(q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2.\end{aligned}$$

Assuming $E(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})|\boldsymbol{X})=0$, ε_{2SRI} is not correlated with any function of $\boldsymbol{S}$. Using the law of iterated expectations, it follows immediately:

$$\begin{aligned}E[(Y-q(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2] =&E[(Y-q(\boldsymbol{X},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2]\\ &+E[(q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2].\end{aligned} \tag{F.1}$$

Since $E[(q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2]\geq 0$, the following inequality is true:

$$E[(Y-q(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2]\geq E[(Y-q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*}))^2] \text{ for } \boldsymbol{\theta}\in\Theta.$$

In order to obtain a unique solution $\boldsymbol{\theta_0}$, it is assumed that $E[(q(\boldsymbol{S},\boldsymbol{\theta_0},\boldsymbol{\alpha^*})-q(\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\alpha^*}))^2]>0$ for all $\boldsymbol{\theta}\in\Theta,\boldsymbol{\theta}\neq\boldsymbol{\theta_0}$, which implies that the inequality is strict. □

Appendix G

Observational study of mortality risks between new users of conventional and atypical antipsychotics

G.1 Objectives

To compare the performance of the IV models in a real-life database study, they are applied to a study of the risk of death between new users of conventional and atypical antipsychotics (APs) among elderly patients. As atypical antipsychotics are thought to be less sedating and less likely to cause extrapyramidal disorders, it is suspected that they are selectively prescribed to frail patients (Brookhart et al., 2007; Schneeweiss et al., 2007). Furthermore, factors as cognitive and physical impairment are typically not recorded in administrative health databases and therefore, the conventional Cox model will be biased due to unmeasured confounding.

G.2 Methods

G.2.1 Study design

A retrospective study in a cohort of patients who initiated treatment with antipsychotics between 2005 and 2012 is conducted using claims data extracted from GePaRD. Patients are included in the cohort if they are continuously insured for at least 365 days before their first AP dispensation, are at least 65 years old and have no diagnosis of a malignant cancer, except nonmelanoma skin cancer, in this period. Patients with a diagnosis of cancer before the first AP prescription are excluded to avoid residual confounding introduced by selective prescribing of conventional antipsychotic medications as antiemetics, because these patients are more likely to die independent of drug use. Patients with multiple dispensations of APs at cohort entry and patients with missing information on the prescribing physician are excluded, because they cannot be included in the IV analyses. Cohort entry is defined as the first dispensation of an AP (index AP) during study period if all inclusion criteria are met, and patients are followed until either end or interruption of insurance time ($>$ 3 days), death, 180 days after cohort entry, or end of the study period (December 31, 2012).

G.2.2 Exposure and instruments

Patients are classified as users of either conventional or atypical APs, depending on the first prescription (N05A (excl. lithium) and promethazine (excl. dermal/topical formulations)) at cohort entry.
The binary instrument is an indicator variable, assigning the value of 1 or 0 if the prescription written to the most recent patient by the same physician is for a conventional or atypical AP, respectively. If two or more prescriptions are filled at the same day, one was randomly picked to determine the preference.

G.2.3 Outcome

The outcome is defined as all-cause mortality if the reason for hospital discharge or the reason for deregistration from the insurance is coded as death.

G.2.4 Confounders

Confounders include sociodemographic characteristics such as age at cohort entry and sex, comorbidities and co-medications. Comorbid conditions are measured within the 365 days before index drug use. Information on relevant conditions is obtained from inpatient and outpatient diagnoses. History of co-medications are assessed based on the date of dispensation within the 365 prior to cohort entry except anxiolytics, opioids, hypnotics and sedatives and antibacterials for systemic use which are assessed within 182 days prior to cohort.

G.2.5 Statistical analyses

To investigate whether the first IV assumption is satisfied, three measures of the strength of the association between the IV and the actual treatment are calculated. The three measures are: the partial F-statistic for the adjusted IV effect, the squared partial correlation r^2, and the estimated effect of the IV on the probability of treatment, quantified as the adjusted difference in prevalence (per 100 patients) (Rassen et al., 2009a,b). Neither the second nor the third IV assumption can be empirically verified, but the plausibility of both can be explored. To check whether the second assumption may be considered as being valid, the balance of measured covariates across levels of the treatment and levels of the instrument are assessed using a linear model as large imbalances in measured covariates may signal potential confounding by unmeasured covariates (Davies, 2015; Jackson and Swanson, 2015). Furthermore, estimates are adjusted for the year of AP prescription and standard errors are calculated robustly accounting for clustering by physician. As the third assumption cannot be explored based on the data, the association between the IVs and

benzodiazepines that are prescribed at the same day as the AP prescription is examined using a logistic model, adjusted for age, sex and year of the index AP prescription. A violation of the third assumption would occur if the physician's AP preference is associated with the concomitant prescribing of a potentially hazardous medication in the elderly. To account for the clustering by physician, the parameters and standard errors are estimated using a robust generalized estimating equation approach and a working variance-covariance matrix with an exchangeable structure. The conventional Cox model and models used in simulations (Section 4.4.3) are employed. All models are adjusted for the same selected confounders.

G.3 Results

Table G.1: Comparison of baseline characteristics of new users of conventional and atypical antipsychotics

Variable	Conventional AP (N=156,675) (%)	Atypical AP (N=74,399) (%)
Age $\geq$ 80 years	72,070 (46.0%)	31,897 (42.9%)
Male	45,629 (29.1%)	23,206 (31.2%)
Myocardial infarction	12,464 (8.0%)	4,533 (6.1%)
Congestive heart failure and cardiomyopathy	48,415 (30.9%)	19,030 (25.6%)
Atrial fibrillation	31,137 (19.9%)	12,159 (16.3%)
Other cardiac arrhythmias and conduction disorders	42344 (27.0%)	18,600 (25.0%)
Valvular disorders (incl. endocarditis)	23,177 (14.8%)	10,044 (13.5%)
Peripheral vascular disease	31,284 (20.0%)	13,394 (18.0%)
Ischemic stroke and sequelae	22,177 (14.2%)	9,795 (13.2%)
Other cerebrovascular disease	40,322 (25.7%)	21,627 (29.1%)
Dementia	53,086 (33.9%)	27,541 (37.0%)
Pulmonary circulation disorders	6,695 (4.3%)	2,281 (3.1%)

Table G.1: (continued)

Variable	Conventional AP (N=156,675) (%)	Atypical AP (N=74,399) (%)
Rheumatic arthritis/ collagen vascular disease	15,883 (10.1%)	7,288 (9.8%)
Extrapyramidal disorders	32,747 (20.9%)	20,446 (27.5%)
Other neurological disorders	12,778 (8.2%)	6,071 (8.2%)
Diabetes	46,013 (29.4%)	20,271 (27.2%)
Paraplegia / hemiplegia and immobility	16,427 (10.5%)	6,369 (8.6%)
Renal failure	29,354 (18.7%)	11,181 (15.0%)
Obesity	20,941 (13.4%)	9,142 (12.3%)
Coagulopathy	9,468 (6.0%)	3,698 (5.0%)
Weight loss	7,581 (4.8%)	3,238 (4.4%)
Fluid and electrolyte disorders	38,217 (24.4%)	14,633 (19.7%)
Deficiency anemia	10,130 (6.5%)	4,093 (5.5%)
Alcohol abuse	6,560 (4.2%)	2,483 (3.3%)
Drug abuse	4,396 (2.8%)	1,672 (2.2%)
Senility / nursing home residence	19,997 (12.8%)	7,701 (10.4%)
Infectious diseases	73,559 (47.0%)	30,737 (41.3%)
Any fracture of lower extremities	4,486 (2.9%)	1,780 (2.4%)
Venous thromboembolism and insufficiency	19,496 (12.4%)	8,930 (12.0%)
Pneumonia	14,287 (9.1%)	4,696 (6.3%)
Ventricular arrhythmia	2,235 (1.4%)	816 (1.1%)
Surgery	24,022 (15.3%)	8,858 (11.9%)
Other psychoses	6,728 (4.3%)	8,413 (11.3%)
Obsessive compulsive disorders	433 (0.3%)	349 (0.5%)
Agitation	10,856 (6.9%)	3,470 (4.7%)
Insulin	12,448 (7.9%)	4,847 (6.5%)
Cardiac glycosides	17,139 (10.9%)	6,503 (8.7%)
Beta-adrenergic agonists	4,609 (2.9%)	2,076 (2.8%)
Calcium antagonists	42,197 (26.9%)	18,078 (24.3%)

Table G.1: (continued)

Variable	Conventional AP (N=156,675) (%)	Atypical AP (N=74,399) (%)
ACE inhibitors	70,415 (44.9%)	31,366 (42.2%)
Angiotensin II antagonists	26,158 (16.7%)	11,622 (15.6%)
Lipid lowering drugs	42,371 (27.0%)	19,926 (26.8%)
Glucocorticoids	21,768 (13.9%)	7,686 (10.3%)
Opioids	32,304 (20.6%)	10,983 (14.8%)
Anti-parkinson drugs	10,942 (7.0%)	9,903 (13.3%)
Anxiolytics	24,557 (15.7%)	10,036 (13.5%)
Hypnotics and sedatives	55,759 (35.6%)	6,646 (8.9%)
Anti-dementia drugs	15,626 (10.0%)	11,069 (14.9%)
Respiratory drugs	24,589 (15.7%)	8,478 (11.4%)
Antibacterials for systemic use	46,194 (29.5%)	17,537 (23.6%)
Antineoplastic agents and immunosuppressants	1,416 (0.9%)	533 (0.7%)
Non-steroidal anti-inflammatory drugs	61,317 (39.1%)	27,612 (37.1%)
Charlson Comorbidity Index > 2	60,171 (38.4%)	25,113 (33.8%)
Hospitalized time $> 5\%$	41,860 (26.7%)	19,612 (26.4%)
Number of drug classes > 8	65,502 (41.8%)	26,528 (35.7%)

Table G.2: Strength of the association of the actual treatment with the instrumental variable

Instrument	N	Partial r^2	Partial F-statistic	Effect IV on exposure in % (95% CI)
IV based on the previous patient	231,074	0.028	6567.7	15.7 (15.3; 16.1)

Table G.3: Adjusted association between the instrumental variable and the coprescription of benzodiazepines at the same day as the index AP prescription

Instrument	Odds ratio (95% CI)	p-value
IV based on the previous patient	0.96 (0.91; 1.00)	0.0687

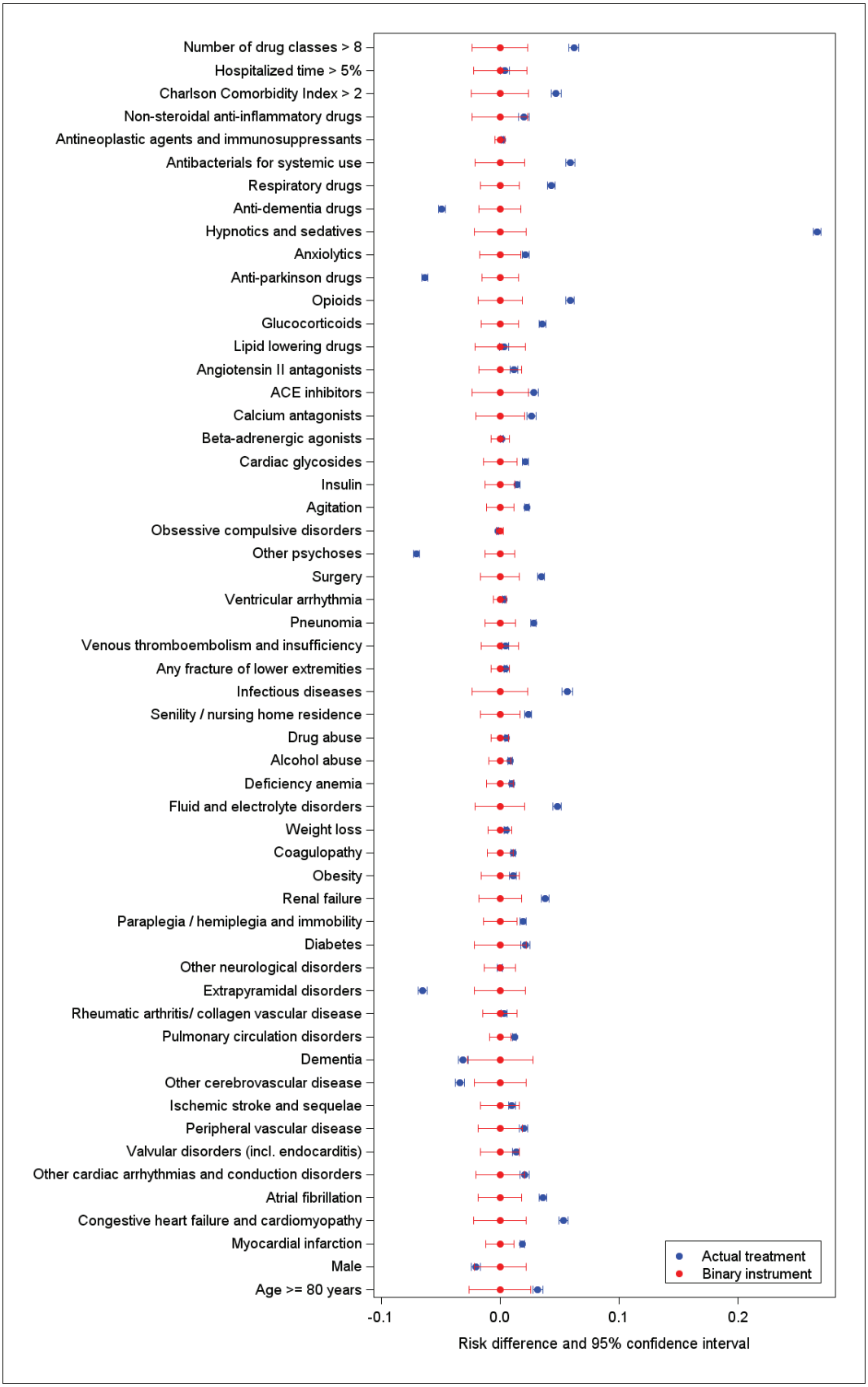

Figure G.1: Balance of measured covariates across levels of the treatment and levels of the instrument

Bibliography

Aalen, OO (1980). A model for non-parametric regression analysis of counting processes. *Lecture notes in statistics-2: mathematical statistics and probability theory*. Ed. by W Klonecki, A Kozek, and J Rosinski. New York: Springer, 1–25.

– (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* 8, 907–925.

Abrahamowicz, M, Bjerre, LM, Beauchamp, ME, LeLorier, J, and Burne, R (2016). The missing cause approach to unmeasured confounding in pharmacoepidemiology. *Statistics in Medicine* 30, 1001–1016.

Angrist, JD (2001). Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical pratice. *Journal of Business and & Economic Statistics* 19, 2–16.

Angrist, JD and Imbens, GW (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90, 431–442.

Angrist, JD, Imbens, GW, and Rubin, DB (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.

Arfè, A, Scotti, L, Varas-Lorenzo, C, Nicotra, F, Zambon, A, Kollhorst, B, Schink, T, Garbe, E, Herings, R, Straatman, H, Schade, R, Villa, M, Lucchi, S, Valkhoff, V, Romio, S, Thiessard, F, Schuemie, Martijn, Pariente, A, Sturkenboom, M, and Drugs (SOS) Project Consortium, G Corrao On be-

half of the Safety of Non-steroidal Anti-inflammatory (2016). Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: nested case-control study. *Bristish Medical Journal* 354, i4857.

Austin, PC (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46, 399–424.

Bierens, HJ (2004). *Introduction to the Mathematical and Statistical Foundations of Econometrics (Themes in Modern Econometrics)*. New York: Cambridge University Press.

Breslow, NE (1982). Design and analysis of case-control studies. *Annual Review of Public Health* 3, 29–54.

Brookhart, MA, Wang, PS, Solomon, DH, and Schneeweiss, S (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 17, 268–275.

Brookhart, MA, Rassen, JA, Wang, PS, Dormuth, C, Mogun, H, and Schneeweiss, S (2007). Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Medical Care* 45, 116–122.

Burgess, S (2015). Commentary: Consistency and collapsibility: are they crucial for instrumental variable analysis with a survival outcome in mendelian randomization? *Epidemiology* 26, 411–413.

Burne, RM and Abrahamowicz, M (2016). Martingale residual-based method to control for confounders measured only in a validation sample in time-to-event analysis. *Statistics in Medicine* 10, 4588–4606.

Concato, J, Peduzzi, P, Holford, TR, and Feinstein, AR (1995). Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. *Journal of Clinical Epidemiology* 48, 1495–1501.

Cox, DR (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 34, 406–424.

Cuzick, J, Sasieni, P, Myles, J, and Tyrer, J (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69, 565–588.

D'Agostino, RB (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 17, 2265–2281.

Davies, NM (2015). An even clearer portrait of bias in observational studies? *Epidemiology* 26, 505–508.

Davies, NM, Smith, GD, Windmeijer, F, and Martin, RM (2013). COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: an instrumental variable analysis. *Epidemiology* 224, 352–362.

Didelez, V, Meng, S, and Sheehan, NA (2010). Assumptions of IV methods for observational epidemiology. *Statistical Science* 25, 22–40.

Fahrmeir, L, Kneib, T, and Lang, S (2007). *Regression: Modelle, Methoden und Anwendungen*. Berlin: Springer-Verlag.

Garbe, E and Suissa, S (2014). Pharmacoepidemiology. *Handbook of epidemiology*. Ed. by W Ahrens and I Pigeot. New York; Heidelberg; Dordrecht; London: Springer, 259–291.

Greenland, S (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 29, 722–729.

Greenland, S, Pearl, J, and Robins, JM (1999). Causal diagrams for epidemiologic research. *Epidemiology* 10, 37–48.

Grimes, DA and Schulz, KF (2002). Bias and causal associations in observational research. *Lancet* 359, 248–52.

Guertin, JR, Rahme, E, and LeLorier, J (2016). Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *European Journal of Clinical Pharmacology* 72, 1497–1505.

Heckman, JJ and Robb, R (1985). Alternative methods for evaluating the impact of interventions. *Journal of Econometrics* 30, 239–267.

Hennessy, S (2006). Use of Health Care Databases in Pharmacoepidemiology. *Basic & Clinical Pharmacology & Toxicology* 98, 311–313.

Ionescu-Ittu, R, Delaney, JAC, and Abrahamowicz, M (2009). Bias-variance trade-off in pharmacoepidemiological studies using physician-preference--based instrumental variables: a simulation study. *Epidemiology* 17, 260–267.

Jackson, JW and Swanson, SA (2015). Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology* 26, 498–504.

Joffe, MM (2009). Exhaustion, automation, theory, and confounding. *Epidemiology* 20, 523–524.

Johnston, KM, Gustafson, P, Levy, AR, and Grootendorst, P (2008). Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Statistics in Medicine* 27, 1539–1556.

Kollhorst, B, Behr, S, Enders, D, Dippel, FW, and Garbe, E (2015). Comparison of basal insulin therapies with regard to the risk of acute myocardial infarction in patients with type 2 diabetes: an observational cohort study. *Diabetes, Obesity and Metabolism* 17, 1158–1165.

Kollhorst, B, Abrahamowicz, M, and Pigeot, I (2016). The proportion of all previous patients was a potential instrument for patients? actual prescriptions of nonsteroidal anti-inflammatory drugs. *Journal of Clinical Epidemiology* 69, 96–106.

Li, J, Fine, J, and Brookhart, A (2015). Instrumental variable additive hazards models. *Biometrics* 71, 122–130.

Lin, NX, Logan, S, and Henley, WE (2013). Bias and sensitivity analysis when estimating treatment effects from the Cox model with omitted covariates. *Biometrics* 69, 850–860.

Loeys, T, Goetghebeur, E, and Vandebosch, A (2005). Causal proportional hazards models and time-constant exposure in randomized clinical trials. *Lifetime Data Analysis* 11, 435–449.

MacKenzie, TA, Tosteson, TD, Morden, NE, Stukel, TA, and O'Malley, AJ (2014). Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Health Services and Outcomes Research Methodology* 14, 54–68.

Martens, EP, Pestman, WR, Boer, A de, Belitser, SV, and Klungel, OH (2006). Instrumental variables: applications and limitations. *Pharmacoepidemiology and Drug Safety* 18, 562–571.

Martinussen, T and Vansteenlandt, S (2013). On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis* 19, 279–296.

McCullagh, P and Nelder, JA (1989). *Generalized Linear Models*. London: Chapman & Hall.

Miller, AB, Goff Jr, DC, Bammann, K, and Wild, P (2014). Cohort studies. *Handbook of epidemiology*. Ed. by W Ahrens and I Pigeot. New York; Heidelberg; Dordrecht; London: Springer, 259–291.

Newey, WK and McFadden, D (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*. Ed. by RF Engle and D McFadden. Vol. 4. Amsterdam: Elsevier, 2111–2245.

Palmer, TM, Sterne, JA, Harbord, RM, Lawlor, DA, Sheehan, NA, Meng, S, Granell, R, Smith, GD, and Didelez, V (2011). Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *American Journal of Epidemiology* 173, 1392–1403.

Pearce, N and Greenland, S (2014). Confounding and interaction. *Handbook of epidemiology*. Ed. by W Ahrens and I Pigeot. New York; Heidelberg; Dordrecht; London: Springer, 659–684.

Peduzzi, P, JConcato, Kemper, E, Holford, TR, and Feinstein, AR (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49, 1373–1379.

Pigeot, I and Ahrens, W (2008). Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. *Pharmacoepidemiology and Drug Safety* 17, 215–223.

Rassen, JA, Brookhart, MA, Glynn, RJ, Mittleman, MA, and Schneeweiss, S (2009a). Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of Clinical Epidemiology* 62, 1226–1232.

– (2009b). Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *Journal of Clinical Epidemiology* 62, 1233–1241.

Ray, WA (2003). Evaluating medication effects outside of clinical trials: new-user designs. *American Journal of Epidemiology* 158, 915–920.

Robins, J and Rotnitzky, A (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 91, 763–783.

Robins, J, Hernán, MA, and Brumback, B (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* 11, 550–560.

Rothman, KJ and Greenland, S (1998). Precision and validity in epidemiologic studies. *Modern Epidemiology*. Ed. by KJ Rothman and S Greenland. 2nd. Philadelphia: Lippincott-Raven, 115–134.

Schneeweiss, S (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety* 15, 291–303.

Schneeweiss, S, Setoguch, S, Brookhart, MA, Dormuth, C, and Wang, PS (2007). Risk of death associated with the use of conventional versus atypical

antipsychotic drugs among elderly patients. *Canadian Medical Association Journal* 176, 627–632.

Schneeweiss, S, Rassen, JA, Glynn, RJ, Avorn, J, Mogun, H, and Brookhart, MA (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 20, 512–522.

Schröder, C, Dörks, M, Kollhorst, B, Blenk, T, Dittmann, RW, Garbe, E, and Riedel, O (2017). Outpatient antipsychotic drug use in children and adolescents in Germany between 2004 and 2011. *European Child & Adolescent Psychiatry* (Epub 2016 Sep 13).

Seber, GAF and Wild, CJ (1989). *Nonlinear regression*. New York: John Wiley & Sons.

Stürmer, T, Schneeweiss, S, Avorn, J, and Glynn, RJ (2005). Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *American Journal of Epidemiology* 162, 279–289.

Suissa, S (2007). Immortal time bias in observational studies of drug effects. *Pharmacoepidemiology and Drug Safety* 16, 241–249.

Swanson, SA, Miller, M, Robins, JM, and Hernan, MA (2015). Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology* 24, 414–420.

Tchetgen Tchetgen, EJ, Walter, S, Vansteelandt, S, Martinussen, T, and Glymour, M (2015). Instrumental variable estimation in a survival context. *Epidemiology* 26, 402–410.

Terza, JV, Basu, A, and Rathouz, PJ (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27, 531–543.

Toh, S, Rodríguez, LA García, and Hernán, MA (2011). Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: An application using electronic medical record data. *Pharmacoepidemiology and Drug Safety* 20, 849–857.

Uddin, MJ, Groenwold, RHH, Ali, MS, de Boer, A, MAB, KCB Roes, Chowdhury, and Klungel, OH (2016). Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *International Journal of Clinical Pharmacy* 38, 714–723.

VanderWeele, TJ (2011). Causal mediation analysis with survival data. *Epidemiology* 22, 575–581.

Walker, AM (1996). Confounding by indication. *Epidemiology* 7, 335–336.

Wooldridge, JM (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd. Cambridge: MIT Press.

CURRICULUM VITAE

Bianca Kollhorst
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Department of Biometry and Data Manangement
Achterstr. 30, 28309 Bremen, Germany
Phone: +49 421 218-56980
Email: kollhorst@leibniz-bips.de

EDUCATION

2013-present	Ph.D. candidate in Statistics, University Bremen
2003-2008	Diploma in mathematics,
	Carl von Ossietzky University, Oldenburg

PROFESSIONAL POSITIONS

2010-present	Research associate at BIPS
2008-2010	Statistician at SCIderm GmbH (CRO) - Hamburg, Germany

PUBLICATIONS

Peer-reviewed publications

Jobski K, Schmedt N, Kollhorst B, Krappweis J, Schink T, Garbe E. Characteristics and drug use patterns of older antidepressant initiators in Germany. *European Journal of Clinical Pharmacology.* 2017;73(1):105-113.

Schröder C, Dörks M, Kollhorst B, Blenk T, Dittmann R, Garbe E, Riedel O. Outpatient antidepressant drug use in children and adolescents in Germany between 2004 and 2011. *Pharmacoepidemiology and Drug Safety.* 2017; (Epub 2016 Nov 20).

Schröder C, Dörks M, Kollhorst B, Blenk T, Dittmann R, Garbe E, Riedel O. Outpatient antipsychotic drug use in children and adolescents in Germany between 2004 and 2011. *European Child & Adolescent Psychiatry.* 2017; (Epub 2016 Sep 13).

Arfè A, Varas-Lorenzo C, Nicotra F, Scotti L, Zambon A, Kollhorst B, Schink T, Garbe E, Herings R, Straatman H, Schade R, Villa M, Lucchi S, Parienteia, Pariente A, Valkhoff VE, Romio S, Thiessard F, Schuemie M, Sturkenboom MCJM, Corrao G, on behalf of the SOS project consortium. Non-steroidal anti-inflammatory drugs and risk of heart failure in four European countries: Nested case-control study. *British Medical Journal.* 2016;354:i4857.

Douros A, Jobski K, Kollhorst B, Schink T, Garbe E. Risk of venous thromboembolism in cancer patients treated with epoetins or blood transfusions. *British Journal of Clinical Pharmacology.* 2016;82(3):839-848.

Enders D, Kollhorst B, Engel S, Linder R, Verheyen F, Pigeot I. Comparative risk for cardiovascular diseases of dipeptidyl peptidase-4 inhibitors vs. sulfonylureas in combination with metformin: Results of a two-phase study. *Journal of Diabetes and its Complications.* 2016;30(7):1339-1346.

Kollhorst B, Abrahamowicz M, Pigeot I. The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs. *Journal of Clinical Epidemiology.* 2016;69:96-106.

Schmedt N, Jobski K, Kollhorst B, Krappweis J, Schink T, Rüther E, Garbe E. Treatment patterns and characteristics of older antipsychotic users in Germany. *International Clinical Psychopharmacology.* 2016;31(3):159-169.

Schmedt N, Kollhorst B, Enders D, Jobski K, Krappweis J, Garbe E, Schink T. Comparative risk of death in older adults treated with antipsychotics: A population-based cohort study. *European Neuropsychopharmacology.* 2016;26(9):1390-1400.

Jobski K, Kollhorst B, Schink T, Garbe E. The risk of opioid intoxications or related events and the effect of alcohol-related disorders: A retrospective cohort study in German patients treated with high-potency opioid analgesics. *Drug Safety*. 2015;38(9):811-822.

Kollhorst B, Behr S, Enders D, Dippel F-W, Theobald K, Garbe E. Comparison of basal insulin therapies regarding the risk of acute myocardial infarction in type 2 diabetic patients: An observational cohort study. *Diabetes, Obesity & Metabolism*. 2015;17(12):1158-1165.