Real-Time Monitoring of Healthcare Interventions in Routine Care

Effectiveness and Safety of Newly Introduced Medicines

THOMAS CARS
Before market authorization of new medicines, their efficacy and safety are evaluated using randomized controlled trials. While there is no doubt about the scientific value of randomized trials, they are usually conducted in selected populations with questionable generalizability to routine care.

In the digital data revolution era, with healthcare data growing at an unprecedented rate, drug monitoring in routine care is still highly under-utilized. Although many countries have access to data on prescription drugs at the individual level in ambulatory care, such data are often missing for hospitals. This is a growing problem considering the clear trend towards more new and expensive drugs administered in the hospital setting. The aim of this thesis was therefore to develop methods for extracting data on drug use from a hospital-based electronic health record system and further to build and evaluate models for real-time monitoring of effectiveness and safety of new drugs in routine care using data from electronic health records and regional and national health care registers.

Using the developed techniques, we were able to demonstrate drug use and health service utilization for inflammatory bowel disease and to evaluate the comparative effectiveness and safety of antiarrhythmic drugs.

With a rapidly evolving drug development, it is important to optimize the evaluation of effectiveness, safety and health economic value of new medicines in routine care. We believe that the models described in this thesis could contribute to fulfil this need.

**Keywords:** Electronic Health Records, Comparative Effectiveness Research, Comparative Safety Research, Sequential drug monitoring, propensity score, real-world data, infliximab, TNF-inhibitors, dronedarone, amiodarone, flecainide

**Thomas Cars, Department of Medical Sciences, Cardiovascular epidemiology, Akademiska sjukhuset, Uppsala University, SE-75185 Uppsala, Sweden.**

© Thomas Cars 2016

ISSN 1651-6206
ISBN 978-91-554-9707-1
urn:nbn:se:uu:diva-304324 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-304324)
To Fia, Emil, Hjalmar and Wille
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
Contents

Introduction .......................................................................................................................... 11
 Monitoring and data on drug use ....................................................................................... 12
   History of drug monitoring in Sweden ........................................................................... 12
   Electronic Health Records ............................................................................................... 16
 Evidence for efficacy and safety ....................................................................................... 18
   Generalizability of randomized controlled trials ......................................................... 18
   Adaptive licensing ........................................................................................................... 19
 Comparative effectiveness research ................................................................................... 19
 Observational studies ......................................................................................................... 19
   Confounding .................................................................................................................... 20
   Propensity score ............................................................................................................. 21
 Unique opportunity for comparative effectiveness studies in Sweden ............................... 25
 Hypotheses and aims .......................................................................................................... 26
 Methods ............................................................................................................................. 27
   Data sources .................................................................................................................... 27
     Electronic Health Records ........................................................................................... 27
     Patient Registers ........................................................................................................ 29
     Registers at Statistics Sweden ................................................................................. 30
 Evaluating the potential of Electronic Health Records to monitor drug use in hospitals (paper I) ................................................................................................................ 30
 Analysis of drugs used in both outpatient and hospital settings (paper II) ......................... 32
 Developing a model for real-time monitoring of drugs (papers III and IV) ......................... 33
   Sequential recruitment and monitoring ......................................................................... 34
   Estimation of propensity score ....................................................................................... 35
   Study population ............................................................................................................ 35
   Outcome definition ........................................................................................................ 36
 Statistical analysis ............................................................................................................ 36
   Statistical analyses - studies III and IV ........................................................................ 37
 Ethical considerations ........................................................................................................ 39
Results and discussion ........................................................................................................40
  Paper I ..........................................................................................................................40
  Paper II .........................................................................................................................43
  Paper III .........................................................................................................................45
  Paper IV ........................................................................................................................48

General discussion and conclusions ..............................................................................53
  Real-time monitoring ....................................................................................................53
  Generalizability .............................................................................................................53
  Data completeness ..........................................................................................................54
  Residual imbalance and hidden bias ..............................................................................55
  Other models for real-time monitoring .........................................................................56
  Toward real-time monitoring .......................................................................................56

Future perspectives ........................................................................................................57

Summary in Swedish (sammanfattning på svenska) .........................................................58

Acknowledgements .........................................................................................................61

References .......................................................................................................................63
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT</td>
<td>Alanine Aminotransferase</td>
</tr>
<tr>
<td>ATC</td>
<td>Anatomical Therapeutic Chemical Classification System</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s Disease</td>
</tr>
<tr>
<td>CER</td>
<td>Comparative Effectiveness Research</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>DDD</td>
<td>Defined Daily Dose</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated Glomerular Filtration Rate</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic Health Record</td>
</tr>
<tr>
<td>EMA</td>
<td>European Medicines Agency</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard Ratio</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory Bowel Disease</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>ITT</td>
<td>Intention to Treat</td>
</tr>
<tr>
<td>KarDa</td>
<td>Data Warehouse at Karolinska University Hospital</td>
</tr>
<tr>
<td>LISA</td>
<td>Longitudinal Integration Database for Health Insurance and Labour Market Studies</td>
</tr>
<tr>
<td>NAHIT</td>
<td>National Alliance for Health Information Technology</td>
</tr>
<tr>
<td>NPR</td>
<td>National Patient Register</td>
</tr>
<tr>
<td>OTC</td>
<td>Over the Counter</td>
</tr>
<tr>
<td>SPDR</td>
<td>Swedish Prescribed Drug Register</td>
</tr>
<tr>
<td>PIN</td>
<td>Personal Identity Number</td>
</tr>
<tr>
<td>PS</td>
<td>Propensity Score</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized Controlled Trial</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor Necrosis Factor</td>
</tr>
<tr>
<td>UC</td>
<td>Ulcerative Colitis</td>
</tr>
<tr>
<td>ULN</td>
<td>Upper Limit of Normal</td>
</tr>
<tr>
<td>VAL</td>
<td>Stockholm Regional Healthcare Data Warehouse</td>
</tr>
<tr>
<td>WCSD</td>
<td>Weighted Conditional Standardized Difference</td>
</tr>
</tbody>
</table>
Introduction

Drug therapy is the most frequent intervention in all health care systems[1] and the use of medicines is forecast to increase by 24% from 2015 to 2020 world-wide[2]. With a rapidly evolving drug development scene on the brink of the genomic medicine era, we will see many new specialized drugs introduced into clinical practice. In 2018, biological therapies (i.e. drugs derived from living material) are expected to account for more than 50% of top-selling drugs[3]. These drugs will likely be an effective treatment option in terms of improved quality of life and reduced duration of illness. However, they may also have severe side-effects and are more expensive compared to other drugs, which will cause major economic stress for the health sector.

Before marketing new drugs, their efficacy and safety are evaluated using randomized controlled trials (RCTs). Although there is no doubt about the scientific value of RCTs, they are usually conducted in selected populations with questionable generalizability to routine care. Regulators and payers in healthcare have therefore begun to demand monitoring of the use, effectiveness, safety and economic value of medicines in real world settings[4-7].

In the digital data revolution, when healthcare data is growing at an unprecedented rate[8], drug monitoring in routine care is still highly under-dimensioned. Although many countries have data on prescription drugs at the individual level in ambulatory care[9, 10], such data is often missing for hospitals[9, 11]. This is a growing problem, since more new and expensive drugs will be introduced in hospital care.

Electronic health records (EHRs) could provide an opportunity to monitor hospital drugs. However, systems for structured data capture and functions to ensure the accuracy and completeness of collected EHR data have been lacking thus far[12] and may be one reason why information from EHRs is still underutilized[13].

Since all counties in Sweden had implemented EHR systems by 2012[14], we hypothesized that data on drugs used in the hospital setting in the EHR systems could be used to monitor drugs administered in hospitals. We also hypothesized that, accompanied with other patient data, such data could be used
for real-time monitoring of the effectiveness, safety and health-economic value of newly marketed drugs in routine care.

Monitoring and data on drug use

Monitoring of drug use can be traced back to the United States in the 1930s[15]. The pharmaceutical industry expressed the need for data on drug use to monitor their products and to identify areas for future drug development. This paved the way for the establishment of large databases to track drug prescriptions and sales[16, 17]. From a global perspective, pharmacoepidemiology, which is defined as the study of the use and effects of drugs in large populations, can be dated to the early 1960s[18]. As a result of the thalidomide disaster[19], there were growing concerns about adverse drug reactions and governments realized that data on drug use was needed. The initial focus of pharmacoepidemiology was to evaluate the safety of drugs, but this area has shifted to also include studies of effectiveness, health-economic aspects and rational use of drugs[20]. In the past two decades, pharmacoepidemiology has grown rapidly as computer-based data sources have become available and with the increased development of epidemiologic techniques[17, 21, 22]. Since the volume of available digital healthcare data is currently growing at an increasing speed, pharmacoepidemiological studies on effectiveness, safety and cost-effectiveness will become even more powerful and valuable in the near future.

History of drug monitoring in Sweden

Already in the 1960s, there were ongoing discussions of implementing an individual-level register of dispensed prescription drugs in Sweden[23]. The Swedish board for adverse drug reactions realized that this would facilitate the assessments of adverse drug reaction reports and tried to establish an individual-level drug register, but unfortunately found it to be too complicated at that time.

Despite the lack of individual-based drug data, Sweden has a long history of drug monitoring using aggregated national pharmacy sales data. Since 1977, we have been able to monitor national drug sales using data from the National Corporation of Swedish Pharmacies (Apoteket AB, formerly Apoteksbolaget AB)[24]. Since 1985, it has been possible to separate drug sales to hospitals and drug sales in the outpatient setting[24, 25]. On January 1, 1997, a new Swedish reform regarding drugs was implemented with the general aim of increasing the quality of drug prescriptions and counteracting the large rise in drug costs of the previous years[26]. As a consequence of this reform, the Prescription Registration Act was implemented[27]. It was now mandatory to
include the patient’s personal identity number (PIN) [28] on all drug prescriptions. This was due to the reform transferring cost liability from the government to the county councils (official autonomous health care administrative regions), and the PIN was used to identify the patient’s place of residence for billing purposes. The PIN also made it possible to analyze age and sex distributions for drug therapies. However, according to the Prescription Registration Act, the PINs were not to be stored in such a way that enabled links to other registers. In October 2002, it became mandatory to include a unique code for each prescribing healthcare provider[29]. This allowed for drug monitoring on a care unit level.

In the absence of a national, individual-level drug register during these 40 years, two internationally well-known regional registers of prescribed drugs were established: the County of Jämtland Project and the Tierp Study. Initiated in 1970, the County of Jämtland Project involved the selection of one seventh of the population of Jämtland; participants’ drug prescriptions were continuously monitored[30, 31]. This project has been used for longitudinal studies of drug use, and data from the study were also used for data simulations for the development and implementation of the new Swedish reimbursement system for drugs. In 1972, Uppsala University Centre for Primary Care Research initiated the Tierp Study, in which data on health service and drug prescriptions were collected for about 20,000 individuals in the area of Tierp[32]. Among other topics, this project has been used to study antidepressants and antidiabetics drug use.

About 40 years after the initial discussion regarding a national, individual-based register on prescription drugs, the Swedish Prescribed Drug Register (SPDR) was established in July 2005[33]. This milestone in Swedish drug monitoring enabled pharmacoepidemiological studies on an individual level for ambulatory care, with record linkage possibilities to other registers and databases. From July 2010, each region had access to data on dispensed prescription drugs for the residents and caregivers in their county.

Data on drug use in hospitals - the missing link
As described above, we have seen gradually improved possibilities for monitoring drugs used in ambulatory care. Unfortunately, we have not seen the same improvements in the monitoring of drugs administered in hospital settings. In Sweden, drugs used in hospitals are usually delivered from the pharmacy and stored in cabinets in the ward/department. These drugs are intended for use for any patient, i.e., they are not ordered for a specific patient. Basically, possibilities for access to national statistics on drugs administered in hospitals has not improved since 1985 and drug use can only be followed in terms of costs and volumes. Other data sources for hospital drugs in Sweden exist, for example in national quality registers. Sweden has over 100 healthcare quality registers aiming to collect individual-based data to improve
the quality of patient care[34]. These registers have significantly improved healthcare and an advantage of these registers, compared to other data sources for observational research, is that they often include patient-reported data[35]. However, quality registers are usually disease-specific and are therefore not suitable for evaluating specific drugs, especially if they are used for several indications. Most of these registers also have limited coverage due to voluntary participation, which may raise questions about their generalizability.

A lack of individual-level hospital data on drug use is common in many countries. In a review aiming to identify hospital-based databases used in observational studies, only 12 databases were identified in Europe, the US and Asia[11]. A more recent project aiming to outline nationwide administrative databases for drug monitoring in Europe identified 18 countries with data on drug use in hospitals/inpatients[9]. However, only one country had data including patient identifiers enabling links to other registers. The corresponding figure for outpatient databases was 27 countries, 10 of which had record linkage possibilities. (Figure 1).

**Sources of nationwide drug utilization databases**

![Sources of nationwide drug utilization databases](image_url)

Figure 1. Sources of nationwide outpatient and inpatient databases on drug consumption in Europe. Countries filled with green or orange have national data on outpatient and inpatient drug utilization respectively. Countries with a lined pattern have drug utilization data with record linkage possibilities. These maps are adapted from the PROTECT-project (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European ConsorTium) and this map is constantly changing [9, 36].
There are many likely explanations for the absence of individual-level drug data in hospitals. Prescriptions in hospitals are generally more complex and hold more information than outpatient prescriptions. For example, hospital prescriptions often include detailed information about the administration and they may also include more than one substance (solutions). There are also several EHR systems on the market and standards and infrastructure for EHRs have been lacking[37, 38] and some prescriptions in hospitals are still not computerized.

Lack of individual-level data for hospital-based drugs is a large and increasing problem, since we are observing a trend toward more hospital-administered drugs. In the past 15 years, we have seen an increase in hospital drug use both in terms of cost and volume (Figure 2).

![Figure 2. Trend in hospital drug expenditure and volumes in Sweden measured as the total cost (million EUR) and the total amount of sold drug packages to the hospitals. Data provided from the national system for drug statistics (Concise) [39].](image-url)
This trend is expected to continue as 42% of the substances in drug development are biologics, compared to 8% biologics on the market today (Figure 3) [40]. Most of these are likely to be introduced in the hospital setting.

![Figure 3. Proportion of biologicals of currently marketed drugs compared to substances in pharmaceutical companies’ pipeline (companies represented in the NASDAQ Biotech Index)[40].](image)

**Ongoing work to develop a national register on hospital-administered drugs**

To provide a complete overview of drug treatment in the Swedish population, data on outpatient prescription drugs needs to be complemented with drugs administered to patients in the hospital setting. Therefore, one action in the Swedish national pharmaceutical strategy is to “make follow-up of hospital drugs possible at an individual level,” i.e., to develop a national register of hospital-administered drugs[41]. This work, initiated in 2011, has resulted in 4 published reports evaluating the technical and legal aspects of such a register[42-45]. This work has also included practical data extractions. The progress is slow however, and if the goal is a register with complete overview of drugs in near time, the relevant authorities will have to intensify these efforts.

**Electronic Health Records**

According to the definition developed by the US National Alliance for Health Information Technology (NAHIT), EHRs are defined as “an electronic record of health-related information on an individual that conforms to nationally recognized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff across more than one healthcare organization” [46]. EHRs are recorded as part of patient care and typically include information on the patient’s demographics, contact information, vital
signs, medical history, current and past medications, x-ray reports and laboratory data.

Because patient records are a central part of the healthcare process, since the 1990s, Sweden’s county councils have started to implement electronic support for this process. However, implementing EHRs was easier for primary care than in hospitals and it was especially hard to implement the EHRs’ prescribing modules for drug use in inpatient care. However, in 2012, all county councils had implemented EHRs in all healthcare areas (Figure 4, left panel). Most Swedish county councils have also chosen to implement one single EHR system in their region (Figure 4, right panel)[14, 47].

Electronic health records with integrated prescribing modules provide an opportunity to include hospital drugs in pharmacoepidemiological studies[48]. An advantage of using drug information from EHRs is that they may include information on the administered dose and therefore enable a precise estimate of drug exposure.

However, systems for structured data capture and functions to ensure the accuracy and completeness of collected EHR data have been lacking and may be one reason why information from EHRs continues to be underutilized[12]. Other reasons include access to information, data reliability, confidentiality and data security[49].

Figure 4. Counties marked in green represents regions in Sweden that have implemented EHRs (left panel). The right panel points out counties that have chosen to implement the same EHR-system for the whole region. Adapted from Lars Jerlvall and Thomas Pehrsson (eHälsa i landstingen)[14, 47].
Evidence for efficacy and safety

The general consensus is that causal inference should be based on randomized studies[50]. Due to the randomization process, treatment and controls are expected to be similar both in terms of observed and unobserved characteristics. The goal is to balance the treatment groups, and ideally, the only differences should be the treatment assignment. If the randomization process is successful and the outcome differs significantly between patient groups, one can conclude that the observed difference is caused by the treatment (causal inference).

The counterfactual model has become the standard for evaluating causality in epidemiology and medical studies[51]. A counterfactual is a potential outcome defined as what would have happened to the same subject in the absence of treatment. For example, in an experimental drug study, the same subject cannot be assigned to both treatment (t) and control (c) (unless it is a crossover study). When a subject is assigned to treatment (t), the outcome (Y) can only occur under one condition in this subject, Y(t). The treatment that this subject does not receive is called the counterfactual treatment (c). For this subject, also a counterfactual outcome, Y(c) exists. However, the counterfactual outcome, Y(c), for this subject is by definition impossible to observe.

Randomized controlled trials work by creating a group that can mimic it. Results from RCTs are usually presented as the difference in outcome between treatment and control (average causal effect: Y(t) – Y(c)). If the average causal effect favors treatment, the population as a whole will perform better under treatment; however, it does not guarantee that a specific subject will benefit from the treatment.

Generalizability of randomized controlled trials

Pre-market authorization evaluations of drugs are usually designed to provide necessary evidence of efficacy and safety to bring a drug to the marketplace. Although the double-blind RCT is the most reliable design for causal inference, it is usually conducted in selected populations often different from patients treated with the drug in clinical practice, where drugs are often used for longer time periods and by elderly populations with more comorbidities and multiple concomitant drugs[52, 53]. The results from RCTs may therefore have limited generalizability to clinical practice. Pre-market trials are also often limited in sample size and are generally powered to detect the most common adverse events[54, 55]. Consequently, rare or late adverse effects are not expected to be identified in clinical trials.
Adaptive licensing

Bringing a new drug molecule to the marketplace can take up to 10 or 15 years. Therefore, to improve timely access for patients to new medicines, the EMA has launched “adaptive licensing.” Adaptive licensing is defined by Eichler et al. as "a prospectively planned, flexible approach to regulation of drugs and biologics. Through iterative phases of evidence gathering to reduce uncertainties followed by regulatory evaluation and license adaptation, adaptive licensing seeks to maximize the positive impact of new drugs on public health by balancing timely access for patients with the need to assess and to provide adequate evolving information on benefits and harms so that better informed patient care decisions can be made”[56, 57]. In this initiative, patients are granted earlier access to medicines at the cost of less randomized evidence.

Comparative effectiveness research

Comparative effectiveness research (CER) is the direct comparison of existing healthcare interventions, aiming to produce evidence regarding the effectiveness and safety of medical products[58]. A drug’s efficacy is defined as “the extent to which a specific health intervention produces a beneficial result under ideal conditions,” (i.e. a response to the question: “Can it work?”) whereas effectiveness is defined as “the extent to which a specific health intervention produced a beneficial result when deployed in the field under routine conditions” (i.e. a response to the question: “Does it work in practice?”)[59]

The ultimate goal of comparative effectiveness research is to improve health by developing and disseminating evidence to patients, healthcare professionals and policy-makers regarding the effectiveness of specific interventions. There are a number of methods for conducting CER, such as systematic reviews, meta-analyses, experimental studies and non-experimental studies, including retrospective and prospective observational studies[60].

Observational studies

Comparative effectiveness studies using observational data can have a great scientific impact when RCTs are not feasible or when RCTs have not included the population or outcome of interest. However, assessing causality using these study designs is challenging. In observational studies, we don’t have the randomization mechanism and therefore many factors may influence treatment assignment. Consequently, observational studies face many challenges and it is much harder to draw causal conclusions from these studies. However,
recent years have seen advancements in methods for causal inference in observational studies and some of the challenges might be overcome using various methodological approaches[61-63].

Confounding
Confounding is a situation in which the causal association between exposure and outcome is distorted by a third variable[64]. Confounding is characterized by the lack of comparability between study groups and therefore a threat to the internal validity of the study. Minimizing confounding bias is a central task for researchers conducting observational studies. Confounders can be divided into measured and unmeasured confounders. Measured confounders can be addressed by using appropriate statistical methods, while unmeasured confounding (a form of hidden bias) is caused by unobserved characteristics and therefore more difficult to handle[65].

A special type of confounding that may occur in pharmacoepidemiological studies is confounding by indication[66] (also referred to as channeling). Confounding by indication is a term used when conditions that determine the selections of drugs are also linked to the outcome. A commonly used example for confounding by indication is calcium channel blockers and risk of myocardial infarction[67]. When used for hypertension, calcium channel blockers showed a higher risk of myocardial infarction compared to beta-blockers and thiazide diuretics. However, this finding may have been a result of confounding by indication since patients treated with calcium channel blockers were more likely to have peripheral vascular disease and diabetes mellitus (conditions motivating the choice of a calcium channel blocker over a beta-blocker or a thiazide diuretic). Since prescribing drugs is a dynamic process, characteristics for treatment may change during a drug’s lifecycle. When this happens, i.e. when confounding by indication changes over time, calendar time might be a confounder or a proxy for other confounders[68, 69]. This phenomenon needs to be addressed when specifying the causal model.

Frequently used unsystematic methods for handling confounding can lead to bias by leaving out important confounders, but also by adjusting for variables that are not confounders, which can potentially worsen the bias situation. Causal diagrams, also known as directed acyclic graphs (DAGs), comprise one tool to help select a bias-minimized model in epidemiological studies[70, 71]. DAGs are used to visualize the model for the hypothesized causal relationship, and to help identify variables for which to adjust.
Propensity score

An increasingly popular strategy for maximizing the causal inference potential in observational studies is the propensity score (PS), a method developed by Paul Rosenbaum and Donald Rubin in 1983[72]. The PS is defined as the conditional probability of receiving treatment given selected covariates with the goal of balancing covariates between study groups (treatment and control)[73]. Rosenbaum and Rubin showed that if treatment assignment is strongly ignorable (i.e. no unmeasured confounding) given the observed covariates, an unbiased estimate of the average treatment effect could be estimated adjusting for the PS alone, rather than a vector of confounders, which is often of high dimension. The ignorable treatment assignment assumption means that no systematic, unobserved and pretreatment differences related to the outcome exist between treated subjects and controls[73].

The estimated PS for a subject can be denoted by Pr(z|x), where z is the treatment (0/1) and x is observed covariates. Since the PS is a probability, it ranges from 0 to 1. If two subjects have the same PS, they both have the same chance of receiving treatment. However, these two subjects may differ substantially in observed covariates.

In a randomized clinical trial where subjects are assigned to treatment or control by chance in a 1:1 ratio, each subject will have a PS of 0.5 since each subject will be randomly assigned to either treatment with a 50% probability.

Methods for PS adjustments

Four primary methods for adjusting on the PS are used:

1. stratification,
2. matching,
3. covariate adjustment and
4. weighting.

The most common methods are matching and covariate adjusting using the PS[74, 75].

Estimating the propensity score

Several methods for estimating the PS have been proposed, but logistic regression is most commonly used[76]. Despite the large and increasing use of propensity scores, there is no general consensus regarding procedures to determine which covariates are to be included in the estimation of PS. However, a common advice is to only include true confounders and “potential confounders” (covariates that affect the outcome but not necessarily the exposure) in
the PS estimation[77, 78]. It is also advised to tend toward over-including co-variables to avoid leaving out important confounding variables[77]. One may also include interactions and polynomial effects, such as cubic splines, in the PS estimation[79-81].

**Overlap in the propensity score**

After estimation of the PS, plotting the distributions in PS between the study groups is considered standard practice (Figure 5). The area of overlap in the PS distribution presents individuals with a comparable PS and is defined as the “region of common support”. Since it is usually difficult to achieve covariate balance for individuals outside the region of common support, these subjects are often excluded from the analyses[79].

![Figure 5. Region of common support for PS between drug A and drug B.](image)

**Assess balance in observed covariates**

Since the goal of PS is to achieve covariate balance between treated and control groups, assessing balance is a fundamental step in the PS process. Several methods to assess balance have been proposed. One frequently used method is to estimate standardized differences (Figure 6), which are defined as the difference in means between treated subjects and controls divided by an estimate of the standard deviation in treated subjects and controls[82]. Standardized difference can be calculated for both continuous and dichotomous covariates and represents the number of standard deviations by which the two groups
Compared to t-tests and chi2-tests, which are also common when evaluating covariate balance, the standardized difference is not influenced by the sample size and balance can be evaluated on a common scale\[80\]. Standardized differences are easy to apply in the context of stratification and matching on the PS.

\[
\begin{align*}
\text{标准化差异} &= \frac{|E(T) - E(C)|}{\text{方差} + \text{方差}} \times 100% \\
\end{align*}
\]

For covariate adjustment on the PS, which is one of the most frequently used PS model in the literature, evaluation of covariate balance is rarely seen in the studies. This is likely explained by the fact that balance diagnostics in this context were not developed until 2008\[82\]. In 2008, Austin proposed the weighted conditional standardized difference (WCSD), which is a weighted modification of standardized differences. In the WCSD, each covariate is regressed on three variables: (1) treatment status \((T)\), (2) the estimated PS and (3) an interaction between treatment status and PS (to allow the mean difference between treated subjects and controls to be different for different values of the PS). In the regression model \((T + PS + T \times PS)\), for each value of the PS, the mean response for that covariate is estimated assuming first that all subjects were treated and further that they were controls. Thus, using the fitted regression model, two predicted values are calculated for each subject, one assuming that the subject was treated and one assuming that the subject was untreated (control). The absolute standardized difference for mean response in treated and untreated groups is calculated for each study subject and averaged over the distribution of the PS in the study sample.

Figure 6. Example figure of standardized differences in baseline covariates between treated subjects and controls, before and after matching on the PS. (A standard difference <0.10 indicates negligible imbalances)\[82\].
There is no general consensus on threshold values for the standardized difference or WCSD to indicate covariate imbalance. Some suggest that standardized difference should be less than 0.25[84], while others suggest that a standardized difference less than 0.1 denotes a negligible imbalance[82, 85-87].

**Sensitivity analysis**

The legitimacy of causal inference in observational studies is based on the assumption that no unmeasured confounding exists. However, this is a very strong assumption and therefore PS analysis should be accompanied by sensitivity analysis to investigate how the finding of the study may be affected by the presence of unmeasured confounding. The first sensitivity analysis of unmeasured confounding in observational studies can be traced back to the 1950s, when Cornfield et al. quantified the role of unmeasured confounding in a study of smoking and lung cancer[88]. Since Cornfield, several methods have been proposed for sensitivity analysis of how robust the estimate is to the presence of unmeasured confounding[65, 89-91].

One commonly used approach that is possible to apply to a variety of statistical models was proposed by Lin et al.[92]. This approach involves evaluating how powerful an unmeasured confounder would have to be to change the observed results.

**Why the propensity score?**

One might ask: why use PS instead of performing a conventional regression in which all covariates are entered into the model? According to Rubin[93], both methods should lead to the same conclusion. In a review by Shah et al. comparing studies evaluated by both conventional regression and propensity score, there was disagreement in only 10% of the cases[94]. Another review by Stürmer et al. found that only 13% of the estimates from PS differed by more than 20% from estimates from a regression model[95].

However, PS offers some advantages over conventional regression. Since the goal of PS is to balance covariates, one is not concerned about over-parameterizing the model, allowing for a more complicated model that includes interaction and polynomial effects[76, 80]. Goodness-of-fit diagnostics can be performed more easily in PS models than in complex multivariate regression[96]. Adjusting for PS is also an effective way to control for covariates when there are few observations or events[97]. This is because only one vector, the propensity score, is controlled for, which minimizes the issue of high dimensionality. Simulation studies have also suggested that PS models are more robust than conventional regression if the models are misspecified[98]. Since PS modeling includes covariate balance checks, one is also alert in situations in which covariates do not overlap between treated subjects and controls, which may be more difficult to observe in conventional regressions.
Personally, I think that the biggest advantage of PS is that you are forced to carefully think through your model, for example: “Who is exposed?”, “What are the grounds for exposure?” and “Have the characteristics of exposure changed over time?” Another major advantage arises from the fact that estimating the PS is an iterative process in which you re-specify the PS model until sufficient balance is achieved.

Unique opportunity for comparative effectiveness studies in Sweden

Sweden has unique opportunities for observational research. The advantages lie in the country’s civic registration system[28] and the fact that all residents have universal access to healthcare with a negligible co-payment for healthcare visits, hospitalizations and drugs[99]. The variety of nationwide health registers, quality registers, registers on socio-demographics and socio-economics and complete implementation of EHR systems allows for research on large populations with minimal loss of follow-up.
Hypotheses and aims

We hypothesized that the data on drugs used in hospital settings in EHR systems can be used to monitor drugs administered in hospitals. Furthermore, we hypothesized that these data can be used to build a generic model for real-time monitoring of effectiveness and safety of newly marketed drugs.

Specifically, we aimed to:

- extract data on drug use from a hospital-based EHR system and evaluate how these data can be used to monitor drugs administered in the hospital setting (paper I);
- use the extraction models developed in paper I to analyze drug utilization patterns and healthcare consumption in inflammatory bowel disease (IBD) (paper II);
- build and evaluate a generic model for real-time monitoring of the effectiveness of newly marketed drugs (paper III), and
- expand and modify the generic model developed in paper III to also monitor the safety of new drugs (paper IV).
Methods

All studies included in this thesis are based on data for Stockholm county, with its 2.2 million inhabitants accounting for 23% of the Swedish population[100].

Data sources

Electronic Health Records

**TakeCare and TakeCare Intelligence**

TakeCare (CompuGroup Medical, Stockholm, Sweden) is an EHR system developed in the early 1990s and is presently the most widely used EHR system in the region[101]. The EHR installation in the Stockholm Healthcare Region has over 50,000 active users representing several groups of healthcare professionals. The installation handles more than 4 million patient records and covers more than 88% of inpatient care and 75% of primary care[102].

The architecture of TakeCare is based on several modules using a shared platform (Figure 7). To provide a standardized interface for data analysis, the TakeCare installation is integrated with the relational database TakeCare Intelligence (MS SQL Server; Microsoft Corporation, Redmond, USA). The contents of the TakeCare Intelligence relational database are incrementally updated every 24 hours using a validation process of data, securing correct structure and data quality (Figure 8).
Figure 7. Architecture of the EHR-system TakeCare (modules) (Figure provided from CompuGroup Medical Sverige AB)[103].

Figure 8. Process for transferring data from the EHR-system to the relational SQL-database. (Figure provided from CompuGroup Medical Sverige AB)[103].
Karolinska University Hospital Database (KarDa)
To perform hospital-specific analyses, Karolinska University Hospital has its own relational database, KarDa (MS SQL Server; Microsoft Corporation, Redmond, USA), which includes data that have been generated and belong to the hospital. KarDa includes data from TakeCare as well as from other patient information systems, such as one for radiological information. Clinical data in TakeCare are automatically extracted from TakeCare Intelligence and loaded into KarDa when the daily update of TakeCare Intelligence is completed.

MediDoc
For study II, we also extracted and included data from the EHR system MediDoc, which was developed in the mid-1990s and has mainly been used in primary healthcare settings. MediDoc contains support for the usual tasks, such as appointment scheduling, documentation, billing, prescriptions and reports[104].

Patient Registers

Regional administrative healthcare data warehouse (VAL)
The regional healthcare data warehouse of Stockholm County Council (called VAL) includes information on all contacts with healthcare financed by the county council[105-107]. Data for primary care, secondary care and hospitalizations are available from the 1980s (inclusion of diagnoses from primary care from 2003). The International Classification of Diseases Version 10 (ICD-10)[108] has been used since 1997. VAL also contains demographic information on patient age, sex, migration status and death. Information on prescription drugs dispensed in the ambulatory setting have been included since July 2010. The drug dispensing data come from the same source as the Swedish Prescribed Drug Register, with a population coverage of over 99%[33]. The Anatomical Therapeutic Chemical (ATC) classification system is used to code dispensed drugs[109]. Drug use can also be identified in VAL using procedure codes and ATC codes in outpatient specialist and hospitalization data[110].

National Patient Register
In 1964, the Swedish National Board of Health and Welfare started to collect data on hospital discharge diagnoses in the National Patient Register (NPR)[111]. In 1984, participation became mandatory for all county councils, and in 1987, all inpatient care in Sweden was included. More than 99% of hospital stays are registered in the NPR and a recent review estimated the overall validity to be 85-95%[112]. Since 2001, NPR has also covered outpa-
tient visits from all Swedish hospitals and other specialist clinics in ambulatory care; however, information on primary care is not yet included. Data in the NPR include PIN, patient demographics, caregiver information, date of admission/discharge or visit, diagnosis (primary and secondary) and clinical procedures. Diagnoses are coded according to the current version of International Classification of Disease (ICD)[108]. When data are entered into the NPR, checks are performed to secure the quality of the data.

Prescribed Drug Register
The Swedish Prescribed Drug Register (SPDR) at the National Board of Health and Welfare holds information on dispensed drugs since 1999[111]. In July 2005, the SPDR was updated to also include the PIN, enabling the linking of data to other registers. The SPDR includes information on the dispensed unit including drug identification number, number of dispensed units, costs, patient demographics (including age and sex), date of prescription and dispensation and code for caregiver. Drugs are classified according to the Anatomical Therapeutic Chemical classification system. Population coverage in SPDR is estimated to be over 99%[33]. The register does not include data on over-the-counter (OTC) medications and coverage is not complete for vaccines and drugs used in nursing homes.

Registers at Statistics Sweden
The Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA) is a database held at Statistics Sweden[113]. It includes information on all Swedish citizens from 16 years and up; information is registered in Sweden on December 31, each year. It combines information from several sociodemographic registers and variables including: country of birth, education level, income and occupational status.

Evaluating the potential of Electronic Health Records to monitor drug use in hospitals (paper I)
A prerequisite for studies on drug use in hospitals is that data is recorded in the EHR systems and can be extracted and analyzed. Therefore, in this first seminal paper, we developed and evaluated two different models for data extraction of drug use from a hospital-based EHR system. We chose the tumour necrosis factor (TNF) inhibitor infliximab as a study drug example, since it is given as an intravenous infusion and is therefore a predominantly hospital-based drug[114]. Infliximab was also chosen for the possible challenge of extracting data on this drug, due to the fact that infliximab is a concentrate for a solution with sodium chloride; that is, the prescription holds two substances
(infliximab and sodium chloride). Therefore, documentation of infliximab in the EHR system is likely to be more complex than a regular prescription. The Drug and Therapeutics Committee[115] in the Stockholm County Council was also interested in analyzing the utilization pattern of infliximab, since infliximab accounted for a substantial proportion of total drug expenditure in the region.

The study included all patients with at least one administration of infliximab recorded in EHRs (TakeCare) at Karolinska University Hospital between January 1, 2007 and December 31, 2010. Data were extracted from the Karolinska University Hospital Database (KarDa), including data from TakeCare generated at Karolinska University Hospital. In this study, we extracted data from the two EHR modules Medication and Patient administration in TakeCare (Figure 7). A summary of the database tables and variables from which data have been extracted is presented in Figure 9.

![Figure 9. Database tables and variables from which data about infliximab drug use have been extracted.](image)

Furthermore, two different techniques were developed: (1) an automatic extraction procedure and (2) a semi-automatic procedure. In the semi-automatic procedure, four steps were added to complete, transform and correct the automatically generated data using pre-defined algorithms handling the following situations:

**Duplicate infusions:** In situations where two or more infusions of infliximab were recorded with the same dosage for the same patient for a day, it
was assumed to be an incorrect registration in the EHR system and only one infusion was included.

Unterminated infusions: If an infusion of infliximab had not been terminated in the EHR system, it was classified as ongoing and consequently not included as performed in the table administrations. Unterminated infusions were identified in the database table infusions. Unterminated infusions were mainly explained by incorrect registrations in the EHR system, and most of those infusions were probably administered. These unterminated infusions were therefore analyzed and if they were started at a time when the patient was scheduled for an infusion of infliximab, we assumed they were executed and included them.

Identification of diagnoses: If data on the primary diagnosis for a treatment episode/visit were missing, we used different algorithms to search for a credible diagnosis among the patient’s previous treatment episodes.

Transformations of dosages: For calculation of dosages, the prescribed dosage was used. To obtain uniform calculable data, dosages were transformed to a uniform format.

Extracted data included: encrypted PIN, age, sex, time of prescription and administration, healthcare unit responsible for the administration, prescribed and administered dose, time of admission and discharge. The diagnosis for the treatment episode/visit at which infliximab was administered was also extracted using algorithms that match the drug administration to the corresponding treatment episode/visit.

Analysis of drugs used in both outpatient and hospital settings (paper II)

Since there is a general lack of individual-based data on drug use in hospitals, many pharmacoepidemiological studies only include information on pharmacy-dispensed drugs for outpatient care. In study II, we therefore used the data extraction techniques developed and evaluated in study I to analyze drug treatment in a therapeutic area where drug treatment includes both prescription drugs and drugs administered in the hospital setting. Since TNF inhibitors are an important component of treatment in inflammatory bowel disease (IBD) and are used as both prescription drugs and in hospitals, we chose to analyze drug use in IBD.
Study II was designed as a cross-sectional study in which both drug therapy and health service utilization were analyzed among patients with IBD in Stockholm county. The main data source was VAL, which includes information on healthcare use and dispensed prescription drugs. VAL may also include information on selected hospital-administered drugs[110]. However, this information has questionable coverage. Therefore, to achieve a complete overview of drug use in IBD, data on TNF inhibitors were also extracted from the EHR system TakeCare using the relational database TakeCare Intelligence. Data on TNF inhibitors were included from the three major hospitals in Stockholm involved in the care of IBD: Karolinska University Hospital, Södersjukhuset AB and Danderyds sjukhus AB. Within the study period, these three hospitals accounted for 94% of all TNF inhibitors in Stockholm county (measured as defined daily doses [DDD]) [109].

VAL had incomplete information on diagnoses for one large private gastroenterology specialist clinic (Stockholm Gastro Centre). To address this limitation, information on all patients with IBD was therefore extracted directly from the clinic’s EHR system (MediDoc, CompuGroup Medical, Sweden). Data from TakeCare and MediDoc were linked to VAL data using the PIN.

The study population comprised all individuals with a diagnosis of IBD recorded either in primary or in secondary care from January 1, 1997 to December 31, 2012. The following ICD-10 codes were used to select IBD patients: K50, Crohn’s disease; K51, Ulcerative colitis and K523, Indeterminate colitis. We restricted our study population to only those patients who have had at least two diagnoses recorded by a physician on two separate occasions[116]. For this population, information on inpatient and outpatient health service use, IBD-related surgery and IBD-related drug treatment were analyzed.

Developing a model for real-time monitoring of drugs (papers III and IV)

In studies I and II, we developed and explored the possibilities of using data from EHR to study drug use in hospitals. However, these two studies only used a small fraction (drug use, diagnoses and demographics) of all the data stored in the EHR databases. The next aim of this thesis was to build and evaluate a generic model of real-time monitoring of new drugs using data from EHRs. To achieve this objective, we continued by extracting data from other parts of the EHR system (clinical measurements, clinical procedures and laboratory information).
We extracted data from the EHR system TakeCare using the relational database TakeCare Intelligence. Data from Karolinska University Hospital, Södersjukhuset AB and Danderyds sjukhus AB were used. In studies I and II, we only included EHR data from hospitals, but since TakeCare is also the principal EHR system for primary care in Stockholm county, we extracted and included data from primary healthcare units as well. Data extracted included patient demographics, information on prescribed and administered drugs, information on diagnoses and clinical procedures and information on clinical measurements and laboratory tests.

For a complete overview of drug exposures, covariates and outcomes for this real-time drug monitoring model, we linked data from EHR with VAL, SPDR, NPR and Statistics Sweden (LISA).

For developing and evaluating this model, we chose the recently developed antiarrhythmic agent dronedarone as a study drug example. Dronedarone is an antiarrhythmic agent approved by the FDA in 2009[117] and marketed in Sweden in 2010[118]. It was approved to help maintain normal sinus rhythm in patients with a history of atrial fibrillation. The developed model was based on prospectively collected data, but to emulate real-time monitoring the model was built as if dronedarone were marketed today.

Sequential recruitment and monitoring
We built a generic sequential cohort model for real-time head-to-head (treated subjects and controls) comparisons of new drugs. The model was sequentially updated in order to evaluate data as they are collected. To emulate prospective sequential analysis, we chose to update the model every six months, but the model can be updated at any desired frequency. In each six-month period (recruitment cycle), new users of the new treatment and controls were included and added to the cohort to continuously increase the study population size (Figure 10). At the end of each recruitment cycle, a propensity score (PS) was estimated for all patients. Furthermore, all patients were monitored for study outcomes, which were evaluated at the end of each recruitment cycle using an intention-to-treat (ITT) approach[119]. As a sensitivity analysis, patients were also evaluated according to the on-treatment approach.
Figure 10. Sequential recruitment. In each recruitment cycle, new users of the treatment and controls were included and added to the cohort to continuously increase the size of the study population.

Estimation of propensity score

A PS was estimated for all patients initiated on the new treatment during all cycles up to the point of analysis, and all patients initiated on the control drug during the same period. Logistic regression was used to estimate the PS.

Covariates included in the PS estimation were chosen based on knowledge of treatment of arrhythmias and information in the Summary of Product Characteristics for the drugs. We restricted covariate selection to only include factors affecting both treatment selection and the outcome (true confounders) or factors strongly related to the outcome (potential confounders). We used causal diagrams (www.dagitty.net)[120] to identify bias-minimized models. In order to accommodate changing prescribing patterns over time, we included the recruitment cycle and recruitment cycle-by-covariate interactions in the PS estimation[68, 69].

Study population

In Study III, new users of dronedarone (treated) were compared to new users of flecainide (controls) with regard to drug effectiveness. Flecainide was chosen as comparator because it is widely used in Sweden and globally for preservation of sinus rhythm in atrial fibrillation patients without structural heart disease. Flecainide was also chosen since no head-to-head comparisons between dronedarone and flecainide exist and are unlikely to be conducted. In study IV, which evaluated potential liver toxicity, dronedarone (treated) was compared to its predecessor amiodarone (control). Amiodarone was chosen since dronedarone is an analogue of amiodarone and structurally modified to
reduce the risks of toxic side effects. In the clinical trial program[121] for dronedarone, the frequencies of abnormal laboratory hepatic tests were similar between dronedarone and placebo groups. Post-approval, the FDA issued a warning about cases of rare but severe liver injury in patients treated with dronedarone[122].

Initiators of treated subjects and controls were primarily identified using EHRs. We also identified new dispensations of these drugs in the SPDR since these antiarrhythmic agents may be prescribed and initiated outside the hospital setting (and are therefore not always recorded as administrations in the EHRs).

Outcome definition

Study III
In the pivotal RCTs on dronedarone, EURIDIS and ADONIS[123], the primary outcome was recurrence of atrial fibrillation. To mimic this endpoint, we created a combined endpoint of the first occurrence of any of the following three conditions: (1) first hospitalization with atrial fibrillation as main discharge diagnosis; (2) first cardioversion, or (3) first recorded atrial fibrillation ablation.

Study IV
Adverse hepatic effects of dronedarone compared with amiodarone were evaluated using (1) survival analyses for liver-related events and (2) analyses of changes in mean levels of liver enzymes. For the survival analyses, we considered a composite endpoint defined as the first occurrence of any of the following conditions: (1) first recorded diagnosis of liver disease (ICD-10: K70-K77); (2) first recorded procedure code for liver surgery (procedure code: JJ)[124], or (3) first recorded elevation of the liver enzyme ALT (Alanine Aminotransferase) above 3 times the upper limit of normal (ULN)[125]. For evaluation in changes in mean levels of liver enzymes, we analyzed differences in ALT levels before and after initiation of dronedarone or amiodarone. Patients with at least one recorded ALT within 15 months before index were identified and monitored until the end of each recruitment cycle. The highest recorded value of ALT after initiation of treatment was selected.

Statistical analysis
In all studies, numbers and proportions were calculated for categorical variables and means, medians, standard deviations (SD) and interquartile ranges (study II and IV) were reported for continuous variables. In study II, chi-
square testing was used to analyze differences in drug treatment and healthcare use between Crohn’s disease and ulcerative colitis. Lorenz curves were also generated to assess skewness in healthcare use among patients with IBD.

**Statistical analyses - studies III and IV**

At the end of each recruitment cycle, Cox proportional hazards models were used to compare the incidence of the outcome between treated subjects and controls. Studies III and IV used an unadjusted Cox model (in the total cohort) and a model adjusting for a 3-knot restricted cubic spline of the PS (in the total cohort with PS in the region of common support). The assumption of proportional hazards was assessed using the ASSESS option of PHREG in SAS. Study III also evaluated the outcome in a cohort matched on the PS. Propensity score matching was done on the logit of the propensity score using calipers of width equal to 0.2 of the standard deviation of the logit of the PS[126], without replacement. Study IV also included a population consisting of patients with a recorded value of ALT before and after initiation of drug therapy. Logarithmic transformation was performed for ALT levels. Adjusted mean ALT changes were estimated (adjusted for 3-knot restricted cubic spline of the PS, pretreatment ALT and unbalanced covariates) and a ratio between ALT levels before and after initiation of treatment was calculated.

Covariate balance was assessed using standardized difference (total cohort and PS-matched cohort) and weighted conditional standardized difference (PS-adjusted cohort). Covariates with standardized differences or weighted conditional standardized differences of >0.1 were considered to have residual imbalance[82, 87]. If residual imbalance for baseline covariates was observed, PS-adjusted or PS-matched Cox models were automatically adjusted for these unbalanced covariates.

Missing data were imputed in five datasets using multiple imputation with chained equations[127]. Multiple imputation is a useful statistical technique for analysis of missing data that takes the uncertainty of missing data into account. This is done by creating several plausible imputed datasets and eventually combining the results obtained from each of them.

Before including data in each recruitment cycle, covariates were validated using logical checks and excluded if they were outside predefined threshold values.

To address the potential issue of residual confounding, we performed a sensitivity analysis to investigate how strong an unmeasured confounder would have to be to change the treatment estimate (study III)[92].

37
Software for data extractions and statistical analyses
SAS (version 9.2-9.4) (SAS Institute, Cary, NC) was used to build the models and for all analyses. Data from the EHR database was extracted using SQL Server Management Studio (Version 2005-2012) (MS SQL Server, Microsoft Corporation, Redmond, USA).
Ethical considerations

The Declaration of Helsinki constitutes the foundation for the examination of medical research ethics[128]. In 2008, the Ethical Review Act[129] was updated to clarify the term “research,” and the handling of personal data was redefined. Research involving sensitive personal data must be examined by the Ethical Review Board.

For the 4 studies included in this thesis, ethical approval was given from the Regional Ethical Review Board at Karolinska Institutet[130], Stockholm (2011/1076-31-3, 2014/781-31-1, 2011/1139-31-3, 2013/1978-32, 2014/1661-32). Permission to include data from EHRs was also given from the chief physicians of included caregivers.

All analyses were performed using de-identified data. Data has been stored in a password and firewall-protected server at the Healthcare Services Committee Administration at the Stockholm County Council.
Results and discussion

Paper I

In study I, which aimed to develop a method for data extraction of hospital-administered drugs, we demonstrated the potential of using EHRs to monitor drug therapy in the hospital setting. The extracted data included 11 different variables and data were complete for seven variables including encrypted PIN, age, time of administration, time of prescription, time of admission, time of discharge and healthcare unit. Information about sex was missing in one patient only.

Both automatic extraction and semi-automatic extraction generated the same number of patients treated with infliximab between 2007 and 2010 (Figure 11). However, the semi-automatic procedure generated 6.7% more infusions compared with the automatic procedure. Due to the additional data processing steps, the results of the semi-automatic procedure should be seen as the most reliable.

We chose to extract both the prescribed and administered dose for infliximab. However, dosage calculations were not possible for the administered doses, due to the fact that they were unstructured on account of the manual registration of this information when terminating the infusion in the EHR system. Information on the prescribed dose was extracted for >99.9% of all infusions successfully matched the corresponding treatment episode or visit. About 4% of these dosages required additional data processing before calculations could be performed. Dosage calculations were therefore excluded from the automatic procedure, which aimed to evaluate the performance of a fully automatic process. In the semi-automatic method, the prescribed dosage was extracted for >99.9% of all matched infusions. After excluding unstructured and non-interpretable dosages, numeric information on prescribed dosages was found for 98.3% of all infusions. Mean dosages per diagnosis are presented in Figure 12.
Figure 11. Comparisons of results with automatic and semi-automatic extraction procedures for patients administered infliximab in Karolinska University Hospital.

For analyses of the diagnostic profiles of infliximab, we used the primary diagnosis for the corresponding treatment episode/visit where infliximab was administered. For the automatic procedure, a corresponding diagnosis was found for 87.6% of all infusions (Figure 11). The corresponding figure for the semi-automatic model was >99.9%. When analyzing diagnostic profiles between the automatic and semi-automatic models, we observed large discrepancies, especially for inflammatory bowel disease (Figure 12). This is because many treatment episodes/visits at which infliximab was prescribed to treat Crohn’s disease or ulcerative colitis were not provided with a primary diagnosis and consequently recorded as having a missing diagnosis. The algorithms used in the semi-automatic model were able to find a corresponding diagnosis for all infusions except for one. It is possible that the recorded diagnoses for treatment episodes with infliximab did not always reflect the truth, because the patient may have been administered infliximab when admitted to
the hospital for another reason. However, the steps used in the semi-automatic procedure showed the potential for finding a valid diagnosis.

![Figure 12. Diagnostic profiles and dosages for infliximab. Dosage calculation for the automatic procedure only.](image)

* In the semi-automatic extraction procedure, different algorithms have been used to search and replace a missing diagnosis. ** Possible off-label use or that a correct diagnosis for the treatment with infliximab has never been registered.

In a following project we performed a validation analysis comparing the units of sold infliximab with the prescribed dosage of infliximab. This validation found that 97.3% of all sold units of infliximab were prescribed, which must be considered a high degree of coverage[45].

This study has shown the potential to monitor hospital drugs using data from EHRs. However, some manual processing of the data (development of algorithms) is needed before data can be analyzed. The study findings, including shortages in data, have been communicated to prescribers with the expectation that establishing a feedback loop from researchers to clinicians can gradually improve the data entry.
Paper II

Using data from population-based registers and EHRs, study II described drug treatment and use of healthcare services in IBD. The most important finding of study II for this thesis was that the data-extracting techniques developed and evaluated in study I could be used and linked with data from other registers to obtain the precise prevalence of biological drug use in IBD. We found that 5.7% (Crohn’s disease (CD): 9.6%, ulcerative colitis (UC): 2.9%, p<0.0001) of all IBD patients were treated with biologicals (Figure 13). The corresponding number in a recent nationwide Swedish study was 3.2% (CD: 5.2%, UC: 1.3%)[116]. The large discrepancies observed in biological treatment may be somewhat influenced by different study time periods (2013 in our study and 2010 in the nationwide study) and different regions (Stockholm versus Sweden). However, the higher estimate in our study is most likely explained by the inclusion of biologicals administered in the hospital setting. Such administrations were lacking in the other study. Therefore, to generate a precise prevalence estimate of drugs used as both prescription drugs and in hospitals, it is necessary to include data from EHRs.

![Figure 13. Inflammatory bowel disease drug treatment in Stockholm at January 1, 2013.](image)

We estimated the prevalence of IBD to 0.65% which was similar to the results from the earlier Swedish nationwide study[116]. The prevalence of IBD in Sweden is therefore among the highest in the world[131], comparable to estimates from Canada, the country with the highest reported IBD incidence and prevalence rates[132].
Furthermore, we found a highly disproportionate use of hospital services among IBD patients, where only a small part of the patient groups accounted for all hospitalizations (Figure 14).

![Figure 14. Lorenz curve displaying the proportion of inpatient healthcare utilization during 2013 among all patients in each inflammatory bowel disease group (n=5690 Crohn’s disease, n=7409 ulcerative colitis and n=817 inflammatory bowel disease unclassified).](image)

We also observed that among patients undergoing surgery, only 23.8% of patients with CD and 8.8% of patients with UC were treated with biologicals before the operation. Given that biologicals represent the recommended treatment step before surgery[133, 134], in some but not all cases this might indicate a possible under-use of biologicals. This finding has led to a thorough review of the medical records of IBD patients in Stockholm county to provide further insight into the health care these patients receive.

In a sub-analysis aiming to identify use of corticosteroids in the IBD population, 85% were not treated with corticosteroids, suggesting that a high proportion of patients are in steroid-free remission. However, 4% of the total IBD population was considered high users of corticosteroids. These patients may be considered for a more aggressive maintenance treatment for IBD instead.

Finally, we found that patients with CD comprise the group with the highest medical needs. Patients with CD used significantly more healthcare resources (including outpatient visits, hospitalizations, and surgeries) than patients with UC. Twice as many CD patients received immunomodulators compared with
UC patients, and CD patients were treated with biologicals three times more often. These results highlight the fact that CD remains a challenge and further efforts are needed to improve care for these patients.

Paper III

In study III, we developed a generic model for automatized real-time monitoring of drugs. Patients were sequentially recruited and in the last recruitment cycle of the study, 1249 patient with dronedarone and 840 patients with flecainide were included (Figure 15). Of those, 95% had a PS within common support and were included in the cohort for PS-adjusted analysis. In analyses matching on the PS instead, only 46% of initiators of dronedarone were possible to match to initiators of flecainide. The fact that over 50% of the population is excluded in the PS-matched analyses might lead to a more selected sample, diminishing two of the great benefits of real-world comparative effectiveness studies: the generalizability of results and the possibility to monitor unusual events. Allowing a wider caliper (than 0.2 of the SD) might increase the number of matched patients with dronedarone, but instead increase imbalances in covariates. Since the PS-adjusted model included 95% of the total study cohort, this model might therefore have a higher external validity than the PS-matched model. However, PS-adjusted models might be more sensitive to covariate imbalances[135]. Therefore, it may be preferable to evaluate treatment estimates using different PS techniques and that is the approach we have adopted.

In the total cohort, we observed imbalances for 17 out of 30 (57%) baseline covariates. Dronedarone users were older, had more comorbid conditions and used more health services. However, these imbalances could be addressed in the PS-matched model in which no residual imbalance was observed after matching on the PS. In the PS-adjusted cohort, residual imbalances were observed for 5 covariates (Figure 16). These imbalanced covariates were automatically adjusted for in the Cox models. Residual imbalances in the PS-adjusted model are likely explained by an under-dimensioned cohort that is unable to balance all covariates.
Figure 15. Inclusion rate of dronedarone and flecainide in the total (unadjusted), PS-adjusted and PS-matched cohorts

Figure 16. Standardized differences in baseline covariates between dronedarone and flecainide for the total (unadjusted), PS-adjusted and PS-matched cohorts (in the last recruitment cycle). (A standardized difference <0.10 indicates negligible imbalances).
The outcome, recurrence of atrial fibrillation, occurred in 697 initiators of dronedarone (52 of 100 person years) and 409 initiators of flecainide (34 of 100 person years).

In Figure 17, hazard ratios according to the ITT approach are plotted for each of the 8 recruitment cycles. A significant difference in favor of flecainide was seen early in the study. In analyses after the last recruitment cycle, the PS-adjusted model and the PS-matched model produced similar results (PS-adjusted HR 1.47 [1.27-1.70]; PS-matched HR 1.48 [1.25-1.76]) for recurrence of atrial fibrillation in initiators of dronedarone vs. flecainide. The corresponding HR for the unadjusted analyses was 1.35 (1.19-1.52). This lower HR in the unadjusted analyses can be explained by the observation that recurrence of atrial fibrillation was driven mainly by age, with younger patients seeking care for AF to a higher extent, whereas dronedarone users were older than flecainide users.

![Figure 17. Sequential analyses for recurrence of atrial fibrillation between dronedarone and flecainide.](image)

The sensitivity analysis for unmeasured confounding showed that to affect the observed findings, an unmeasured binary confounder should have an hazard ratio for the outcome of at least 3 and a difference in prevalence between dronedarone and flecainide of at least 20%.
We chose the ITT principle as our main analysis strategy since ITT reflects the actual clinical scenario for effectiveness determinations. One argument against ITT is that it may produce conservative results because of dilution due to non-compliance. We therefore complemented our ITT analyses with on-treatment analyses. As expected, on-treatment models produced higher hazard ratios than the ITT models in this study (HR for PS-adjusted: 1.67 [1.41-1.97]), HR for PS-matched: 1.63 [1.32-2.00]).

Compared to the PS-matched analyses, the PS-adjusted analyses presented more precise (narrower CI) and more stable treatment estimates. This is explained in part by a larger study population in the PS-adjusted cohort.

Using data from EHRs, administrative health databases and population demographics registers, we have built and evaluated an automatized model for real-time monitoring of effectiveness of new drugs. We demonstrate that this automatized model produces consistent results with increasing precision over time as data on the new drug accumulates.

**Paper IV**

In this study, we amended the automatized model developed in study III to be able to monitor the post-marketing safety of drugs. Dronedarone was compared to its predecessor amiodarone, and over 8 recruitment cycles, we identified 2,296 eligible patients initiating treatment with dronedarone (n=1,244) and amiodarone (n=1,052) (total cohort). Of the total cohort, 91.5% had a PS within the region of common support and were included in the PS-adjusted analyses (PS-adjusted cohort). We observed a higher proportion of unbalanced baseline covariates in the dronedarone vs. amiodarone comparisons in study IV compared with the dronedarone vs. flecainide comparisons in study III. In the total cohort, 21 out of 24 covariates (88%) were unbalanced (Figure 18). Balance diagnostics for the PS-adjusted cohort showed residual imbalance for 4 covariates. Unbalanced covariates were included in the Cox models and in analyses of the adjusted mean changes of ALT levels.
In the secondary study population aimed at analyses of changes in mean ALT levels, we identified 1,061 patients (n=493 dronedarone, n=568 amiodarone) with at least one value of ALT measured before and one measured after initiation of treatment. Of those, 935 (88.1%) had a PS within the region of common support and were also included in the PS-adjusted analyses of changes in mean ALT.

In the survival analyses, liver-related events were rare and similar in patients treated with both dronedarone and amiodarone. The composite outcome occurred in 56 patients (25 initiators of dronedarone [9.4 per 1000 person years] and 31 initiators of amiodarone [13.4 per 1000 person years]). In analyses of the last recruitment cycle, the HR for the PS-adjusted analysis was 0.74 (0.39-1.42) (Figure 19). In sensitivity analyses using an on-treatment approach, the HR for the PS-adjusted analysis was 0.88 (0.35-2.25)
Figure 19. Sequential analyses for incidence of liver-related events between dronedarone and amiodarone.

Separating each included condition in the composite outcome (diagnosis of liver disease, surgery on the liver or ALT >3 ULN), we found that 76% of the outcomes for dronedarone patients and 65% of the outcomes for amiodarone patients were due to elevations in ALT (>3 ULN). Consequently, 24% in the dronedarone group and 35% in the amiodarone group were either a diagnosis of liver disease or a surgical procedure on the liver.
In analyses of changes in liver enzymes, we observed an increase in adjusted mean ALT for both dronedarone and amiodarone (Figure 20). This increase was more pronounced ($p=0.001$) for dronedarone patients, which saw a 44% increase in mean ALT in the last recruitment cycle. The corresponding number for amiodarone was 21%. In on-treatment analyses, a 37% and 12% increase in mean ALT was observed for dronedarone and amiodarone, respectively ($P<0.001$). When analyzing more substantial elevations in ALT, we observed ALT values of >5 ULN for 0.7% in the dronedarone group and 0.9% in the amiodarone group using the on-treatment approach. These numbers were in concordance with findings from the clinical trials of dronedarone, where 0.8% of dronedarone patients had an ALT value of >5 ULN according to on-treatment analysis. Extreme ALT values above >25 ULN were observed for 4 amiodarone patients. ALT values above 25 were observed for 1 patient in the amiodarone group according to on-treatment analysis.

Figure 20. Adjusted mean changes in ALT before and after initiation of dronedarone or amiodarone (presented as a post/pretreatment ALT ratio). *No patients within PS region of common support.
In study IV, we expanded on the work from study III and demonstrated a strategy for real-time follow-up of safety outcomes for newly marketed drugs. The model can be used for early detection of signs of side effects by monitoring indicators of drug toxicity, such as liver enzymes, or by monitoring hard outcomes. The seemingly discrepant results for mean ALT levels (higher among dronedarone patients) and extreme levels (slightly higher among amiodarone patients) illustrate the usefulness of our model. The model can be built to monitor central measures (means, medians) or extreme levels for different physiological fields. Extreme reactions are likely to be more important than slight shifts in central measures for e.g. liver transaminases or eGFR (estimated glomerular filtration rate), whereas a slight elevation in e.g. mean blood glucose or blood pressure may be enough to take a preventive drug off the market. We envision monitoring of both extreme reactions and central measures to be warranted in the case of new drug classes with limited safety information.
General discussion and conclusions

Data from EHRs have been called “a goldmine for research.”[136, 137] Although these information systems collect huge amounts of valuable information on healthcare, EHRs are still underutilized for evaluation of quality of care and research.

Using the developed extraction techniques and model for real-time monitoring we were able to demonstrate drug use and health service utilization for inflammatory bowel disease, the comparative effectiveness and safety of antiarrhythmic drugs.

Real-time monitoring

This thesis is entitled “Real-time monitoring of healthcare interventions in routine care.” However, “real-time” is a relative concept. In the best of worlds, structured monitoring of drugs and other healthcare interventions should be monitored in real time. Until that milestone is reached, we have to settle with monitoring in “near” real time. Of the available data sources for drug monitoring, information in EHRs and in the population registry is the closest to real time. Since the medical records are recorded as part of patient care, EHRs are instantly updated. The EHR database used in this thesis was updated daily. Therefore, if our proposed model were to be used in combination with this EHR system (for exposure and morbidity) and the population registry (for mortality), drugs could theoretically be monitored on a daily basis. Such frequent evaluation might seldom be necessary, although it can be useful for early detection of signs of adverse events.

Generalizability

The developed models for extractions and for routine monitoring have been evaluated using biological drugs and drugs used for the treatment of atrial fibrillation. Since the models are generically built, they are also applicable to other therapies, as long as the interventions are traceable in EHRs. Comparative studies will also depend on the potential to find a relevant and measurable outcome and an appropriate comparator substance.
It should be mentioned that the propensity score is not only used to perform non-randomized studies on drugs. We observe an increased use of propensity scores in psychological, educational and economic analyses[76, 138, 139]. Since the ultimate goal of the PS is to achieve balance between study groups, our proposed model should therefore also be applicable to study other healthcare interventions, such as surgical procedures and diagnostic tools.

In this thesis, we primarily worked with one EHR system. Since there are many different EHR systems on the market (6 major systems exist in Sweden)[47] with differences in infrastructures, the scripts, codes and algorithms developed in these studies will have to be slightly modified for use in other systems. However, we have built the real-time monitoring system as generically as possible and developed the model in one of the major analysis software packages on the market. In practice, if raw data is extracted, cleaned and transformed according to pre-defined standards, data can easily be imported into the monitoring model.

Data completeness

For building the model for automatized real-time monitoring of drugs, we primarily used data from EHRs, but supplemented them with data from the regional healthcare data warehouse of Stockholm (VAL), and registers at the National Board of Health and Welfare and Statistics Sweden. The EHR system TakeCare used in these studies has overall coverage in Stockholm county of about 75% of outpatient care and 88% of inpatient care[102]. Had we only included data from TakeCare, we may have therefore missed information for some study participants. The VAL database includes diagnoses and procedures for all contacts in health care financed by the county council. By including data from VAL, coverage of diagnoses and clinical procedures could be increased. However, clinical measurements and laboratory information are not included in VAL and therefore, this missing data was imputed using multiple imputation. Since VAL is updated on a monthly basis, the real-time concept will move from a daily basis to a monthly basis if key data are to be obtained from VAL.

We also included data from the National Patient Register and the Prescribed Drug Register at the Swedish National Board of Health and Welfare and from Statistics Sweden. These registers were included to obtain information on Stockholm residents seeking care outside of Stockholm county to avoid missing out on important information on covariates, drug treatment or outcomes. They were also included to obtain information on population demographics.
However, including these national data sources will further increase the time in the “real time” concept due to the process of linking records.

In several sensitivity analyses aimed at evaluating how excluding some of the databases will affect the outcomes (data not shown), we found that including only EHRs and VAL produced similar results compared to including all data sources used in this study. This may suggest that these two data sources (EHR and VAL) are sufficient to use when performing analyses in Stockholm county, at least for analyses of antiarrhythmic treatment. Furthermore, by including only EHRs and VAL, real-time monitoring can take place on a monthly basis. The actual data sources used will, of course, depend on the specific research question, and one may also need additional data sources that were not used in this thesis (e.g. disease-specific quality registers and including patient-reported data).

Residual imbalance and hidden bias

In the automatized monitoring model (studies III and IV) we used directed acyclic graphs to choose which covariates to include in the estimation of the propensity score. Before estimating treatment effects, goodness-of-fit diagnostics were performed. These balance checks of covariates showed, overall, that the majority of covariates were balanced in the PS analyses. In study III, the PS adjusted model showed that 5 covariates were unbalanced. However, the observed imbalances were moderate since the largest weighted conditional standardized difference was 0.19. Some have suggested that the standardized difference should be less than 0.25, although the trend is that a standardized difference of less than 0.1 is considered standard. We chose to address unbalanced covariates by including them in the Cox regression models and in the estimation of adjusted mean changes in ALT.

Hidden bias concerns unobserved characteristics of the study population that are unbalanced between study groups. This is not considered an issue in randomized trials of sufficient size, since the logic of randomization is that both observed and unobserved characteristics should be balanced between treated subjects and controls. More conventionally, hidden bias is a limitation in observational studies. We have tried our best to address this via a comprehensive data extraction from the data sources and a stringent procedure for model specification using state-of-the-art causal inference techniques; but missing data due chiefly to unstructured charting, for example of tobacco habits, and echocardiography findings, underpins the residual confounding risk.
Other models for real-time monitoring

To our knowledge a few other models for real-time sequential monitoring have been proposed [61, 91, 140, 141]. These have mainly been based on administrative claims data and focused on monitoring adverse drug reactions.

Toward real-time monitoring

With country-wide use of EHRs, national healthcare registers with virtually complete coverage and the fact that all residents have access to health care, Sweden can have a strong position in the post-market monitoring of effectiveness, safety and the health-economic value of drugs. The presence of these unique data sources, the unique PIN for data linkage, and analysis with minimal loss to follow-up will overcome limitations that exist in most other countries.

However, to take a lead in post-marketing monitoring of drugs, the Swedish government should intensify the work with standards and infrastructure for healthcare information to be able to include nationwide EHR data in analyses of the quality of care and research. This work must be accompanied by discussions of regulations regarding how to protect personal integrity.

We have presented how data on drug use from EHRs can be extracted and used to routinely monitor drug therapy. With a rapidly evolving drug development, it is important to optimize the evaluation of effectiveness, safety and health economic value of new medicines in routine care. We believe that the models described in this thesis could contribute to fulfil this need.
Future perspectives

During the work developing and evaluating these models for sequential monitoring of drugs we have identified areas to be further investigated and evaluated.

In this thesis, we evaluated the model according to drug’s effectiveness and safety. The next step will be to analyze the performance of the model when evaluating cost-effectiveness of drugs.

We will also investigate how this model works with other data types and sources of data, for other healthcare interventions than drugs, legal and technical aspects of using data from EHRs for routine monitoring and how this model can be used in the structured introduction of drugs in Sweden[142].

Since RCTs often have strict inclusion criteria, one major benefit of CER in a real-world setting is the generalizability of the results. Therefore, one goal must be to include as many patients and follow-up periods as possible in the analysis. However, the propensity score framework is poorly evaluated in multilevel models and therefore patients often only occur once in PS-analysis. We are therefore planning to investigate and evaluate how PS can be used in a multilevel setting when the same patients may start treatment with two or more different drug therapies.

It would also be of great interest to identify countries with similar conditions as Sweden and establish collaborations to perform cross-national sequential monitoring of drugs. However, since we have seen that drug data in hospitals is lacking in most European countries such cross-national collaboration on routine monitoring of drug effectiveness and safety might begin evaluating prescription drugs. For this, the Nordic countries have a unique potential for cross-national pharmacoepidemiological studies[143].

Innan ett läkemedel når marknaden har dess effekt och säkerhet utvärderats i randomiserade kliniska studier. Dessa studier anses vara den gyllene standarden för att utröna effekten av olika behandlingar. Ofta görs dock dessa studier på en selekterad studiepopulation som inte alltid motsvarar de patienter som kommer att behandlas med läkemedlet i klinisk rutinvård. I klinisk vardag används de nya läkemedlen ofta under en längre period än vad som studerats och av patienter med annan samtidig sjuklighet och läkemedelsbehandling. Det är därför inte säkert att de resultat man sett i kliniska studier av ett nytt läkemedel motsvarar de resultat man får i den kliniska vardagen.


I det första delarbetet i denna avhandling utvecklade och utvärderade vi en metod för att kunna använda data från elektroniska journalsystem för att kunna följa läkemedel som ges på sjukhus. Vi använde anonymiserade data från journalsystemet TakeCare som är Stockholms läns landstings huvudsakliga journalsystem och vi valde att extrahera data för det biologiska läkemedlet *infliximab* som bl.a. används vid reumatism. Resultaten visade att det är möjligt att via journalsystemet TakeCare kunna följa hur det läkemedlet *infliximab* används då vi bland annat kunde studera vilka patientpopulationer som använder detta läkemedel, i vilka doseringar samt på vilken indikation.

I det andra arbetet använde vi metoden från första delarbetet för att bland annat analysera läkemedelsbehandling vid inflammatorisk tarmsjukdom (IBD). Läkemedelsbehandlingen vid IBD består av både läkemedel som förskrivs via recept men även av biologiska läkemedel som ofta administreras på sjukhus eller vid sjukhusanknuten vård. Tidigare studier kring läkemedelsbehandling vid IBD har främst baserats på receptförskrivna läkemedel och det har därför varit svårt att få en helhetsbild över hur behandlingen vid IBD ser ut. I vår studie inkluderade vi data både från elektroniska journalsystem och från läkemedelsregistret som innehåller information om uthämtade läkemedel på recept. Genom att använda båda dessa informationskällor kunde vi få en bra bild över läkemedelsbehandlingen vid IBD.


Att utvärdera hur läkemedel fungerar i rutinsjukvård är betydligt svåreare än i en randomiserad klinisk prövning. Syftet med randomiseringen är att få jämförbara grupper då det är slumpen som avgör om patienten får behandling A eller B. I klinisk rutinvård kan man inte göra en sådan randomisering och det är många olika faktorer som påverkar valet av läkemedel A eller B och därmed blir grupperna ofta inte direkt jämförbara. Detta kan dock till stor del hanteras genom olika statistiska metoder.

Modellen framtagen i studie III och IV bygger på att man kontinuerligt utvärderar ett nytt läkemedel mot ett annat läkemedel i rutinsjukvård. Jämförelseläkemedlet kan tex vara bästa tillgängliga behandling som fanns innan det nya läkemedlet började användas. Genom statistiska analyser...
försöker man efterlikna en randomiserad studie genom göra grupperna så jämförbara som möjligt, vad avser riskfaktorer, tidigare läkemedelsbehandling, bakomliggande sjukdomar mm.

I denna avhandling har vi visat att det är möjligt att använda data från elektroniska journalsystem för att analysera läkemedelsanvändningen på sjukhus. Vi har även visat att genom att använda dessa data kan vi kontinuerligt följa upp nya läkemedel med avseende på dess nytta och säkerhet i rutinsjukvård.
Acknowledgements

I would like to express my sincere gratitude to everyone who has supported me, inspired me, and contributed to make this thesis possible, with special thanks to:

**Johan Sundström**, my main supervisor, for introducing me into epidemiology! Thank you for your enthusiasm, engagement, extensive skills and brilliant ideas. You are a true visionary and it has been a privilege having you as my supervisor!

**Björn Wettermark**, my co-supervisor, for introducing me and teaching me about pharmacoepidemiology. It has been a privilege to gain from your never-ending energy and enthusiasm and your unique knowledge in the field of pharmacoepidemiology!

**Martin Neovius**, my co-supervisor, for brilliant ideas and helpful comments along the road! Looking forward working with you in the field of health economics.

**Lars Lindhagen**, my “statistical therapist”. I am so impressed of your brain and skills and grateful for all advice during this journey!

**Rickard Malmström, Bo Vikström, Ulf Bergman, Bo Ringertz, Robert Löfberg, Mikael Lördal and Jonas Schwieler**, my co-authors, for providing excellent ideas and comments that substantially improved the papers.

**Lars L Gustafsson**, for your encouraging support during my first study.

All colleagues at hälsö- och sjukvårdsförvaltningen at SLL.

My “läkemedelsstatistik” friends at SLL. *Desirée Loikas, Elin Dahlén, Gunnar Ljunggren, Maria Juhasz Haverinen, Helena Ramström, Pia Frisk, Sten Ronge* and *Irene Eriksson*, for all the nice company and interesting discussions!

**The RUDE-group** and Torbjörn for sharing your expertise in data integrity.
Marie-Christine Berglund, Per Lindström, Gunnar Ekeving, Mattias Öhlén and Qemajl Imeri for always kind support regarding databases and SQL.

Göran Lord (“the king of SAS-arrays”) and Per Haglund, for kindly responding to my questions regarding SAS and VAL.

Emil Hagström, for geeky discussions, great statistical company in London, and for teasing me about everything!

Maria Bertilsson, Frida Kåver, Anna Stendahl and everybody else at UCR for nice discussions and company during my year at UCR.

Sven-Åke Lööv, for a productive collaboration when working with introduction and follow up of new drugs in SLL, during the INNOVIS-project.

Per Nasfi, for the cover illustration.

Karolinska Universitetssjukhuset Danderyds sjukhus, Södersjukhuset and Stockholm läns sjukvårdsområde for kindly contributing with data for my studies.

My dear friends! I feel incredibly lucky to have you in my life!

And last:

Lars and Maria, for your love and for always being there!

The Renn-Grass family with a special thanks to Gunilla for all help with looking after the children and correcting my English grammar!

Otto, Karin, Ninna and Stefan with families for your love and support!

And finally, Fia, Emil, Hjalmar and Wille. You are my everything!
References


120. DAGitty, *Drawing and analyzing causal diagrams (DAGs)*. Available from: http://www.dagitty.net/ [Accessed: September 26, 2016].


A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)