DEPARTMENT OF MEDICINE, SOLNA
Karolinska Institutet, Stockholm, Sweden

PHARMACOEPIDEMIOLOGIC STUDIES OF DRUG SAFETY IN PEDIATRIC CHRONIC INFLAMMATORY DISEASE

Viktor Wintzell

Karolinska Institutet

Stockholm 2021

Published by Karolinska Institutet.
Printed by US-AB

ISBN 978-91-8016-239-5
Cover plot shows the tree structure of diagnoses from data mining analysis of adverse events; results are presented in section 4.3.2.

Pharmacoepidemiologic Studies of Drug Safety in Pediatric Chronic Inflammatory Disease

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Viktor Wintzell

The thesis will be defended in public on June 16, 2021, 14:00 at Inghesalen, Widerströmska huset, Karolinska Institutet, Solna and online (find link at ki.se).

Principal Supervisor:
Docent Björn Pasternak
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

Co-supervisors:
Doctor Henrik Svanström
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

Professor Jonas F. Ludvigsson
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Opponent:
Professor Morten Andersen
University of Copenhagen
Faculty of Health and Medical Sciences
Department of Drug Design and Pharmacology

Examination Board:
Professor Fredrik Nyberg
University of Gothenburg
School of Public Health and Community Medicine

Docent Sara Öberg
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Professor William O. Cooper
Vanderbilt University Medical Center
Department of Pediatrics, Department of Health Policy

To Ana, Ella

ABSTRACT

Safety evidence for use of pharmaceutical drugs in children, including treatments for serious conditions such as chronic inflammatory diseases, is generally scarce. Off-label use is common and clinicians need to rely on evidence from adults when prescribing to children. This is concerning because safety profiles might differ; the metabolism, distribution and absorption of drugs vary between children and adults.

The overall aim of this thesis was to develop new, relevant, and pediatric-specific drug safety evidence for treatments of chronic inflammatory disease; both addressing specific safety concerns and screening for signals of previously unknown adverse events. Sub-aims were to evaluate the feasibility of these types of safety studies in the Scandinavian setting and to examine the differences between alternative pharmacoepidemiologic study designs. We conducted analyses based on data from Swedish and Danish national registers covering a source population of 5.3 million children; including 21,000 patients with confirmed chronic inflammatory disease.

In the first study, the aim was to investigate if there is an association between the use of azathioprine and the risk of acute pancreatitis in Swedish and Danish patients with pediatric inflammatory bowel disease (IBD). We found that azathioprine was associated with a 6-fold increased risk of acute pancreatitis during the first 90 days following treatment initiation, compared to no use, based on a sample of 8725 patients (n=3574 azathioprine users).

In the second study, we investigated if there is an association between use of tumor necrosis factor-alpha (TNF-α) inhibitors and the risk of serious infection in patients with pediatric IBD in Denmark. We found no significant association between current use of TNF-α inhibitors and the risk of serious infection, based on 2817 patients (n=618 TNF-α inhibitor users), in comparison with no use.

The aim of the third study was to perform data mining to detect previously unknown adverse events of TNF-α inhibitors in children with IBD or juvenile idiopathic arthritis (JIA) in Denmark. We used tree-based scan statistics on 1284 incident diagnoses identified during follow-up and found two significant signals, dermatologic complications and psychiatric adjustment disorders. Neither of these signals were considered relevant for further investigation.

In the fourth study, we systematically described and compared various pharmacoepidemiologic designs, in particular alternatives to the active comparator new user design. We used target trial emulation as a common framework and drew two conclusions. That eligibility is the key design element that differentiates the designs and that many factors influence the choice of an ideal comparator, including indication, available comparator drugs, treatment patterns, potential effect modification, and sample size.

In the fifth and final study, we investigated if there is an association between use of TNF-α inhibitors and the risk of serious infection in Danish patients with JIA. Based on 4493 JIA patients (n=578 TNF-α inhibitor users), we found that current use of TNF-α inhibitors was associated with a two-fold increased risk of serious infection, compared to methotrexate.

In summary, we provided data on three current drug safety concerns in children with chronic inflammatory disease; we showed that Scandinavian health registers are suitable for both targeted and adverse-event signal detection studies; and finally, we provided guidance on the factors that need to be considered when selecting comparators in pharmacoepidemiologic studies.

LIST OF SCIENTIFIC PAPERS

I. Wintzell V, Svanström H, Olén O, Melbye M, Ludvigsson JF, Pasternak B
Association between use of azathioprine and risk of acute pancreatitis in children with inflammatory bowel disease: a Swedish–Danish nationwide cohort study.
The Lancet Child & Adolescent Health; 2019; 3(3):158-165

II. Wintzell V, Svanström H, Melbye M, Jess T, Olén O, Ludvigsson JF, Pasternak, B
Use of tumour necrosis factor-α inhibitors and the risk of serious infection in paediatric inflammatory bowel disease in Denmark: a nationwide cohort study.
The Lancet Gastroenterology & Hepatology; 2019; 4(11):845-853

III. Wintzell V, Svanström H, Melbye M, Ludvigsson JF, Pasternak B Kulldorff, M
Data mining for adverse events of tumor necrosis factor-alpha inhibitors in pediatric patients: tree-based scan statistic analyses of Danish nationwide health data.
Clinical Drug Investigation; 2020; 40(12):1147-1154

IV. Wintzell V, Svanström H, Pasternak B
Selection of comparator group in observational drug safety studies – alternatives to the active comparator new user design.
Submitted

V. Wintzell V, Svanström H, Melbye M, Ludvigsson JF, Pasternak, B
Association between use of tumor necrosis factor-alpha inhibitors and the risk of serious infection in juvenile idiopathic arthritis – a Danish nationwide cohort study.
Manuscript

Papers are referred to by their roman numerals.

LIST OF RELATED SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

Ueda P, Wintzell V, Melbye M, Eliasson B, Svensson Am, Franzén S, Gudbjörnsdottir S, Hveem K, Jonasson C, Svanström H, Pasternak B
Use of incretin-based drugs and risk of cholangiocarcinoma: Scandinavian cohort study
Diabetologia; 2021 in press

Wang Yh, Wintzell V, Ludvigsson Jf, Svanström H, Pasternak B
Association between proton pump inhibitor use and risk of asthma in children
JAMA pediatrics; 2021;175(4):394-403

Flam B, Wintzell V, Ludvigsson Jf, Mårtensson J, Pasternak B
Direct oral anticoagulant use and risk of severe COVID-19
Journal of internal medicine; 2021; 289(3):411-419

Wang Yh, Wintzell V, Ludvigsson Jf, Svanström H, Pasternak B
Association between proton pump inhibitor use and risk of fracture in children
JAMA pediatrics; 2020; 174(6):543-551

Pasternak B, Wintzell V, Eliasson B, Svensson Am, Franzén S, Gudbjörnsdottir S, Hveem K, Jonasson C, Melbye M, Svanström H, Ueda P
Use of glucagon-like peptide 1 receptor agonists and risk of serious renal events: Scandinavian cohort study
Diabetes care; 2020; 43(6):1326-1335

Pasternak B, Wintzell V, Melbye M, Eliasson B, Svensson Am, Franzén S, Gudbjörnsdottir S, Hveem K, Jonasson C, Svanström H, Ueda P
Use of sodium-glucose co-transporter 2 inhibitors and risk of serious renal events: Scandinavian cohort study
BMJ (Clinical research ed.); 2020; 369:m1186

Pasternak B, Ueda P, Eliasson B, Svensson Am, Franzén S, Gudbjörnsdottir S, Hveem K, Jonasson C, Wintzell V, Melbye M, Svanström H
Use of sodium glucose cotransporter 2 inhibitors and risk of major cardiovascular events and heart failure: Scandinavian register based cohort study
BMJ (Clinical research ed.); 2019; 366:l4772

Pasternak B, Wintzell V, Furu K, Engeland A, Neovius M, Stephansson O
Oral fluconazole in pregnancy and risk of stillbirth and neonatal death
JAMA; 2018; 319(22):2333-2335

CONTENTS

LIST OF ABBREVIATIONS

The following abbreviations have been used in the thesis summary:

5-ASA	5-aminosalicylic acid
ACNU	Active comparator new user
ATC	Anatomical Therapeutic Chemical
ATE	Average treatment effect
ATT	Average treatment effect in the treated
CD	Crohn's disease
CI	Confidence interval
DAG	Directed acyclic graph
eCDF	empirical cumulative density function
HR	Hazard ratio
IBD	Inflammatory bowel disease
ICD	International Statistical Classification of Diseases and Related Health Problems
IPC	Inverse probability of censoring
IPT	Inverse probability of treatment
IR	Incidence rate
IRR	Incidence rate ratio
JIA	Juvenile idiopathic arthritis
LISA	Longitudinal Integration Database for Health Insurance and Labour Market Studies
LLR	Log likelihood ratio
MTX	Methotrexate
NSAID	Nonsteroidal anti-inflammatory drug
OR	Odds ratio
PS	Propensity score
RA	Rheumatoid arthritis
RCT	Randomized controlled trial
RR	Risk ratio
SD	Standard deviation
SGLT2	Sodium glucose cotransporter 2
SMR	Standardized mortality ratio
TNF-α	Tumor necrosis factor-alpha
UC	Ulcerative colitis

1 INTRODUCTION

> *"The committee believes that it is unethical to adhere to a system which forces physicians to use therapeutic agents in an uncontrolled experimental situation virtually every time they prescribe for children. Furthermore, it is not only ethical but also imperative that new drugs to be used in children be studied in children under controlled circumstances so the benefits of therapeutic advances will become available to all who may need them"*
>
> –American Academy of Pediatrics[1]

The formal testing of pharmaceutical drugs before they are used in clinical practice was established during the previous century. Historic disasters have prompted the development of new legislation, in particular the death of more than 100 people in 1937 after using a sulfanilamide elixir and the widespread use of thalidomide in the late 1950s that caused deformities in more than 5000 babies.[2] The modern approval of novel drugs is a highly regulated process where efficacy and safety are assessed based on clinical trials in increasingly larger cohorts. Despite this progress, some patient groups are still underrepresented in or even excluded from the testing of new therapies. One such group is children.[3,4]

There is generally less drug safety information in children than in adults; the proportion of drugs used in children without proper labeling has been estimated at 54%.[5] The participation of children in clinical trials is very low due to ethical, practical and financial reasons.[6,7] As a consequence, clinicians treating children need to make decisions based on data extrapolated from adults and clinical judgement rather than specific safety evidence for children. This can lead to suboptimal treatment because children are different from adults with respects to their physiology, organ development, and their drug absorption, distribution and metabolization.[8] Hence the safety profile of a drug can be different in pediatric patients in comparison with adults.

Off-label pharmacotherapy in children is high.[9,10] Since market forces alone have not stimulated the development of pediatric safety evidence several legal and regulatory efforts have been made to increase the inclusion of children in randomized controlled trials (RCT), both in the United States and Europe.[11] The first efforts to stimulate inclusion of children were made in the United States in the late 1990s. Although they

had positive long-term effects and more children have been included in trials, there is still a general lack of safety evidence for pediatric patients. The legal incentives have also been criticized for incurring high societal cost and not prioritizing pediatric-specific studies based on clinical needs.[4,12] Furthermore, results from many clinical trials of children are not published. The proportion of pediatric phase III RCTs that are not published is estimated at 30% and one factor influencing nonpublication is failure to enroll enough patients.[13]

Given the obstacles in conducting RCTs on children and the limited output, observational studies play a critical role in the development of safety evidence for pediatric patients.[14-16] There are several general advantages of observational safety studies, among them the possibility to analyze larger samples with longer follow-up, which allows for higher precision and the study of rare adverse events. This is particularly important when studying children as they have lower prevalence of disease and drug use; resulting commonly in sample size issues in RCTs.[13,17] Studying drug safety from routine clinical care also increases relevance and generalizability of the results. Despite the advantages of observational studies, and that they in many cases represent the only source of pediatric-specific evidence, there is still a shortage of this type of research.[16]

In this thesis project we addressed current drug safety concerns in children with focus on chronic inflammatory diseases. The overall aim was to develop new, relevant, and pediatric-specific drug safety evidence for treatments in pediatric inflammatory bowel disease (IBD) and juvenile idiopathic arthritis (JIA); both addressing specific safety concerns and screening for new signals of adverse events. Sub-aims were to evaluate the feasibility of these types of safety studies in the Scandinavian health care register setting and to investigate the pros and cons of common pharmacoepidemiologic study designs. The specific aims of the papers were:

- *Study I*: To investigate if there is an association between use of azathioprine and the risk of acute pancreatitis in Swedish and Danish children with IBD.

- *Study II*: To investigate if there is an association between use of tumor necrosis factor-alpha (TNF-α) inhibitors and the risk of serious infection in Danish children with IBD.

- *Study III*: To screen for signals of previously unknown adverse events of TNF-α inhibitors in Danish pediatric patients with IBD or JIA by applying data mining methods to nationwide health care registers.

- *Study IV*: To systematically describe and compare alternative pharmacoepidemiologic designs, and to present a case example where the designs are applied in a real-world drug safety assessment to illustrate the differences.

- *Study V*: To investigate if there is an association between the use of TNF-α inhibitors and the risk of serious infection in patients with JIA.

We conducted analyses based on linked data from nationwide health care and administrative registers in Sweden and Denmark. We used a complete binational cohort of 21,000 pediatric IBD and JIA patients with an average follow-up of 4.4 years and individual data on demographics, diagnoses and procedures in specialist care, pharmaceutical drug use, and socioeconomic status of patients' parents.

In the next section of this thesis summary (section 2), we review the safety concerns in pediatric IBD and JIA that were investigated. In section 3, data sources, study designs, and statistical methods are described. Section 4 contains summaries of the background and key results of papers I-V. We discuss the clinical implications, methodological and ethical considerations, and a few points on the future of pediatric pharmacoepidemiology in section 5. Section 6 contains conclusions. In sections 7 and 8, popular science summaries are provided in English and Swedish, respectively. Finally, acknowledgments can be found in section 9 and references in section 10. At the end, papers I-V, including supplementary appendices, are attached.

2 BACKGROUND

In this section we describe the treatment patterns in pediatric IBD and JIA, and review the current literature on the safety concerns that were investigated in this thesis.

2.1 TREATMENT IN PEDIATRIC IBD AND JIA

IBD and JIA represent some of the most common serious health conditions in pediatric patients. IBD is characterized by chronic inflammation of the gastrointestinal tract and typically includes ulcerative colitis (UC), Crohn's disease (CD), and unclassified IBD. The incidence rate of pediatric IBD in Sweden has been estimated at 18.5 per 100,000 person-years,[18] but there is large geographic variation and incidence rates have increased in western countries in recent years.[19,20]

In UC, 5-aminosalicylic acid (5-ASA) is the recommended induction therapy for mild and moderate cases.[21] This treatment is also used as first-line maintenance therapy, although it is often replaced by azathioprine, which is more efficacious.[22] Azathioprine is also one of the recommended maintenance treatments in CD, with pediatric-specific efficacy evidence.[23-25] 5-ASA is also used in CD, but is controversial due to lack of evidence and is only recommended for use in very mild disease.[25]

JIA is a heterogeneous group of autoimmune diseases characterized by arthritis of unknown etiology with onset before the age of 16 years that persists for at least six weeks.[26] There are seven distinct types of JIA: systemic arthritis, oligoarthritis, polyarthritis (rheumatoid factor negative or positive), psoriatic arthritis, enthesitis-related arthritis, and undifferentiated arthritis.[26] Criteria for each diagnosis include the number of joints affected, symptoms and laboratory tests. A meta-analysis estimated the incidence of JIA at 8.2 per 100,000 person-years, standardized to the European population.[27]

In JIA, the treatment strategies vary depending on disease type, but general initial therapy includes nonsteroidal anti-inflammatory drugs (NSAID), which are used to manage symptoms and pain. In active disease, local glucocorticoid injections have been recommended regardless of concurrent therapy and across JIA types.[28] The most widely used disease-modifying drug in JIA is methotrexate (MTX),[29,30] which can be administered both orally and subcutaneously. It has proven efficacy since the early

1990s and is recommended in all subtypes of JIA, although its role in enthesitis-related arthritis is unclear.[31] It is generally recommended to continue MTX, at least initially, following start of treatment with TNF-α inhibitors.[28]

Biologic treatment, in particular TNF-α inhibitors, has revolutionized the treatment of inflammatory diseases and has become increasingly common in the treatment of both children and adults. TNF-α inhibitors have proven efficacy as induction and maintenance therapies in pediatric IBD. In patients with CD and UC, 88% and 73% respond to TNF-α inhibitor treatment while 59% and 29% are in remission after one year, respectively.[32-34] In JIA, efficacy has been established in multiple clinical trials.[35,36]

The recommendations are reflected by the treatment patterns observed in clinical practice. In Danish pediatric patients with disease onset 2007-2016, 5-ASA was more prevalent in UC than in CD patients (Figure 1). The proportions of patients who used 5-ASA in the first five years were 78% and 22% in UC and CD, respectively. Most UC patients, 72%, started 5-ASA treatment in the first year following disease onset. In contrast, thiopurines were more common in CD; where 69% in comparison with 48% in UC used the drug in the first five years. In the first year, 58% of CD patients started thiopurines. Among JIA patients, NSAIDs were used by 83% in the first five years, with 71% using them in the first year. MTX was used by 21% in the first five years. Use of TNF-α inhibitors gradually increased in prevalence over the 5 years following disease onset in all diseases. Their use was most prevalent in CD patients, where 47% used TNF-α inhibitors in the first 5 years, while the proportions in UC and JIA were 27% and 20%, respectively.

There is a general lack of pediatric-specific safety evidence for the most common treatments in pediatric IBD and JIA. Below we summarize the current literature on the safety concerns that were investigated in this project: the potential association between use of thiopurines and the risk of acute pancreatitis in pediatric IBD; and between use of TNF-α inhibitors in pediatric IBD, JIA and the risk of serious infections.

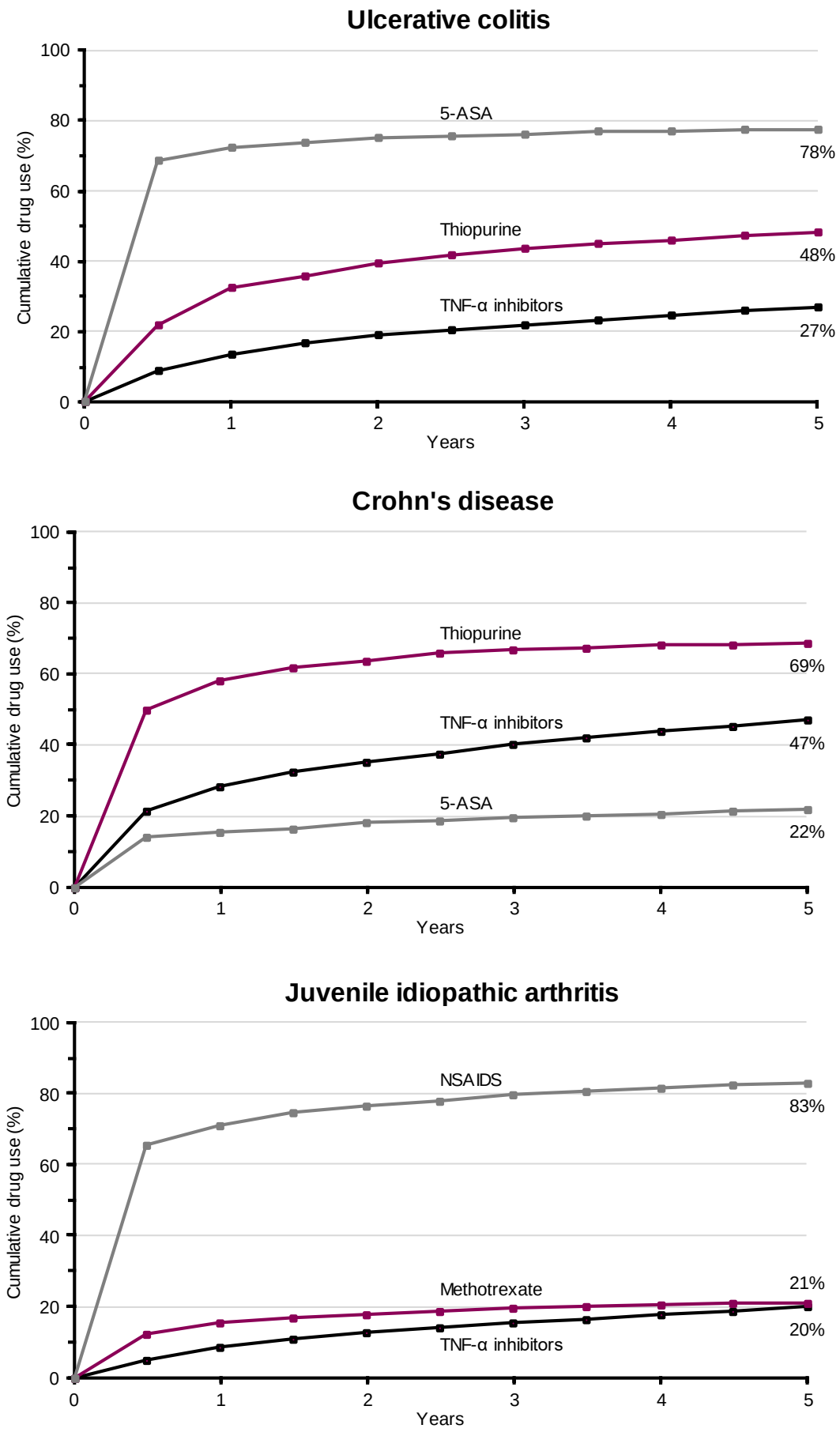


Figure 1. Cumulative drug use in Danish UC, CD and JIA patients during the first five years of disease among patients with disease onset 2007-2016

2.2 THIOPURINES AND THE RISK OF ACUTE PANCREATITIS IN PEDIATRIC IBD

Several RCTs have indicated that acute pancreatitis is an adverse effect of thiopurines In adult IBD, including azathioprine and 6-mercaptopurine, and that the risk is higher among CD patients compared to UC patients. A Cochrane review of RCTs (11 trials; total N=881) on use of azathioprine or 6-mercaptopurine in adult CD concluded that pancreatitis was among the most common adverse events that led to withdrawal of treatment,[24] although the relative risk of pancreatitis was not estimated. Some of the included RCTs showed large risk differences between azathioprine and comparators. In one RCT of 142 adult CD patients randomized to either early azathioprine treatment or conventional therapy, n=7/71 (10%) in the azathioprine group had pancreatitis events versus n=2/71 (3%) in the comparator group (no p-value presented).[37] In another trial, adult CD patients were randomized to either azathioprine or placebo and n=7/68 (10%) and n=0/63 (0%) (p-value 0.01) developed pancreatitis, respectively.[38]

A Cochrane review on adult UC and use of azathioprine and 6-mercaptopurine (6 trials; total n=279) showed a lower absolute risk of pancreatitis (2%) in patients treated with azathioprine compared with the studies in CD.[39] There was no statistically significant difference between treated and controls in the meta-analysis, although the analysis was based on very small numbers (n=3/141 [2%] versus n=0/138 [0%], risk ratio [RR] 4.13, 95% confidence interval [CI] 0.48 to 35.48).

Among observational studies in adult IBD, a recent German study prospectively followed 510 patients with IBD who initiated treatment with azathioprine.[40] There was a larger proportion of CD patients (8.6%) compared to UC patients (3.2%) who developed pancreatitis after azathioprine initiation. Azathioprine-associated pancreatitis occurred after a median of 21 days and less than half of the cases resulted in hospitalization (43%).

In pediatric IBD, only one RCT has been conducted on the use of thiopurines (CD; N=55).[23] In the trial, the combination of 6-mercaptopurine and prednisone was compared with placebo and prednisone and no patients developed symptoms of acute pancreatitis. A few observational studies on thiopurine use in pediatric IBD have been conducted but none of them have included comparator groups. A prospective register study from the United States on pediatric UC (N=197) showed that 2% (n=4) of the patients initiating treatment with either azathioprine or 6-mercaptopurine developed

pancreatitis during follow-up.[41] Other small case series have reported absolute risks of pancreatitis in the range of 1.1-6.4% among pediatric IBD patients initiating thiopurine treatment: n=1/88 (1.1%) (CD)[42], n=2/79 (2.5%) (UC/CD)[43], and n=6/93 (6.4%) (CD)[44].

In summary, safety evidence on the use of thiopurines that is specific to pediatric IBD patients is scarce and inconclusive. None of the studies in pediatric IBD described above have been specifically designed or powered to investigate if thiopurine use is associated with the risk of acute pancreatitis. Despite the general lack of safety evidence, thiopurine is commonly prescribed in pediatric IBD. In study I, we investigated the potential association between use of azathioprine, the most widely used thiopurine in Scandinavia, and the risk of acute pancreatitis in a nationwide Swedish and Danish pediatric IBD cohort.

2.3 TNF-ALPHA INHIBITORS AND THE RISK OF SERIOUS INFECTIONS IN PEDIATRIC IBD AND JIA

TNF-α inhibitors are generally considered safe but whether they increase the risk of infection in pediatric patients is controversial. In adults, a large Cochrane review from 2011 (160 RCTs) showed an increased risk of serious infections (defined in most studies as infections requiring hospitalization) in patients initiating biologics treatment as compared with controls (odds ratio [OR] 1.39, 95% CI 1.18 to 1.64).[45] The increased risk was similar when restricting the analysis to TNF-α inhibitors (116 RCTs; OR 1.41, 95% CI 1.13 to 1.75). RCTs on rheumatoid arthritis (RA) patients dominated the included studies (62 RCTs; OR 1.55, 95% CI 1.23 to 1.95). When the analysis was restricted to IBD patients (12 RCTs) there was no significant association, though the CI was wide (OR 1.28, 95% CI 0.67 to 2.44). In another study, a network meta-analysis of RCTs in RA, the results showed a dose-dependent increased risk of serious infections associated with the use of biologics: significantly higher risk in patients with high dose (OR 1.90, 95% credible interval 1.50 to 2.39) and standard dose (OR 1.31, 95% credible interval 1.09 to 1.58), but not in low-dose patients (OR 0.93, 95% credible interval 0.65 to 1.33).[46] A more recent meta-analysis of RCTs on TNF-α inhibitor use in adult IBD from 2016 (33 RCTs; total n=10,015) found no association between TNF-α inhibitors and increased risk of serious infection or death (OR 0.90, 95% CI 0.69 to 1.17). However, there was an increased risk of any infection, which also included infections in outpatient care (OR 1.21, 95% CI 1.10 to 1.33).[47]

The observational evidence in adult IBD and RA indicates an increased risk of serious infection. A large American study in RA based on claims data (N=31,801) found a significantly increased risk of serious infection associated with initiation of etanercept (hazard ratio [HR] 1.24, 95% CI 1.07 to 1.45) and infliximab (HR 1.39, 95% CI 1.21 to 1.60), as compared with abatacept.[48] Multiple observational studies in IBD have shown similar associations. A prospective study from 2012 (n=3420 exposed), a large retrospective study from 2018 (n=26,255 exposed), and a study of young adults (age 18-29; n=3574 exposed) showed significant associations between TNF-α inhibitor use and risk of serious infection in adult IBD (HRs: 1.43, 95% CI 1.11 to 1.84[49]; 1.71, 95% CI 1.56–1.88[50]; 1.49, 95% CI 1.12 to 1.98[51], respectively). A Swedish observational study in RA (n=4167 exposed) found a statistically significant association between use of TNF-α inhibitors and risk of serious infection during the first year of follow-up (HR 1.43, 95% CI 1.18 to 1.73) but not in year two (HR 1.15, 95% CI 0.88 to 1.51) or year three (HR 0.82, 95% CI 0.62 to 1.08).[52] Similarly, a Danish study in IBD patients (n=1543 exposed) found an increased risk of serious infection in TNF-α inhibitor users shortly following initiation (first 3 months: HR 1.63; 95% CI 1.01 to 2.63) and no significant association during the entire follow-up of 1 year (HR 1.27, 95% CI 0.92 to 1.75).[53] Finally, another large American study based on claims data found no significant associations between TNF-α inhibitor use and risk of serious infection in RA patients (N=10,484; HR 1.05, 95% CI 0.91 to 1.21) or IBD patients (N=2323; HR 1.10, 95% CI 0.83 to 1.46), as compared with non-biologic treatment.[54]

In pediatric patients, two prospective studies in JIA have investigated the association between use of TNF-α inhibitors and risk of serious infection. The first study found a significant association with use of etanercept (n=1414; RR 2.12, 95% CI 1.08 to 4.17) and not with adalimumab (n=320; RR 0.88, 95% CI 0.18 to 4.28).[55] The second study had lower power and found no significant association with etanercept (n=852), HR 1.36 (95% CI 0.60 to 3.07).[56] However, both studies reported higher ratios when investigating the risk of infection based on wider outcome definitions. A recent meta-analysis in JIA (n=1434 exposed) found no statistically significant association between TNF-α inhibitor use and risk of infections overall (OR 1.13, 95% CI 0.76 to 1.69).[57] The review did not investigate the risk of serious infections.

In pediatric IBD, a cohort study on TNF-α inhibitor use in young adults is the only study that presented pediatric-specific (age <18 years) results in a subgroup analysis.[51] The

analysis showed no significant association between TNF-α inhibitor use and risk of serious infections in children, HR 1.12 (95% CI 0.75 to 1.68). A meta-analysis from 2014 in pediatric IBD also showed no increased incidence rate of serious infections in TNF-α inhibitor users (standardized incidence ratio 1.06, 95% CI 0.83 to 1.36).[58] However, the results carry limited relevance due to inclusion of small case series, use of an unsuitable comparator group and lack of confounding control.

The safety evidence of TNF-α inhibitor in pediatric IBD and JIA is limited and concerns have been raised regarding the risk of serious infections. More information is needed to support clinical practice and decisions on optimal treatment strategy. We investigated if there is an association between use of TNF-α inhibitors and the risk of serious infection in Danish pediatric IBD patients in study II, and in JIA patients in study V.

3 MATERIALS AND METHODS

In this section, we summarize the data sources and methods used in studies I-V. The section includes definition of source populations, study populations, study designs and statistical methods. For additional information, see the individual papers and the supplementary material. The methods are discussed in more detail in section 5.2.

3.1 DATA SOURCES

All analyses in this project were based on data from Swedish and Danish health care and administrative registers with nationwide coverage. The general source population for the pediatric analyses (all studies except IV, discussed below) was children (age <18 years) who resided in Sweden (2005 to 2016) or Denmark (2000 to 2016). The population amounted to approximately 5.3 million individuals and was identified through national population registers (The Total Population Register in Sweden and The Danish Civil Registration System).[59-61] From the source population we extracted data on all patients with any diagnosis of chronic inflammatory disease during the general study periods or three years before. We identified 8700 unique patients with IBD diagnosis and 12,600 patients with JIA diagnosis. Patient-level, longitudinal data from multiple registers was linked via personal identification numbers and all data was anonymized before analysis.[62] An overview of the data sources can be found in Figure 2.

In study I, we utilized both Swedish and Danish data that were analyzed separately. Aggregated results were pooled for the main analysis. In studies II, III and V, the analysis was performed in Denmark where the coverage of hospital administered TNF-α inhibitors in the national patient register is close to complete.[63,64] Treatment with biologics in Denmark is administered in specialist care and does not incur any cost for the patient.

From the population registers and multi-generation registers we extracted data on demographics (date of birth, sex, migration, date of death) and linked individuals to their parents.[59-61,65] As the study population was pediatric education level and income of patients' parents or guardians (extracted from Longitudinal Integration Database for Health Insurance and Labour Market Studies [LISA] in Sweden and socioeconomic databases in Denmark) were used as proxies for socioeconomic status.

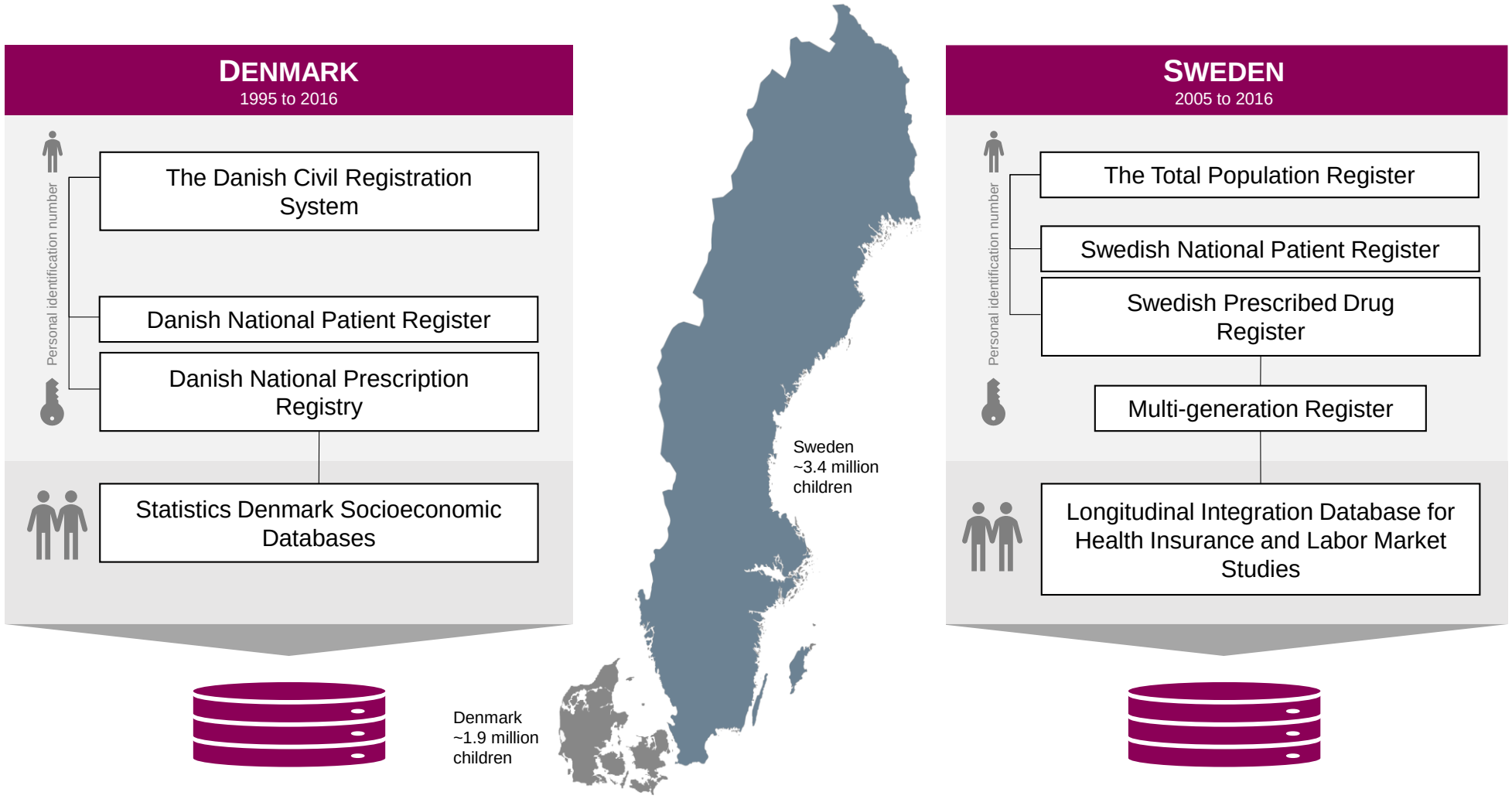


Figure 2. Overview of the data sources and linkages in Denmark and Sweden

The national patient registers in Denmark and Sweden, which include all physician-assigned diagnoses and performed procedures in secondary care (outpatient visits and inpatient admissions), were used to identify the cohort of patients with chronic inflammatory disease and for identifying comorbidities, disease history and outcomes.[35,66,67] The national drug registers, which cover all dispensings of prescription drugs including prescriptions originating from primary care, were used to identify exposure, co-medication and treatment history.[68-70]

A strength of the Danish and Swedish data sources is the nationwide coverage of diagnoses, procedures and drug use from routine clinical care. The general access to health care and subsidized use of prescription drugs in both countries enable population-wide coverage. The sample size and length of follow-up are comparably large for a pediatric cohort. The data sources also have limitations. The Swedish patient register does not have complete coverage of hospital administered drugs, while administration of biologics is captured in the Danish patient register, as described above. There is no structured data on dosing, which can hamper estimation of duration of prescription fills and complicate exposure definitions that are based on prescribed dose. Finally, there is no data on diagnoses from the primary care setting with national coverage.

In study IV, we performed an illustrative case example that was based on a cohort of adult type 2 diabetes patients, derived from the data sources described above (only Sweden). The source population for the analysis was all patients who had filled a prescription of a type 2 diabetes drug (Anatomical Therapeutic Chemical [ATC] A10) during the study period (July 2013 to December 2018). We identified 574,999 unique patients who were eligible at some time point during the study period. The case example was based on a previously published study[71] and was chosen based on the large sample and well-established safety concern (sodium glucose cotransporter 2 [SGLT2] inhibitors and the risk of ketoacidosis) to enable clear illustration of the differences between pharmacoepidemiologic designs.

3.2 STUDY DESIGN

3.2.1 Study populations

All studies used cohort designs; eligible individuals were identified at certain time points (index dates) and followed until event or censoring, whichever occurred first. Patients with confirmed disease were included in the study cohorts.

The pediatric IBD cohorts for studies I, II and III were identified based on data from the national patient registers. Patients with at least two contacts (inpatient or outpatient) with specialist care with a physician assigned IBD diagnosis during or before the study periods were included. In study I, the study period was July 2006-2016 in Sweden and 2000-2016 in Denmark, due to the earlier launch of the drug register in Denmark. The study period in study II, which was only based on Danish data, 2007-2016, started a few years after drug approval to avoid inclusion of early users. In study III, where we screened for signals of new adverse events, the study period started from the approval of TNF-α inhibitors in Denmark (2004-2016).

The Danish JIA cohort, analyzed in studies III and V, was also identified based on the national patient register. In study III, at least two JIA diagnoses from specialist care were required. In study V, the study period was the same as in study II, 2007-2016, and at least two JIA diagnoses in specialist care were required, where the first diagnosis was recorded at age 16 or younger.

In the case example of study IV, we identified type 2 diabetes patients based on drug use rather than diagnosis. At least two filled prescriptions of any diabetes drug (ATC A10) during the study period (July 2013 to December 2018) were required. Using indication-specific drugs for cohort definition had the advantage of also capturing patients who were diagnosed and treated in the primary care setting.

3.2.2 Exposures and comparators

Drug use was identified based on filled prescriptions in the Danish National Prescription Registry and the Swedish Prescribed Drug Register. Records of Hospital administered treatments in Denmark were obtained from the Danish National Patient Register.

In studies I-III, we used no-use episode designs where we identified episodes of new use (no previous use during a fixed look-back period) of the study drug and episodes of no use of that drug for each individual in the study population.[72,73] The study drugs were azathioprine in study I and TNF-α inhibitors in studies II and III. All eligible follow-up, post confirmed disease, contributed to the episodes. Hence, all time with neither current nor recent use of the study drug was divided into mutually exclusive no-use episodes, which made up the comparator. Maximum episode length was one year in study I, and three years in studies II and III (Figure 3). In this design, one patient could contribute to both exposed and no-use episodes. However, because there was no overlap between the episodes and patients with a history of the outcome were excluded, no individual patient could contribute with multiple events.

In study V, we used a modified (or generalized) prevalent new user design[74] where initiators of TNF-α inhibitors were compared with incident and prevalent users of an active comparator, MTX. With this design we were able to use an active comparator, while not excluding initiators of the study drug, TNF-α inhibitors, who were prevalent in the comparator. Further, in relation to the standard prevalent new user design, we did not use time-dependent propensity score (PS) matching, which meant that all exposed observations were retained in the analysis. To form this analytical cohort, we pooled a large set of sequential cohorts identified during the study period.[75,76] The study period was divided into short intervals and one sequential cohort was identified for each interval. Individuals from the source population (described above) who met the eligibility criteria during a given interval were included in that sequential cohort.

In study IV, we sought to demonstrate how different pharmacoepidemiologic designs, including no-use episodes and generalized prevalent new user, can be defined based on the target trial emulation framework and sequential cohorts. In target trial emulation, an observational study is designed by emulating a hypothetical clinical trial, element by element. By being explicit about the emulation, the design of the observational study is transparent and potential biases can be addressed (more information in section 5.2.2).[76] In the case example of study IV, we reanalyzed a previously published drug safety assessment in adults with type 2 diabetes: the association between use of SGLT2 inhibitors and the risk of diabetic ketoacidosis.[71]

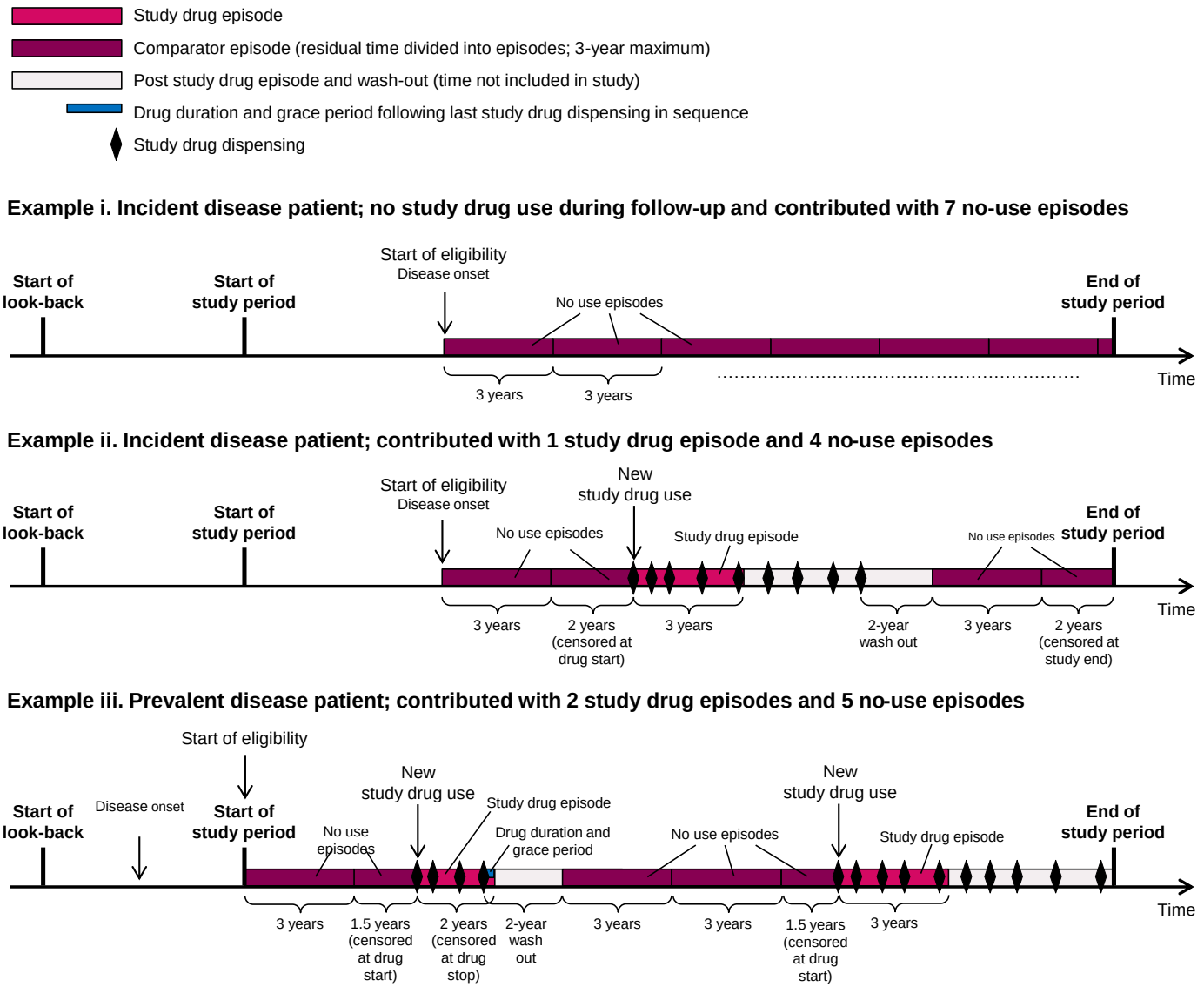


Figure 3. Examples of episodes of new study drug use and no-use episodes with an as-treated design and a maximum follow-up of three years and wash-out of two years

3.2.3 Eligibility and censoring

The general exclusion criteria that were applied to all analyses at baseline (based on fixed look-back periods) were: previous use of the study drug, history of the outcome event (to only study incident events), and residing outside of the country (to make sure that covariate status and data related to exclusion were updated and had been collected equally between patients). In all studies (except the case example of study IV) patients of age ≥18 years at baseline were excluded and patients with no recent hospital contact with a diagnosis for the underlying disease (IBD or JIA) were excluded to avoid including patients without regular contact with health care or with very mild disease. Additionally, in studies II and V, we also excluded patients with history of immunodeficiency, previous use of biologics, or diseases that might require immunosuppressing treatments (e.g. HIV and cancer) to limit the risk of confounding. Due to the non-targeted nature of study III, where we performed data mining for new signals of adverse events, the exclusion criteria were less restrictive.

We used as-treated designs in the main analyses of all studies, i.e. patients were followed as long as they adhered to their baseline treatment strategy. In study I, the maximum length of follow-up in the main analysis was short (90 days) and azathioprine episodes were not censored due to treatment changes, while no-use episodes were censored at initiation of azathioprine, if any. In the as-treated analyses of TNF-α inhibitor use in studies II, III and V, patients were censored at treatment cessation in the TNF-α inhibitor group and initiation of a TNF-α inhibitor in the comparator group, if any. The duration of TNF-α inhibitor use was based on the dosing schedule of treatment guidelines and a grace period of 60 days in studies II and V, and 90 days in study III. Additionally, in study V the duration of subcutaneous TNF-α inhibitors was set to 60 days to account for potential dispensing of biologics for self-administration. In the case example of study IV, we assumed the same durations per prescription fill as in the original study.[71] Other reasons for censoring were end of study period, maximum follow-up (e.g. maximum episode length), migration from Sweden or Denmark, or death.

3.2.4 Outcomes

All analyses were targeted at specific outcome events except study III, where we screened for any event with elevated risk. Outcomes were identified based on

physician-assigned diagnoses in specialist care, derived from the national patient registers. The primary outcome in study I was acute pancreatitis, defined as a contact with specialist care (inpatient or emergency outpatient) with a diagnosis of acute pancreatitis. The primary outcome in studies II and V was serious infection, defined as an inpatient hospital admission with a physician-assigned diagnosis for any infection. In study IV, the outcome of the case example was diabetic ketoacidosis, defined as any contact with specialist care (inpatient or outpatient) with a diagnosis of diabetic ketoacidosis. The date of health care contact or admission was the date of event.

3.3 STATISTICAL ANALYSES

3.3.1 Confounding adjustment

We used propensity score (PS) methods to adjust for confounding in all studies. The PS, which is the conditional probability of treatment, was estimated with logistic regression. In the PS models we included general socio-demographic factors, such as age, sex, socioeconomic status (from parents in the pediatric studies), and measures of health care use. Additionally, we adjusted for disease history, treatment history, and factors related to the severity of the underlying disease, that were potential confounders. Covariate balance at baseline following adjustment was assessed using standardized mean differences, where a difference below 0.1 was considered consistent with well-balanced groups.

In studies I and III, we used 1:1 PS matching with a greedy nearest neighbor algorithm. The caliper (maximum difference in PS between exposed and comparator) was 20% of the pooled standard deviation of the logit PS in study I,[77] and 200% of the same in study III to ensure that all exposed episodes were included in the matched cohort. Additionally, in study III, we required an exact match on the underlying disease of JIA, CD or UC.

In studies II, IV and V, we used different types of PS weighting to adjust for confounding. In studies II and IV, we used standardized mortality ratio (SMR) weights and fine-stratification weights, respectively, to estimate the average treatment effect in the treated (ATT) (Table 1). Only comparator observations were weighted to achieve a similar distribution of covariates as in the exposed. In study V, we used stabilized inverse probability of treatment (IPT) weighting, to estimate the average treatment

effect (ATE) or the marginal effect, in the TNF-α inhibitor initiators and MTX users. In all weighted analyses, observations with PS outside the common range were excluded from the analysis.

In all studies, except the case example of study IV, we performed a priori defined sensitivity analyses to investigate the robustness of the main results, and subgroup analyses to examine effect modification between patient groups.

Table 1. Propensity score weighting methods

Weighting method	Formula	Estimand	Application
Standardized mortality ratio (SMR)	$w_i = \begin{cases} 1, & if\ A_i = 1 \\ \frac{e_i}{1-e_i}, & if\ A_i = 0 \end{cases}$	ATT	Study II
Fine-stratification	$w_i = \begin{cases} 1, & if\ A_i = 1 \\ \frac{N_{A=1;j}\ /N_{A=1}}{N_{A=0;j}/N_{A=0}}, & if\ A_i = 0 \end{cases}$	ATT	Study IV
Stabilized inverse-probability of treatment (IPT)	$w_i = \begin{cases} \frac{e}{e_i}, & if\ A_i = 1 \\ \frac{1-e}{1-e_i}, & if\ A_i = 0 \end{cases}$	ATE	Study V

Note: e_i PS in observation i; e marginal PS; w_i PS weight in observation i; A_i treatment in observation i (1 exposed; 0 comparator); N_A total number of observations with treatment A; $N_{A;j}$ number of observations with treatment A in PS stratum j.

3.3.2 Informative censoring adjustment

We performed as-treated analyses in all studies and patients were censored at deviation from the baseline treatment. To adjust for potential informative censoring, i.e. differential censoring in relation to the prevalence of risk factors for the outcome, we used stabilized inverse probability of censoring (IPC) weighting in studies IV and V (Formula 1).[78] Weights were calculated for a certain patient and time interval of follow-up as the inverse of the conditional (on baseline treatment and confounders) probability of *not* being censored in the previous interval.

$$w_{i,t} = \prod_{i,t=1}^{i,t} \frac{P(C_{i,t} = 0 | C_{i,1} = 0, \dots, C_{i,t-1} = 0, A_i)}{P(C_{i,t} = 0 | C_{i,1} = 0, \dots, C_{i,t-1} = 0, A_i, X_{i,t})}$$

[Formula 1] IPC weights. $w_{i,t}$ stabilized censoring weight for observation i at time t; $C_{i,t}$ censoring status for observation i at time t (1 censored; 0 not censored); A_i baseline treatment for observation i; $X_{i,t}$ vector of baseline and time updated confounders for observation i at time t.

Weights were stabilized by inserting the probability of not being censored conditioned on baseline treatment in the numerator. The conditional probability of censoring was estimated with logistic regression. Final weights used in the analysis were calculated as the product of baseline IPT weights and IPC weights until the time interval analyzed. The weights were truncated at the 1st and 99th percentiles to avoid adjusting for extreme weights. In studies I-III, we performed naïve analyses with no adjustment for potential informative censoring.

3.3.3 Effect estimation

In study I, we used Poisson regression with offset for patient-time to be able to pool aggregate results in subgroups with zero events from the country analyses. With Poisson regression we estimated the incidence rate ratios (IRR) of the outcome associated with exposure. In study II, we used Cox proportional hazards regression to estimate HRs. Robust sandwich estimators were used to account for repeated observations in the weighted pseudo population. We assessed the proportional hazards assumption by testing if an interaction term between exposure and time was significant. In studies IV and V, we estimated the rate ratio of the outcome associated with exposure in weighted or matched pooled logistic regression models, where all sequential cohorts and follow-up intervals were included.[79] The only covariates in the outcome models were baseline exposure and time interval (including polynomials) and we accounted for repeated and dependent outcome events within individuals. IRRs and HRs with 95% CIs not including 1.0 were regarded as statistically significant. In study I, we additionally estimated the absolute rate differences with the formula (IRR-1)*comparator incidence rate,[73] while we used Poisson regression with an identity link in study II. Crude and

adjusted (matched or weighted) cumulative incidence curves were estimated with the complement of the Kaplan-Meier function.

3.3.4 Data mining with scan statistics

In study III, we screened for new signals of adverse events of TNF-α inhibitors in children with IBD or JIA in Denmark. We used physician-assigned diagnosis codes from specialist care (primary and secondary diagnoses; outpatient and inpatient contacts). All diagnoses observed during follow-up, at five levels of the ICD-10 code tree (cuts), were considered as potential adverse events, i.e. from disease chapters (e.g. I00-99 diseases of the circulatory system) to four-position codes (e.g. I47.1 supraventricular tachycardia) (Figure 4). Events were collected from the register at the three- and four-position levels across the entire ICD-10 code system. A limited set of diagnoses were not assessed since they were considered not relevant as potential adverse events (e.g. congenital malformations).

Two analyses were performed. In the first, we used PS matched tree-based scan statistics to compare episodes of TNF-α inhibitor use with episodes of no use. In the second, we performed a self-controlled analysis using tree-temporal scan statistics to compare temporal risk windows within the TNF-α inhibitor episodes. We only analyzed incident events, defined as a code not preceded by the same code on the three-position level. The look-back window in the PS matched analysis was infinite and in the self-controlled analysis the look-back window was three years in relation to the time of the event.

In the PS matched analysis, we used unconditional Bernoulli tree-based scan statistics.[80,81] The exposure of each observed event was assumed to follow a Bernoulli distribution. The null hypothesis was that events in all cuts were equally probable (due to the 1:1 matching) to occur in the TNF-α inhibitor as the no-use episodes, while the alternative hypothesis was that the risk of events to occur in the TNF-α inhibitor episodes was higher, for at least one cut. Find a summary of formulas in Table 2.

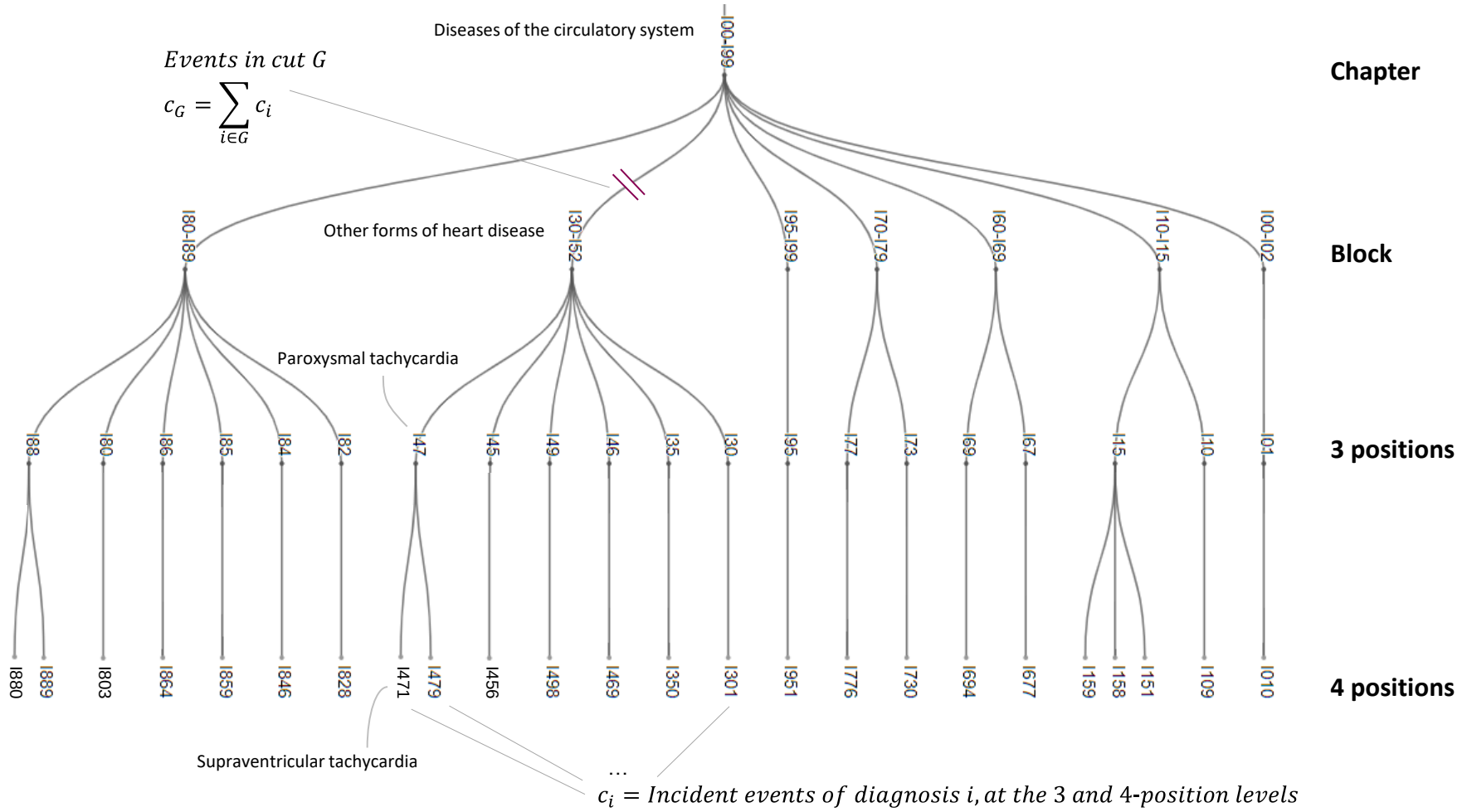


Figure 4. Example of tree-based structure of ICD-10 codes (I00-I99) from Chapter to four-position level. Number of events in cut G, c_G, is calculated as the sum of incident diagnoses at the three and four-position levels below the cut, c_i (one patient cannot contribute with more than one event to each cut)

In the self-controlled analysis, we used a tree-temporal scan statistic, conditioned on the number of events in each cut and assumed that the events were uniformly distributed over follow-up under the null hypothesis. The alternative hypothesis was that the risk was higher in at least one combination of cut and investigated risk window; the log likelihood ratio (LLR) was estimated for each combination. The durations of risk windows were 2 days to 1.5 years (half of the maximum follow-up) and no window was shorter than 20% of the follow-up day it ended on (e.g. windows that ended on day 50 were 10 days or longer).

For each cut in the PS matched analysis and for each cut-risk window in the self-controlled analysis the LLR was calculated. Inference was based on Monte Carlo simulation because there is no simple expression for the sample distribution of the LLRs. See details on the procedure in the box below. Cuts with p-values below 0.05 were considered statistically significant.

3.3.5 Statistical software

The following software packages and applications were used to perform the statistical analyses for studies I-V SAS v9.4 (SAS Institute Inc.), TreeScan v1.4 (www.treescan.org), and D3.js (v5.14.0).

3.3.6 Ethical approval

Studies I, II and IV, conducted based on Swedish register data, were approved by the regional ethics committee in Stockholm (Ref 2016/2029-31/1; 2017/715-31). Ethical approval was not required for studies performed based on Danish register data; those studies were approved by the Danish Data Protection Agency.

Table 2. Definitions and formulas for scan statistics analyses

Item	Tree-Based Unconditional Bernoulli Scan Statistic	Tree-Temporal Conditional Scan Statistic
Expected exposed events under the null hypothesis	$p(c_i + n_i)$ p, proportion of exposed in matching cluster (0.5 if 1:1 matching); c_i, exposed events in node i; n_i, unexposed events in node i.	$k_i \frac{w_j}{T}$ k_i, total exposed events over follow-up in node i; w_j, length of risk window j; T, total length of follow-up.
Events in cut G	$c_G = \sum_{i \in G}(c_i) \quad n_G = \sum_{i \in G}(n_i)$ G, cut; c_G, exposed events in cut G; n_G, unexposed events in cut G.	$c_{G,j} = \sum_{i \in G}(c_{i,j}) \quad k_G = \sum_{i \in G}(k_i)$ G, cut; $c_{G,j}$, exposed events in cut G and risk window j; $c_{i,j}$, exposed events in node i and risk window j; k_G, total exposed events over follow-up in cut G; k_i, total exposed events over follow-up in node i.
Hypotheses	H_0: $\frac{c_G}{c_G + n_G} = p \; for \; all \; cuts \; (G)$ H_1: $\frac{c_G}{c_G + n_G} > p \; for \; at \; least \; one \; cut \; (G)$	H_0: $\frac{c_{G,j}}{k_G} = \frac{w_j}{T} \; for \; all \; cut \; and \; risk \; windows \; (G,j)$ H_1: $\frac{c_{G,j}}{k_G} > \frac{w_j}{T} for \; at \; least \; one \; cut \; and \; risk \; window \; (G,j)$
Log likelihood ratio	$LLR(G) = ln\left(\frac{\left(\frac{c_G}{c_G+n_G}\right)^{c_G}\left(\frac{n_G}{c_G+n_G}\right)^{n_G}}{(p)^{c_G}(1-p)^{n_G}}\right) I\left(\frac{c_G}{c_G+n_G} > p\right)$	$LLR(G,j) = ln\left(\frac{\left(\frac{c_{G,j}}{k_G}\right)^{c_{G,j}}\left(\frac{k_G - c_{G,j}}{k_G}\right)^{k_G - c_{G,j}}}{\left(\frac{w}{T}\right)^{c_{G,j}}\left(\frac{T-w}{T}\right)^{k_G - c_{G,j}}}\right) I\left(\frac{c_{G,j}}{k_G} > \frac{w}{T}\right)$
	I, indicator function which takes value 1 if true, otherwise 0.	
Test statistic	$T = \max_G LLR(G)$	$T = \max_{G,j} LLR(G,j)$

Monte Carlo hypothesis testing

Monte Carlo simulations were performed for the PS matched analysis and the self-controlled analysis, through the following steps (additional definitions in Table 2):

1) A series of random replicas of the real datasets (9999 iterations) were generated under the null hypothesis:
    a. In the *PS matched analysis*, the exposed events (c_i) and unexposed events (n_i) in each leaf were replaced by random values drawn from a Bernoulli distribution with p=0.5, where the total events (c_i+n_i) was fixed. If multiple events occurred in an episode, the events were randomized together to either the TNF-α inhibitor or no-use group. The number of events in each cut (c_G and n_G) was calculated in the real and simulated datasets. One episode could not contribute with more than one event to each cut.
    b. In the *self-controlled analysis*, the day of event for each exposed event was replaced by a random value drawn from a uniform distribution over the patient's total follow-up (events had equal probability of occurring at all days during follow-up). The number of events in each cut-risk window ($c_{G,j}$) was calculated in the real and simulated datasets.
2) The test statistics, *T* (max LLR), were calculated for each random and real dataset.
3) The test statistics from the real data (LLR for the most likely cut from each analysis) were ranked in relation to their corresponding series of simulated test statistics (1 vs 9999 values per analysis). The p-value was estimated based on the rank of the observed test statistic and the number of iterations:

$$pvalue = \frac{r}{S+1}$$

Where r, rank of observed, real test statistic; S, number of iterations. In addition to the most likely cut and cut-risk window, we also estimated p-values for second most likely and so on, including all with p-values < 1.

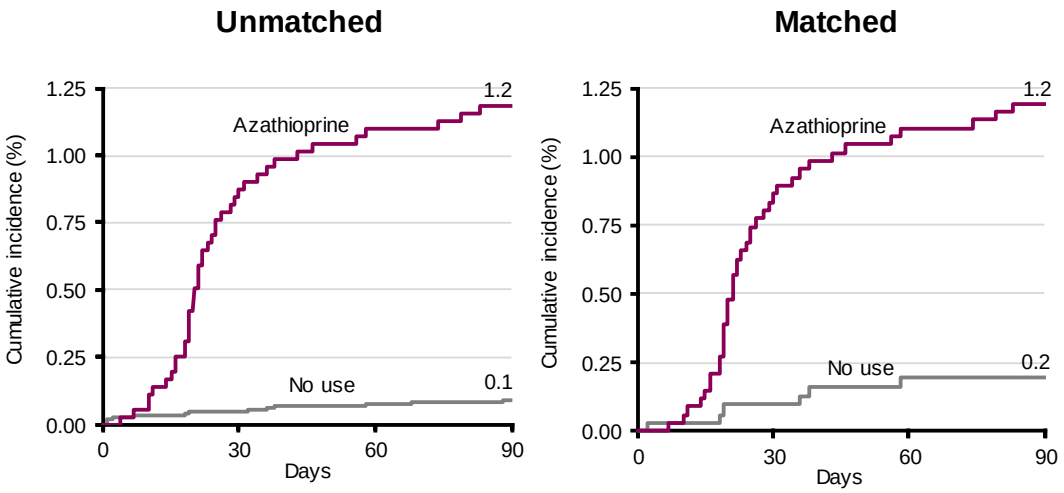
4 SUMMARY OF PAPERS

In this section, we summarize the background and key results of studies I-V. More detailed descriptions can be found in attached papers and supplementary appendices.

4.1 STUDY I: AZATHIOPRINE AND THE RISK OF ACUTE PANCREATITIS IN PEDIATRIC IBD

4.1.1 Background

Previous studies in adult IBD have indicated that use of thiopurines increase the risk of acute pancreatitis. According to studies, between 3 and 7% of patients experience the event in the first few months and risk increases up to 8 times have been observed.[40,82-84] In children the evidence is limited to a few case series and no controlled studies have been reported.[41-44,85] The aim of study I was to investigate if there is an association between use of azathioprine, the most commonly used thiopurine in Scandinavia, and the risk of acute pancreatitis in Swedish and Danish children with IBD.



Number at risk

Azathioprine	3574	3514	3491	3450	3374	3318	3295	3265
No use	18700	17967	17510	16702	3374	3091	2900	2900

Figure 5. Study I: Cumulative incidence of acute pancreatitis in the first 90 days in unmatched and PS matched cohorts of azathioprine and no use episodes

4.1.2 Key results

Based on Swedish and Danish nationwide data, we identified 3374 episodes of new azathioprine use among children with IBD, which were PS matched with episodes of no use. In these pairs, the mean (standard deviation [SD]) age was 14.3 (3.2) years, 55% were male, 57% had CD and 43% had UC or unclassified IBD. During the first 90 days of follow-up, 40 events of acute pancreatitis occurred among the azathioprine treated (incidence rate [IR] 49.1 events per 1000 patient-years) and 6 events occurred in the no-use group (IR 8.4 events per 1000 patient-years) (Figure 5). Use of azathioprine was significantly associated with an increased risk of acute pancreatitis; the IRR was 5.82 (95% CI 2.47 to 13.72). The absolute rate difference was 1.0 (95% CI 0.3 to 2.6) events per 100 patients during the 90-day risk period. In the secondary risk period, days 91-365 following azathioprine initiation, there was no significantly increased risk (IRR 0.99, 95% CI 0.31 to 3.11). The risk of acute pancreatitis appeared to be similar between subgroups, although the results were uncertain due to few no-use events.

4.2 STUDY II: TNF-ALPHA INHIBITORS AND THE RISK OF SERIOUS INFECTION IN PEDIATRIC IBD

4.2.1 Background

Previous observational studies have shown an association between use of TNF-α inhibitors and increased risk of serious infection in adult IBD; commonly defining serious infections as infections requiring hospitalization. Larger studies from various settings, including one prospective study and two retrospective studies, showed significant associations, HRs ranging between 1.43 and 1.71.[4-6] The only controlled study that has presented results for pediatric patients (age <18 years) to our knowledge, found a non-significant association between TNF-α inhibitor use and the risk of serious infection (HR 1.12, 95% CI 0.75 to 1.68) based on insurance claims data from the United States.[4] Hence, more data is needed to support the understanding of this drug safety concern in children. The aim of study II was to investigate if there is an association between use of TNF-α inhibitors and the risk of serious infection in Danish children with IBD.

4.2.2 Key results

Based on Danish nationwide data, we identified 618 episodes of new TNF-α inhibitor use and 2925 no-use episodes, among children with IBD. In the PS weighted cohort, 53% were male, mean (SD) age was 15.1 (1.7) years, 70% had CD and 30% had UC or unclassified IBD. The most commonly initiated TNF-α inhibitor was infliximab (95% of episodes) and the median follow-up time was 1.0 years among TNF-α inhibitor episodes and 2.1 years among no-use episodes. During follow-up, in the unweighted episodes of current TNF-α inhibitor use and no use there were 41 and 262 serious infection events, respectively (Figure 6). This translated to incidence rates of 54.6 and 61.9 events per 1000 patient-years among TNF-α inhibitor and no-use episodes, respectively, in the PS weighted cohort. There was no significant association between use of TNF-α inhibitors and the risk of serious infection, HR 0.81 (95% CI 0.54 to 1.21), and the absolute rate difference was -12.0 (95% CI -28.6 to 13.0) events per 1000 patient-years. Only considering the first 90 days of follow-up the weighted HR was similar, 0.76 (95% CI 0.35 to 1.66). Additionally, a similar result was observed in a replication of the analysis in a small Swedish cohort, weighted HR 0.72 (95% CI 0.28 to 1.83).

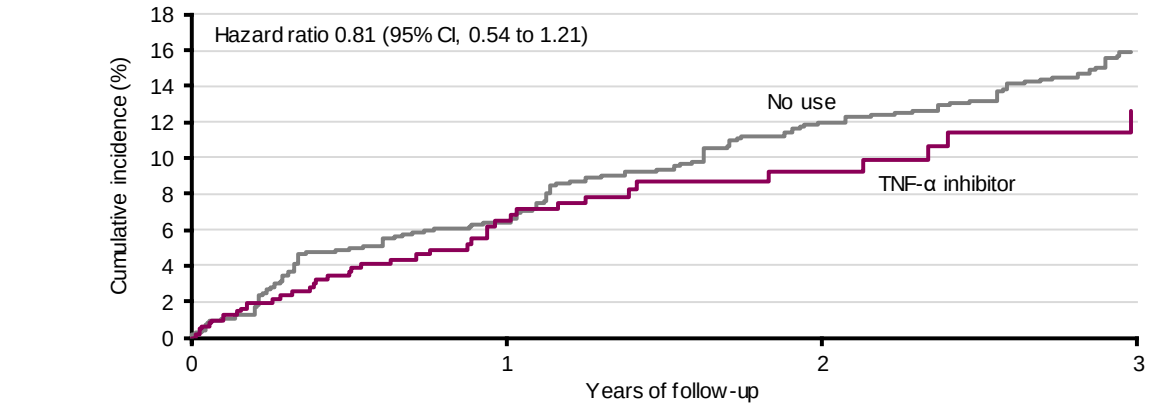


Figure 6. Study II: Cumulative incidence of serious infection in weighted cohort of TNF-α inhibitor and no-use episodes

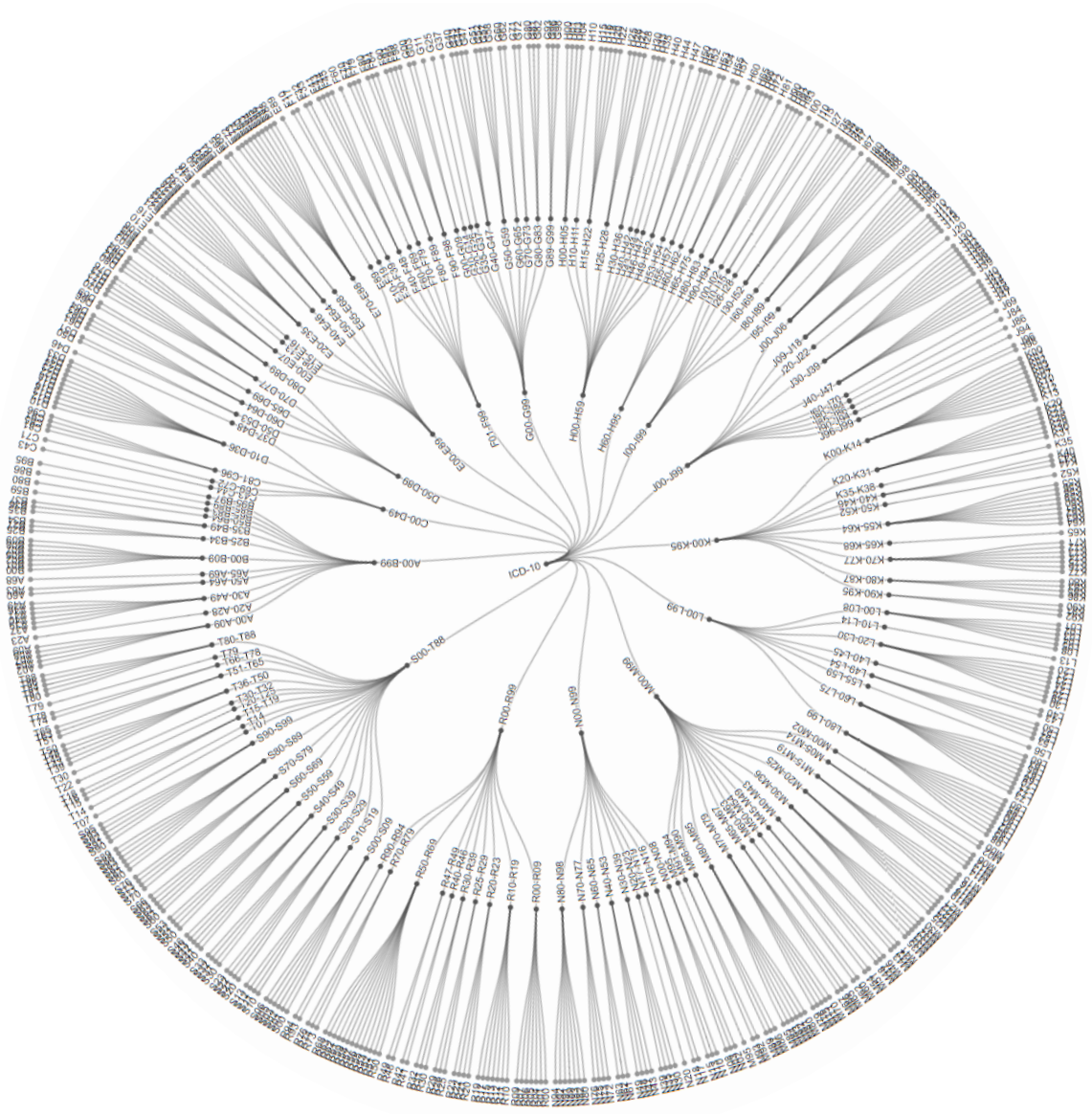
4.3 STUDY III: DATA MINING FOR ADVERSE EVENTS OF TUMOR NECROSIS FACTOR-ALPHA INHIBITORS IN PEDIATRIC PATIENTS

4.3.1 Background

Although TNF-α inhibitors are efficacious and considered safe in adults,[32-34] the pediatric-specific safety data is generally scarce. Previously unknown adverse events can be detected post-market approval when drugs are used by a more heterogenous and larger set of patients in clinical practice. Traditionally, spontaneous reporting systems have been the main source for signal detection. The use of routinely-collected data from health registers is another opportunity, which enables data mining at a large scale with potentially lower risk of reporting bias and confounding. The aim of study III was to screen for signals of previously unknown adverse events of TNF-α inhibitors in Danish pediatric patients with IBD or JIA, applying data mining methods to nationwide health care registers.

4.3.2 Key results

Based on Danish nationwide data, we identified 1310 episodes of new TNF-α inhibitor use in pediatric IBD and JIA patients. In a PS matched tree-based scan statistics analysis with episodes of no use as comparator, we detected two signals of adverse events of TNF-α inhibitors: dermatologic complications (ICD-10: L00-L99, 87 Vs 44 events, risk difference [RD] 3.3%) and psychiatric diagnosis adjustment disorders (ICD-10: F432, 33 Vs 7 events, RD 2.0%) (Table 3). The former events have been described previously in adults and children, while the latter was likely associated with the underlying diseases and their severity, rather than with the treatment. We also performed a self-controlled scan statistics analysis that generated no signals. Hence, no signals of previously unknown adverse events of TNF-α inhibitors in pediatric patients were detected. The analysis showed how Scandinavian health care registers and novel data mining methods can be used to screen for previously unknown adverse events. This type of evidence can play a particularly important role in pediatrics where output of both clinical and observational studies is low.



Cut (ICD-10 code)	TNF-α inhibitor events	No use events	Relative risk	Risk difference (%)	P-value
F432 Adjustment disorders	33	7	4.71	2.0	0.002
L20-L30 Dermatitis and eczema	34	8	4.25	2.0	0.004
F40-F48 Anxiety, dissociative, Stress-related, somatoform, etc.	39	11	3.55	2.1	0.007
F43 Reaction to severe stress, and adjustment disorders	35	9	3.89	2.0	0.008
L00-L99 Diseases of the skin and subcutaneous tissue	87	44	1.98	3.3	0.017

Table 3. Study III, PS matched analysis: Plot; dendrogram on all cuts of the ICD-10 tree with at least one event in the TNF-α inhibitor episodes or the no-use episodes (down to the three-position level). Table; cuts of the ICD-10 tree with significantly high risk in TNF-α inhibitor episodes.

4.4 STUDY IV: SELECTION OF COMPARATOR GROUP IN OBSERVATIONAL DRUG SAFETY STUDIES

4.4.1 Background

The comparator group is a key element of the design in pharmacoepidemiologic studies.[86] The active comparator new user (ACNU) design is a commonly used design where the comparator consists of patients initiating another drug at baseline.[87,88] This design has high potential to reduce various types of bias, but it also has limitations, including the requirement of a suitable comparator drug and strict eligibility criteria. In this study we explored and evaluated the following alternative designs that can be used when ACNU is not optimal: traditional no use, no use episodes, prevalent new user, generalized prevalent new user, and hierarchical prevalent new user. We used target trial emulation as a mutual framework to facilitate comparison of the designs. The specific aims of study IV were to systematically describe and compare alternative pharmacoepidemiologic designs, and to present a case example where the designs are applied in a real-world drug safety assessment to illustrate the differences.

4.4.2 Key results

In this study, we showed how the target trial emulation framework and sequential cohorts can be used to transparently communicate and compare various study designs in pharmacoepidemiology: the key difference between the designs is the eligibility criteria at baseline (Figure 7). From scrutinizing the differences and applying the designs in a case example, we concluded that many study-specific factors influence the selection of optimal comparator, including indication, available comparator drugs, treatment patterns, potential effect modification, and sample size. The ACNU is superior in its potential to reduce confounding and information bias, but if the strict eligibility criteria impair generalizability or statistical precision, a prevalent new user design might be preferable. If there is no suitable active comparator drug available a no use design can be considered. Irrespectively of the chosen design, the risk of bias needs to be critically assessed in each study.

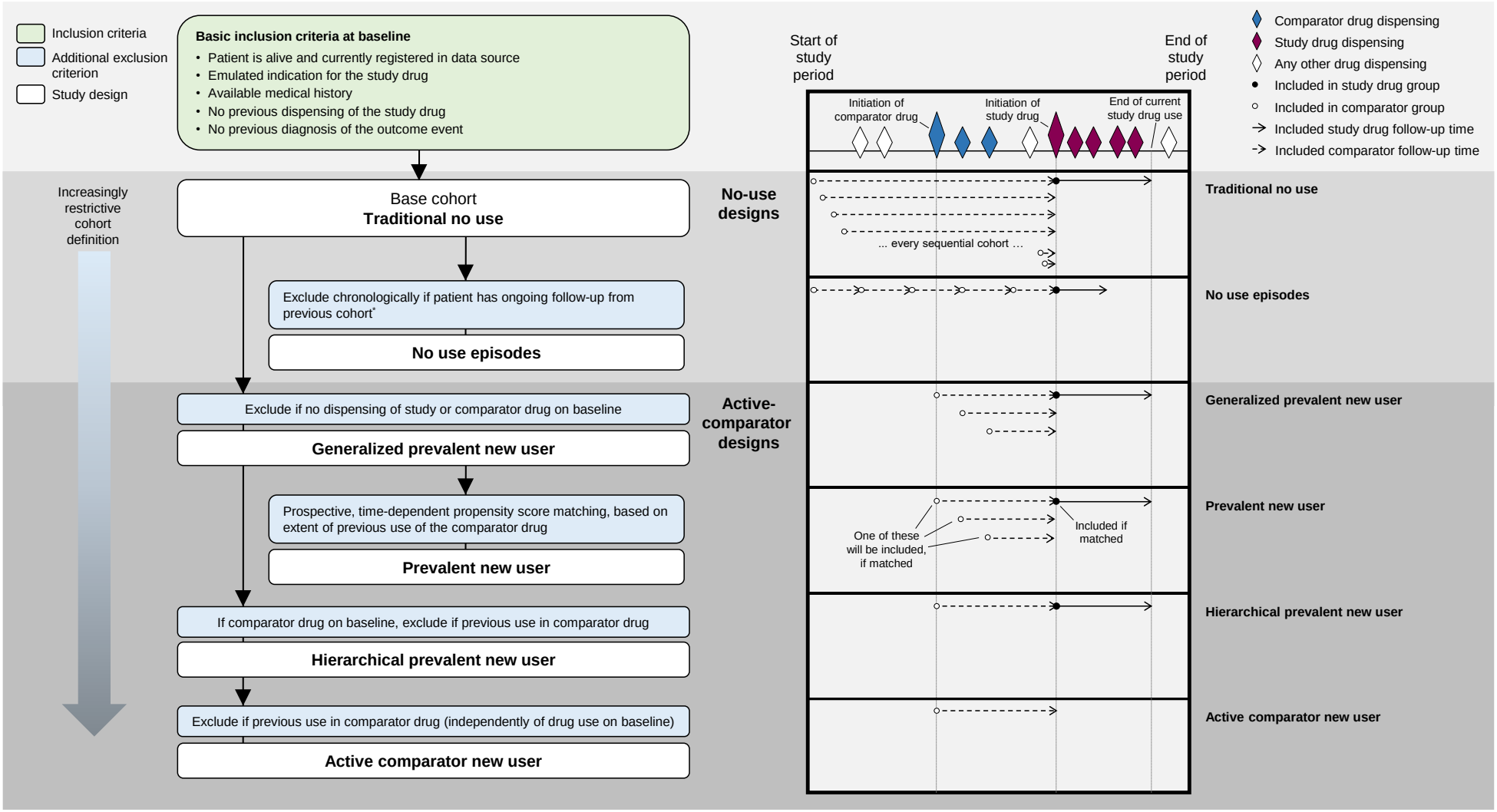


Figure 7. Study IV: Flow chart for identification of eligible patients (left) and eligibility for patient with sequential use of comparator and study drug (right), by alternative study designs. *Exclusion is typically preceded by applying a maximum follow-up, i.e. the episode length.

4.5 STUDY V: TNF-ALPHA INHIBITORS AND THE RISK OF SERIOUS INFECTIONS IN JIA

4.5.1 Background

Previous studies have shown that serious infection is an adverse event of TNF-α inhibitors in adults with rheumatic disease. A meta-analysis of RCTs on biologics found significantly increased risks, separately restricted to TNF-α inhibitors (116 RCTs) and RA patients (62 RCTs).[45] In JIA, there are primarily two previous studies that investigated this safety concern. Both were prospective, observational and analyzed use of etanercept: one found a significant association (n=1414; RR 2.12, 95% CI 1.08 to 4.17)[55] and the other reported an HR of 1.36 with a rather wide CI (n=852; 95% CI 0.60 to 3.07).[56] Hence, the pediatric-specific safety evidence is limited. The aim of study V was to investigate if there is an association between the use of TNF-α inhibitors and the risk of serious infection in patients with JIA.

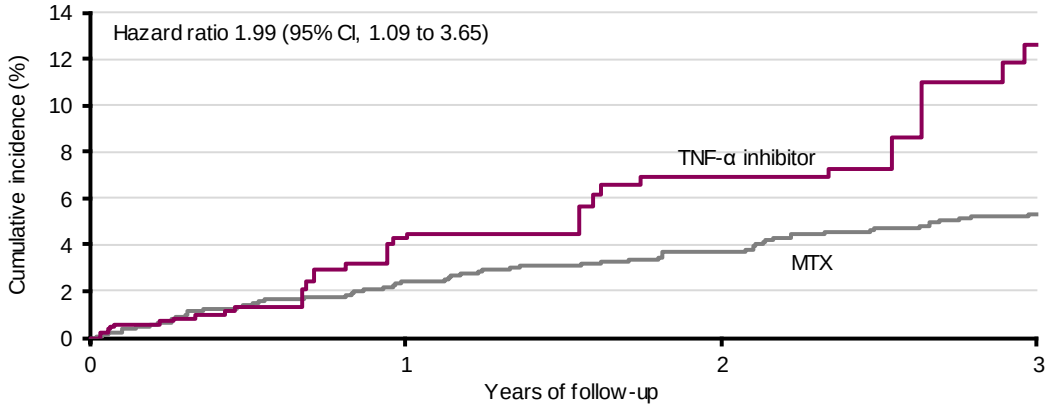


Figure 8. Study V: Cumulative incidence of serious infection in weighted cohort of TNF-α inhibitor and MTX users

4.5.2 Key results

Based on Danish nationwide data, among patients with confirmed JIA we identified 578 initiators of TNF-α inhibitors who met the eligibility criteria. The comparator consisted of 1915 observations of initiators and users of MTX. In the unadjusted cohort, the mean age (SD) was similar between the TNF-α inhibitor and MTX groups, 11.7 (4.2) and 11.8 (4.3) years, respectively. The distribution of females was the same, 71% in both groups.

However, comorbidities and JIA complications were generally more prevalent in the TNF-α inhibitor group. The proportion that was incident to MTX at baseline was 36% and 24% in the TNF-α inhibitor and MTX group, respectively. The mean (SD) follow-up was longer in the MTX group, 2.2 (1.1) compared with 1.2 (1.0) years. During follow-up in the unweighted cohort, we observed 26 events of serious infection among the TNF-α inhibitor users and 75 events in the MTX group. This translated to incidence rates in the TNF-α inhibitor and MTX groups of 4.0 and 1.9 events per 100 patient-years, respectively, in the PS weighted cohort (Figure 8). We observed a significant association between the use of TNF-α inhibitors and the risk of serious infection, HR 1.99 (95% CI 1.09 to 3.65). The site-specific infections with increased risk were respiratory tract infections and infections of the skin and subcutaneous tissue.

5 DISCUSSION

5.1 CLINICAL IMPLICATIONS

In study I, we found that the use of azathioprine in pediatric IBD significantly increased the risk of acute pancreatitis by almost six times during the first 90 days following azathioprine initiation, as compared with no use. The relative risk was similar to the results from two previous studies in adults,[39,82] while the absolute risk was lower than findings from several observational studies and clinical trials in adults.[37-40,83,84,89] This discrepancy could be due to differential risk factors of acute pancreatitis, where gallstone, smoking and alcohol misuse are some of the most important risk factors in adults. Acute pancreatitis is a very rare condition in the general pediatric population and a meta-analysis in adults showed that IBD is a risk factor; a two- to four-fold increased risk depending on IBD subtype was observed, in comparison with the general population.[90] Another similarity with the findings in adults was the short time to onset of acute pancreatitis among azathioprine patients (median 23 days in our study), which supports the notion of an association and a similar mechanism between the patient groups. However, the pathogenesis of thiopurine-induced acute pancreatitis is unknown. A few potential mechanisms have been suggested, including accumulation of toxic metabolites, immunological reactions, and genetic predisposition.[91]

Thiopurine-induced cases of acute pancreatitis in adults have been described as comparably mild.[40] Withdrawal of thiopurine treatment is indicated following onset of acute pancreatitis, which in most cases leads to alleviated symptoms. This was supported by the results of our study where the duration of hospital stay of inpatient acute pancreatitis cases was shorter in the azathioprine episodes in comparison with the no-use episodes (median length of stay was 5.1 and 18.4 days, respectively). The increased risk of acute pancreatitis shortly following treatment initiation supports the current practice of frequent monitoring of pediatric IBD patients during this period. Regular tests of thiopurine-related metabolites and enzymes, including thiopurine methyltransferase (TPMT), supports optimal dosing and reduces the risk of adverse events, such as acute pancreatitis, in these patients.[92]

In studies II and V, we investigated if there is an association between use of TNF-α inhibitors and the risk of serious infections in patients with pediatric IBD and JIA,

respectively. The analysis in pediatric IBD showed no significant association between the use of TNF-α inhibitors and the risk of serious infection, as compared with no use. Whereas the study in JIA found a significant two-fold increased risk associated with TNF-α inhibitors, as compared with MTX.

There are more available pediatric-specific data on this safety concern in JIA than in IBD. Our results in JIA were similar to what has been shown previously in two prospective studies.[55,56] All studies had comparably small sample sizes, but taken together they give a coherent picture of an increased risk of serious infection among JIA patients who initiate TNF-α inhibitors, in particular infections of the respiratory tract and skin. Further, the increase in relative risk is also similar to what has been reported in meta-analyses of RCTs in adults.[45]

In pediatric IBD, there is only one previous controlled study, which was conducted based on insurance claims data in the United States, that offers relevant comparison.[51] Similar to our analysis, this study found no significant association between the use of TNF-α inhibitors and the risk of serious infection in children. The absolute rates of serious infection among users of TNF-α inhibitors were also similar between the studies. Considering the results together, they indicate that use of TNF-α inhibitors might be associated with a smaller increased risk of serious infection in pediatric in comparison with adult IBD, if any at all.

The factors that influence the variable risk of serious infection following use of TNF-α inhibitors need further investigation. Potential aspects are the baseline infection risk, timing of initiation during the disease course, specific biologics used, and use of combination treatments.[93] Intravenously administered infliximab was the predominant TNF-α inhibitor agent in pediatric IBD, whereas subcutaneous adalimumab and etanercept were the most common in JIA. The baseline risk of infection varies greatly between age groups of children and in relation to adults. The characteristics of underlying disease, including its severity and extent, are also important risk factors. The severity of IBD is generally higher in patients with childhood onset IBD in comparison with adult onset.[94] The disease manifestation and prognosis also affect treatment strategies and the timing of initiation of TNF-α inhibitors. As described above (section 2.1), TNF-α inhibitors are used earlier in the disease course and to a larger extent in pediatric IBD than in JIA. Comparisons with adult IBD patients have shown more widespread use of TNF-α inhibitors in children.[18] From a global perspective, it is also

evident that the use of TNF-α inhibitors is more prevalent in the United States in comparison with Europe.[95]

Traditionally bottom-up treatment strategies have been used in both pediatric IBD and JIA, where available pharmacologic therapies form a pyramid with less efficacious but safer treatments at the bottom and more potent but potentially more toxic treatments at the top.[30,96] Treatment starts at the bottom of the pyramid in newly diagnosed patients and is stepped up if tolerability is low or the treatment response is not sufficient. Critical steps in this strategy are the initiations of thiopurines in pediatric IBD, which are more potent than 5-ASA, and MTX in JIA. Followed by the option to add or switch to a TNF-α inhibitor or other biologic at a later point, which represent the top of the pyramid.

This approach has, however, been contested by alternative treatment strategies where either treatment is stepped up faster and initiation of biologics occurs earlier in the disease course or a top-down approach, where biologics are initiated before or concomitantly with thiopurines or MTX.[30,96] The rationale for a top-down strategy is based partly on the, at least short-term, favorable safety profiles of TNF-α inhibitors, and partly on the notion that the efficacy of these treatments and the chance to positively alter the disease course are higher at an early stage.

For particular patient groups, such as severe CD with anal fistula, polyarthritis with high disease activity, and systemic JIA, early initiation of biologics is generally considered favorable.[30,94] In other patient groups, the optimal treatment strategies are less distinct. A challenge in the top-down approach is to distinguish the pediatric IBD and JIA patients with worse prognosis and elevated risk of an aggressive disease course, who could benefit most from early use of biologics.[30,97]

Considering the risks of adverse events is key when determining a suitable treatment strategy. The safety profiles of thiopurines, TNF-α inhibitors, and potential concomitant use in children critically impact the choice and timing of therapies. The results from studies I, II and V can support clinical decision-making and need to be considered when weighing potential benefits against risks when prescribing to children with pediatric IBD or JIA. Important avenues for future research include safety assessments from other health settings, based on larger pediatric cohorts, and with focus on subgroups that

might be at higher risk of acute pancreatitis and serious infections, in particular disease subtypes, age groups, and patients at different steps of the traditional treatment ladder.

5.2 METHODOLOGICAL CONSIDERATIONS

5.2.1 New-user design

Pharmacoepidemiologic studies on the safety and effects of drugs are typically cohort studies based on longitudinal data. A seemingly minor design feature that has significantly improved the quality and relevance of these studies during the last two decades is the new-user design.[98] To make a valid assessment of the effect, new (incident) users of a drug must be studied rather than prevalent users. It is critical to identify the time point of treatment initiation and define time at risk from this point.

There are primarily two advantages of this design. First, the effect of the drug can vary over time and excluding a time interval following initiation can bias the effect estimate, e.g. excluding the time shortly following treatment initiation when studying an acute allergic adverse reaction. There are many examples of effect assessments that have been biased due to not studying new users.[99] We used a new-user design in all studies, I-V. In study I, we found a significant association between the drug and the outcome within the first three months following treatment initiation and no association in the secondary time window, months 4-12. Hence following patients from the initiation of the drug was critical to correctly capture the adverse effect of the drug. Second, patients' covariate status needs to be assessed only based on information available before drug initiation for appropriate confounding control. For example, if we adjust for disease severity status (X) that is evaluated after drug initiation, it could be affected by drug use. If disease severity in turn affects the risk of the outcome (Figure 9; graph a) it is a mediator and we will not estimate the total effect of the drug (A on Y). In another scenario, disease severity post baseline shares cause (U) with the outcome and adjusting for it induces selection bias (Figure 9; graph b). Characterization at treatment start, when patients in most cases have been in recent contact with a physician, also ensures that covariate status is evaluated similarly for all patients.

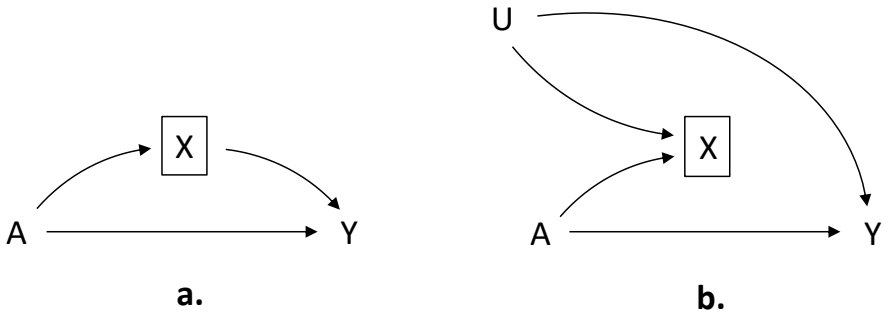


Figure 9. Directed acyclic graphs (DAG) on scenarios of conditioning on post baseline disease status (X), where X is a mediator (graph a) or a collider (graph b). A, drug exposure; Y, outcome; U, unobserved factors causing X and Y.

Identifying the time point of drug initiation can be non-trivial because data on drug use is rarely complete retrospectively, i.e. from birth to observed drug use. In practice a look-back period is used and drug initiation is defined as a filled drug prescription or a drug administration that was not preceded by another prescription fill or administration during a fixed time window before (e.g. two years). The possibility to use a longer look-back window depends on the data source and the duration patients are enrolled. Shorter look-back periods are typically used in analyses based on insurance claims data where disenrollment is more frequent than in national registers (where patients generally are disenrolled only due to emigration). Pediatric patients commonly have shorter disease and treatment history and fewer comorbidities that require a long look-back period to capture. In our data, a large share of patients' history is covered and, in many cases, we have complete look-back to birth. Thus, timing of drug initiation and covariate status can be assessed with higher accuracy in pediatric patients and particularly when using data from national registers.

5.2.2 Target trial emulation

Target trial emulation is a framework for conducting observational studies by relying on concepts and methods from RCTs.[75,76] In this framework, an observational study is designed by mimicking a hypothetical target trial; the experiment that we would have conducted if it was practically possible. The target trial is emulated element by element (e.g. research question, eligibility criteria, treatment assignment, censoring criteria,

outcome definition) to clearly and transparently communicate what is actually done and to avoid unnecessary bias caused by inappropriate design.[75,76] In study IV, we used target trial emulation as a common framework when defining and analyzing differences between alternative pharmacoepidemiologic designs. In study V, we used the concept of a pragmatic trial as a blueprint for the design and used sequential cohorts and a generalized prevalent new user design (described in study IV).

The target trial emulation framework is particularly useful in pharmacoepidemiology, where analyses often share purpose, exposure and terminology with clinical trials – even when the framework is not explicitly used. Evidence from pharmacoepidemiologic studies can be extensions of and complement real clinical trials. Emulating a target trial is sensible because longitudinal and routinely-collected observational data from multiple sources are often complex. Identifying a suitable start and end of follow-up for each patient is not trivial and the lack of prospectively collected data makes patient status at different time points often indistinct. Relying on the structure of clinical trials makes critical design decisions more rigorous and ensures that we analyze retrospective data with a prospective point of view.

Certain aspects of the target trial emulation framework are particularly valuable. Eligibility criteria that are clear, applied consistently and only based on information known at baseline are key. A useful tool to synchronize eligibility, treatment identification and start of follow-up at baseline is sequential cohorts that are defined repeatedly over the follow-up period and resemble a series of repeatedly conducted trials. With this approach, the baseline (time zero; index date) for each sequential cohort is clearly defined and various time-related biases are avoided (Figure 10).[76]

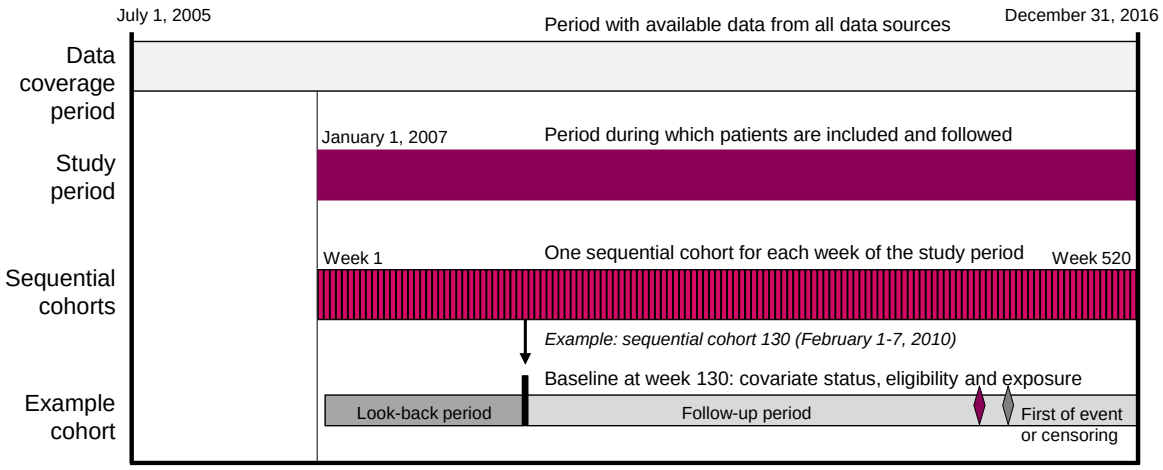


Figure 10. Sequential cohorts defined in discrete intervals over the study period

Further, allowing repeated eligibility and entry over follow-up is an efficient way of using observational data,[100] which is particularly relevant in analyses of small samples such as in pediatric drug safety studies. To be able to follow each patient from multiple start dates that differ in terms of current and previous treatments, history of complications, disease progression, and obviously age, as opposed to selecting only one baseline per patient, adds relevant data to the analysis. The mean length of follow-up in the cohort of patients with confirmed IBD or JIA disease was 4.4 years. However, repeated inclusion of individual patients is not a requirement when using sequential cohorts; patients who once were eligible can be excluded from subsequent cohorts.

5.2.3 Comparator

Exposed patients are compared with a set of patient observations: the comparator (unexposed; control group). Simply put, the ideal comparator contains patients who do not use the study drug but are as similar as possible to those who do, in particular with respect to factors that influence the risk of the outcome. With an ideal comparator the difference in risk between the groups can be attributed to the drug under study. In interventional studies the assignment to study drug and comparator group is randomized, creating balance between the groups on both observed and unobserved factors. In observational studies the comparator definition is a critical element of the design, which is rarely straight-forward and can have a large impact on the results.[86]

In study IV, we investigated the pros and cons of different study designs commonly used in pharmacoepidemiology and the conditions that affect the selection of comparator in a particular study. Specifically, we looked at viable alternatives to the ACNU design, including traditional no use, no use episodes, prevalent new user, generalized prevalent new user, and hierarchical prevalent new user. The ACNU is often described as a robust standard for evaluating safety concerns.[87,88] The ideal active comparator is a drug with the same indication as the study drug, that targets patients with similar disease severity and frailty, and with no known association with the outcome. Only patients who are naïve to both the study drug and the comparator drug are included in the analysis.

An advantage of the ACNU design is that it can reduce confounding, both observed and unobserved, by using patients with the same indication. Another advantage is that patients followed from initiation of the study drug or the comparator drug will be

temporally aligned; they had recent contacts with health care, are in similar phase of disease development and have had similar information collected. The ACNU design can also be used to assess the comparative safety between two drugs where the comparator drug is not necessarily without known association with the outcome.

There are also disadvantages to the ACNU design. First, it can be difficult to find a suitable active comparator drug: comparator drug candidates within the same indication might target different patients or have an effect on the drug studied. Second, only study drug initiators who have not previously used the comparator drug are eligible, which can lead to extensive exclusion, leaving a too small and possibly not representative sample of patients for analysis. This typically occurs when treatments are given in sequence, e.g. when a new drug is introduced and there is channeling to it from the previous standard of care, or when patients switch between treatments due to lack of response or adverse events. In the end, the ACNU design was not used in any of the studies of this project because of the eligibility requirements. For instance, in study I, a potential active comparator drug was 5-ASA. If we had compared azathioprine initiators with 5-ASA initiators using ACNU in study I, we would have excluded 68% of the azathioprine group due to previous 5-ASA use. Given the few events that occurred in these patients, this analysis would have been practically impossible to perform.

This limitation of the ACNU design was addressed by Suissa et al in 2017.[74] In the prevalent-new user design, initiators of the study drug are compared with new and prevalent users of the comparator drug. The study drug initiators (both incident and prevalent to the comparator drug) are matched with comparator initiators and prevalent users in strata based on the extent of previous treatment with the comparator (defined based on treatment duration or number of prescriptions) on time-dependent PS, i.e. stratified PS models. The matching is performed prospectively, starting with the first strata (patients who are incident to the comparator). Individuals with comparator observations can only be matched once, ensuring that outcome events are not accounted for repeatedly in the analysis. Hence, with this design the ACNU cohort (no previous use of the comparator) is analyzed and strata with varying extent of use in the comparator are added, which can be analyzed separately or pooled. In the patients who are prevalent in the comparator, the contrast between switching to or adding the study drug and staying on treatment with the comparator drug is assessed.

In study IV, we also looked at two other prevalent new user designs. The *generalized* prevalent new-user design is less restrictive than the design proposed by Suissa et al, applying the same eligibility criteria but not requiring time-dependent PS matching. Instead, patients can contribute with repeated observations to the comparator group (more than one stratum), which means that follow-up time and potential events are also included repeatedly. This approach is less restrictive in the sense that no exclusion of observations due to lack of match or because another observation from the same individual had already been matched. This can increase efficiency, generalizability and make it possible to use different methods for confounding control, including PS weighting, and estimate different types of effects, e.g. ATT and ATE. In study V, we applied the generalized prevalent new user design, in order to use MTX as active comparator while not excluding TNF-α inhibitor initiators who had previously used this drug (65% of all initiators).

We also described the *hierarchical* prevalent new-user design, which is sometimes simply referred to as a 'new user' design.[54,101-105] In this design, eligibility criteria are applied differentially depending on baseline exposure: patients who are prevalent in the comparator are excluded from the comparator group, but not the study drug group. Analogously, if this design was applied in a clinical trial, both patients who were incident and prevalent to the comparator drug would be enrolled, but only the incident would be randomized (to either the study drug or the comparator drug). The prevalent patients would be automatically assigned to the study drug. If previous treatment with the comparator or characteristics of that treatment history (time since initiation, cumulative dose, etc.) are confounders they cannot be adjusted for due to the deterministic violation of positivity; among prevalent patients the true PS is one.[106]

The potential bias in the hierarchical prevalent new-user design depends on what previous use of the comparator represents and many different scenarios are plausible. In one scenario, use of the comparator increases the risk of the outcome (contrary to the standard criteria for a suitable active comparator), which means that patients with previous use and no previous outcome event represent survivors and potentially have lower risk of the outcome. In a second scenario, there is no effect of the comparator drug, but previous use is a positive proxy for disease severity, which increases the risk of the outcome. In an opposite third scenario, continuous previous use is a negative proxy representing healthy users who have lower risk of the outcome. Irrespective of

what previous use in the comparator represents, previous use of a comparator drug commonly contains information about risk factors for which adjustment is necessary. Despite the risk of bias, the hierarchical prevalent new user design is surprisingly common in pharmacoepidemiology; possibly because it solves the fundamental challenge of using an active comparator while not excluding those who are prevalent in the comparator drug.

Finally, we also investigated the less restrictive no-use designs, where the comparator group consists of patients with the same underlying disease as the study drug initiators and with neither current nor recent use of the study drug.[86] No use designs are commonly misunderstood; possibly due to inappropriate application in the past. When implemented correctly, 'no use' simply means no use of the study drug at a certain time point and during a set look-back period before. It does not mean no use of any pharmaceutical drug or no use of the study drug during the entire study period, which could introduce selection bias. The use of multiple sequential cohorts with repeated baselines during the study period, as described in section 5.2.2, facilitates transparent and unbiased assignment of index dates.

In study IV, we included a traditional no use design that served as a template for the other designs since the eligibility criteria were basic (indication, no previous study drug, no previous event) and all other designs were nested within it. Similarly, to the generalized prevalent new user design, overlapping follow-up time and events are included to use the data in the most efficient way. In contrast, in the no use-episode design that was used in studies I-III, all study drug users and non-users were analyzed in mutually exclusive episodes of follow-up. This was achieved by defining a maximum length of the episodes and adding to the eligibility criteria that a patient observation was excluded if the same patient had contributed a previous episode that was still ongoing. The length of episodes was set to one year in study I and three years in studies II and III. When applying this design, multiple episodes of both study drug use and no-use could be contributed by the same patient, but not more than one outcome event. In practice the no-use episodes design can be very similar to the prevalent new-user design in terms of patient selection for the comparator group, especially if we condition on previous use of a comparator drug. However, important potential limitation of no-use designs in relation to active comparator designs is the risk of information bias and confounding by indication. In study IV, we conclude that no-use designs are in particular

useful when no suitable comparator drug is available, which is not a rare scenario. Active comparator designs are generally preferred and a prevalent new user design can be used when ACNU requires extensive exclusion.

5.2.4 Confounding by indication

Confounding can be a major issue in observational safety studies. If factors that affect both exposure and outcome are not adjusted for, the association between exposure and outcome is confounded (Figure 11; graph a). In pharmacoepidemiologic studies, good clinical practice and tailored prescribing of drugs may lead to confounding by indication – that study drug users are systematically different from comparators with respects to risk factors for the outcome. Despite targeted prescribing, there are many reasons why similar patients do not receive the same treatment. Prescribing patterns can vary between geographical regions, hospitals, individual physicians, time periods and due to patient preference. These variations are key for being able to perform drug safety analyses based on observational data.

Confounding by indication can be a stubborn bias for which adjustment is challenging. The difficulty in identifying comparator patients with the same indication and disease severity as the study drug users is one of the key reasons that active comparators, where confounding is mitigated by design rather than statistical analysis, are useful (as in study V). In studies with no-use comparators (studies I-III), robust confounding control needs to be achieved through the statistical analysis. Typical potential confounders in a drug safety study are age, sex, comorbidities, treatment history, disease stage, and disease severity. If data has not been collected on the identified potential confounders an alternative is to use proxies, i.e. factors that are highly correlated with the potential confounders. For example, a proxy for general health status is health care use, which could be measured as the number of unique drugs used pre-baseline (Figure 11; graph b).

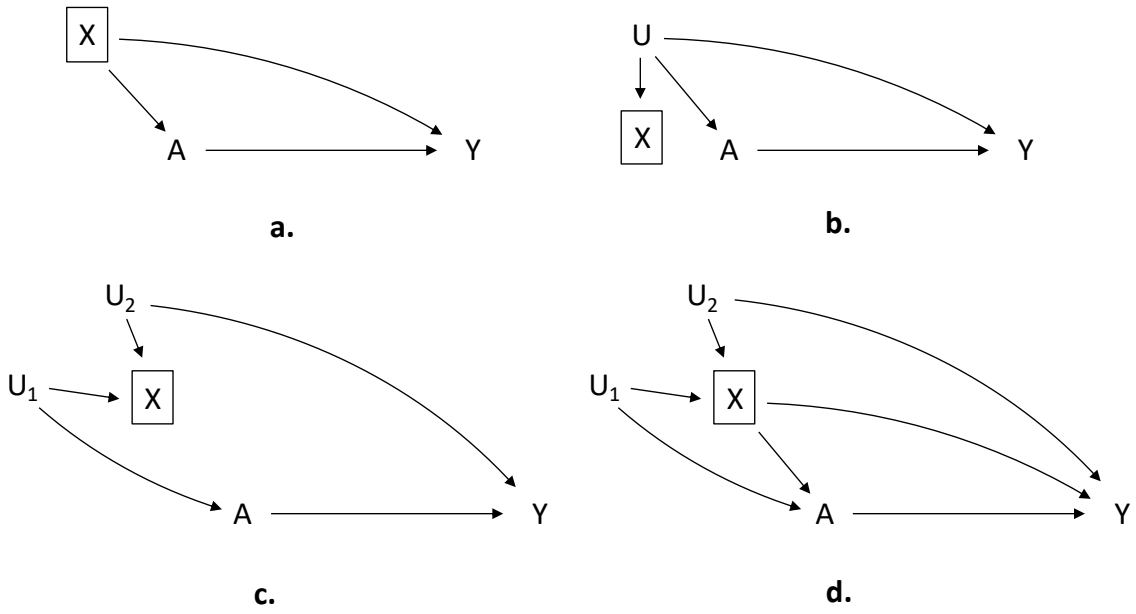


Figure 11. DAGs of confounding adjustment based on adjusting directly for a confounder (graph a), adjusting for a proxy for a confounder (graph b), M bias where the factor that we adjust for is a collider (graph c), and both collider and confounder (graph d). A, treatment; Y, outcome; X, the factor that we adjust for; U, unobserved factor.

5.2.5 Propensity score methods

There are many methods to adjust for confounding, such as stratification, standardization, and regression analysis. PS methods, introduced by Rosenbaum and Rubin in 1983, are common in pharmacoepidemiology.[77,107] The PS measures the propensity of an individual patient to be assigned treatment; the probability of being exposed ($0 \leq e_i \leq 1$). In an RCT where all patients have equal probability of being assigned treatment the PS is 0.5. In observational studies the true PS is unknown, but we can estimate it conditioned on potential confounders in our cohort (Formula 2).

$$e_i = P(A_i = 1|X_i)$$

[Formula 2] PS (e_i) is the conditional probability of being exposed. A_i baseline treatment (1 exposed; 0 comparator); X_i vector of baseline confounders.

The PS is useful in confounding control because it is a balancing score. For each value of the PS, the distribution of potential confounders that the PS was conditioned on is the

same in the exposed and comparator. Hence, information from multiple confounders is collapsed in the PS and balance between exposed and comparator can be achieved by conditioning on it, e.g. through matching or weighting.

The key advantage of PS methods is that we can perform robust confounding adjustment by modelling the probability of the treatment rather the outcome. Due to the balancing property of the PS we can control for many confounders independently of the prevalence of the outcome. If we include too many covariates in relation to the number of events in a multivariable outcome model, we may obtain biased estimates or have convergence problems.[108] In drug safety studies the combination of few outcome events and extensive confounding is common, particularly in pediatric studies. We used PS matching or weighting in all the studies of this project.

5.2.5.1 Propensity score model estimation

A valid PS analysis relies on a correctly specified model of the relationship between treatment assignment and potential confounders. The PS is commonly estimated with logistic regression, which was used in studies I-V, where treatment is the dependent variable and potential confounders are independent variables (Formula 3).

$$e_i = \frac{exp(X_i\beta)}{1 + exp(X_i\beta)}$$

[Formula 3] PS (e_i) estimated with logistic regression. X_i vector of baseline confounders; β vector of coefficients estimated from the data.

Key diagnostics in PS analysis is the crude PS distribution, the overlap between exposed and comparator, and balance of individual covariates in the adjusted cohort. Differences and large separation of the PS distributions between the groups can indicate a misspecified PS model or lack of positivity, i.e. there are levels in the potential confounders where all are exposed or comparator observations. This can in turn be due to the selection of an unsuitable comparator with large differences in covariate status at baseline, which might also indicate imbalance in unobserved factors.

Covariate balance is an intuitive diagnostic of PS model performance. It can be assessed by calculating absolute standardized mean differences for each covariate, both continuous and dichotomous. This measure expresses the difference in means in units of the pooled standard deviation (Formula 4). A difference smaller than 10% is commonly regarded as well-balanced.[77] Standardized differences are preferred over hypothesis tests and p-values because they are not affected by sample size.

$$d = \left| \frac{\bar{x}_{A=1} - \bar{x}_{A=0}}{\sqrt{(s^2_{A=1} + s^2_{A=0})/2}} * 100 \right|$$

[Formula 4] Absolute standardized mean difference (d). $\bar{x}_A$ sample mean in covariate x among observations with treatment A (1 exposed; 0 comparator); s^2_A sample variance in covariate x among observations with treatment A.

We performed this assessment in studies I-V. In the case example of studies IV, we plotted the standardized mean differences (not absolute) of 58 risk factors to show differences between comparators (only adjusted for age and sex). A positively skewed distribution indicated a higher observed risk in the study drug group and a negatively skewed distribution meant a lower risk.

Additionally, a more granular assessment of balance can be performed for continuous variables by comparing empirical cumulative density functions (eCDF) in the exposed and comparator for visual inspection.[109] In Figure 12, this is shown for the covariates disease duration and measures of general health care use for the crude and weighted cohorts in study V. The difference in eCDF can also be quantified with the Kolmogorov–Smirnov statistic, which is the maximum vertical distance between the eCDF in the exposed and the comparator.

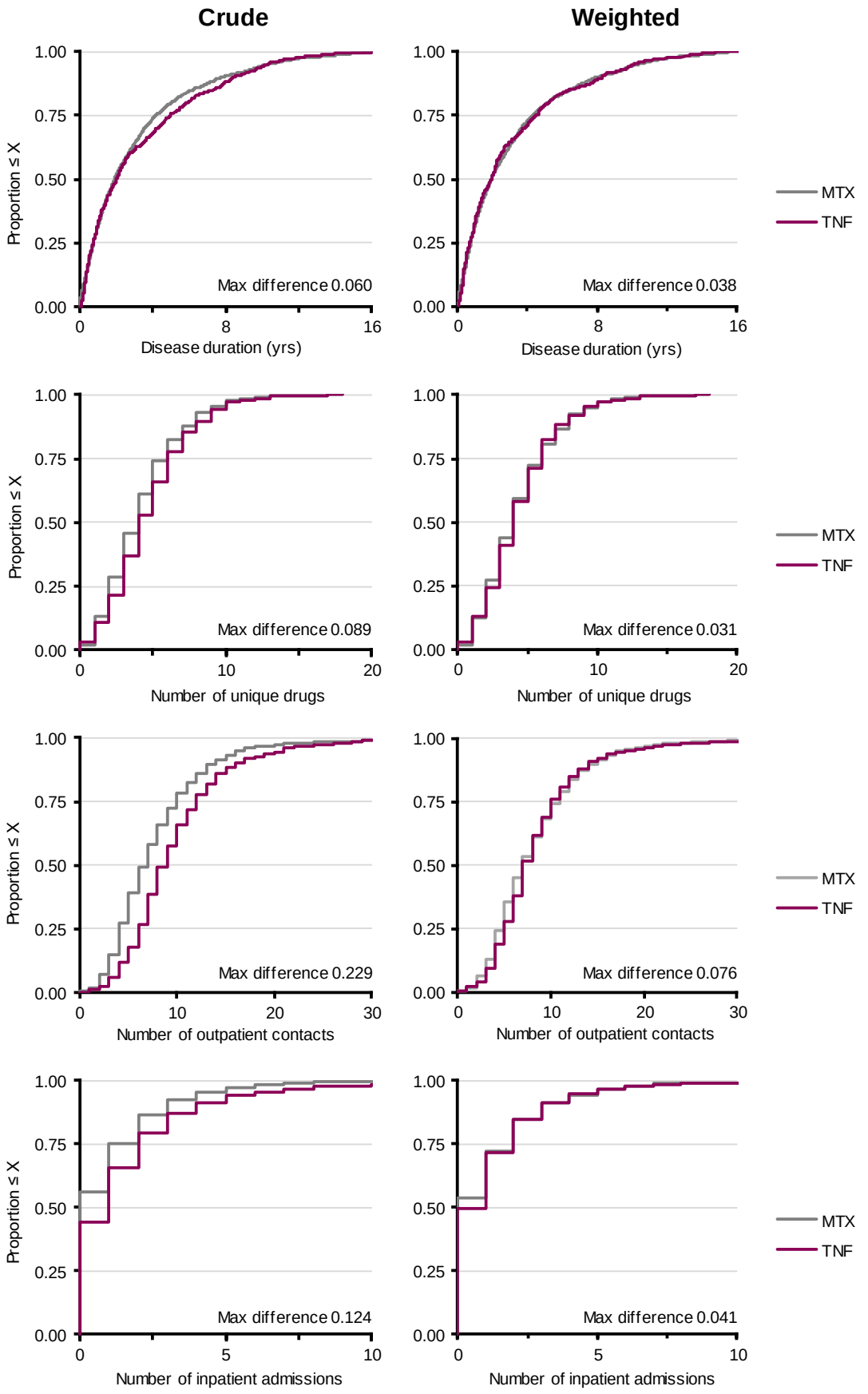


Figure 12. Diagnostics of baseline covariate balance in study V: empirical cumulative density functions of covariates in the TNF-α inhibitor and MTX groups of the crude and weighted cohorts.

An advantage with PS methods is that alternative PS models can be assessed and compared against each other without involving the outcome. In the study cohort, alternative methods for PS estimation, selection of confounders and model specification can be tested in order to optimize balance at baseline, before estimating the association between drug and outcome event. Beyond logistic regression there are numerous more flexible, data-adaptive methods that have been proposed to improve PS estimation and reduce bias, including machine and ensemble learning methods that can be applied with cross validation.[110,111] However, logistic regression is still the most common method for PS estimation in applied pharmacoepidemiologic analyses.

5.2.5.2 Covariate selection

Preferably covariates for a PS model are chosen based on subject-matter knowledge and the assumed causal structure surrounding the drug-outcome relationship.[112] Data is collected and adjustment is made for the identified potential confounders. However, in pharmacoepidemiology, where large and complex secondary data sources commonly are used, prospective data collection is rarely feasible. To improve confounding adjustment, methods for empirical covariate selection have been developed and are commonly used. With these methods covariates are selected, partially or solely, based on observed associations in the data. The rationale for using these methods is that the underlying causal structure is largely unknown and that confounding control can be improved by adjusting for large sets of proxy variables that are associated with both the treatment and outcome. The causal relationships and the role of the proxy variables are not necessarily known.

One of the most common methods for empirical covariate selection is the high dimensional propensity scores[113], which is an algorithm where potential baseline covariates are ranked univariately based on association with treatment and outcome. Those with the highest rank are included in the PS model; based on a predetermined threshold. The potential covariates or proxy variables are derived as dichotomous indicators of history of disease, medical procedures and treatment. Some of the limitations of this method are that covariates are selected independently of each other and the potential for overfitting.[114] However, these issues can be overcome by adapting the covariate selection procedure and using flexible, data-adaptive methods, such as ensemble learning algorithms and penalized regression, rather than an univariate

screener.[115,116] Yet, one limitation that applies to all approaches of empirically selecting covariates is the risk of adjusting for pre-baseline colliders, i.e. opening a backdoor path between treatment and outcome, and introducing selection bias. This bias is commonly known as M bias (Figure 11; graph c) and simulation studies have shown that the potential for this bias is small in relation to the bias caused by lack of adjustment for confounders.[117] Note that a pre-baseline collider can simultaneously be a confounder (Figure 11; graph d), in which case adjustment is generally recommended.

In addition to methods based on only PS estimation there are doubly robust methods, such as targeted maximum likelihood estimation,[118] where both the treatment and outcome are modeled. It has been shown that this two-step procedure can optimize the bias-variance tradeoff in estimation of causal effects.

5.2.5.3 Propensity score matching

In studies I and III, we adjusted for confounding through PS matching. In PS matching, each exposed patient is matched with one or many comparator patients who have similar PS (within a certain absolute caliper) according to a predefined ratio, e.g. 1:1. Various matching algorithms are available and in the most common, greedy nearest-neighbor matching, exposed patients are selected randomly and matched with the comparator patient (among those who have not already been matched) where the difference in PS is minimized.[77] PS matching is a simple and intuitive procedure that allows a transparent presentation of results and balance assessment.

In PS matching we estimate the ATT if all exposed patients in the crude cohort are matched. Due to lack of overlap in PS distribution or too few comparator patients, a completely matched cohort is rare. In practice, the estimand in a PS matched cohort is ATT in the exposed who were matched, which is not necessarily a distinct subset of patients with certain characteristics. The matched patients are indirectly defined by properties of the matching procedure, such as selected confounders, caliper, and comparator group. Exclusion due to lack of match can lead to decreased generalizability and precision.

The selection of caliper in PS matching represents a tradeoff where a small caliper gives less bias at the expense of reduced generalizability and precision. Typically, the caliper is set relative to the dispersion of the estimated PS in the crude cohort, e.g. 20% of the

pooled standard deviation of the logit PS,[77] which was used in study I where 94% of the exposed were matched. In study III, where we performed a data mining analysis we prioritized efficiency over bias reduction. We used a very large caliper and the entire cohort was matched, while maintaining acceptable balance on all covariates.

Under some conditions, in particular if the sample is small a reduced caliper can increase bias, which has been described as the PS matching paradox.[119] However, it has been shown that this is rare if standard caliper sizes (relative to dispersion) are used and it is possible to test if the analysis is susceptible to this issue by varying the caliper.[120]

5.2.5.4 Propensity score weighting

In studies II, IV and V, we adjusted for confounding with PS weighting: SMR, fine stratification weighting, and stabilized IPT weighting, respectively. In weighting, each observation is assigned a weight that is calculated based on the PS (see formulas in previous section 3.3.1) in order to create a weighted pseudo population. Key advantages of PS weighting in relation to matching are less exclusion of exposed observations, flexibility in terms of the estimated effect, and low computational intensity. If there are fewer comparator patients than exposed, weighting is the obvious choice to avoid exclusion.

While lack of overlap in PS distribution between exposed and comparator leads to exclusion in PS matching, in PS weighting it leads to large weights. Extreme weights decrease precision and indicate lack of positivity or that the PS model is misspecified.[109] In contrast to PS matching, positivity violation is avoided by excluding patients with PS outside of the common range in PS weighted analyses. Additionally, patients with weights in the bottom and top percentiles, e.g. in the 1st and 99th, can be truncated to avoid extreme weights. However, this truncation can lead to increased bias.

With SMR and fine stratification weighting we estimate the ATT, while we estimate the ATE with stabilized IPT weighting (see weighting formulas in section 3.3.1). In study II, we used SMR weighting to estimate the ATT because we used a no-use comparator and wanted to estimate the effect in those who actually received TNF-α inhibitors. In the case example of study IV, we used fine stratification weighting to estimate the ATT to

preserve the comparator sample size and estimate an effect that was as similar as possible between the alternative comparator analyses.

Fine stratification has the advantage that the comparator sample size is kept and extreme weights can be avoided by using a lower number of strata (note that there is also an ATE version of fine stratification weighting that was not used in this project). In study V, where we used an active comparator and the treatment groups were fairly similar at baseline we estimated the ATE. IPT weighting is the most commonly used PS weighting method and is a part of g methods (generalized methods which are also applicable in analyses of time-varying exposure) and targeted maximum likelihood estimation.[118]

5.2.6 As-initiated and as-treated analyses

As in clinical trials, patients can be analyzed based on the treatment assignment at baseline (intention-to-treat) or based on the treatment actually received during follow-up, i.e. following the patients who adhere to the assigned treatment strategy and as long as they adhere to it (per-protocol). In the observational setting, the analyses corresponding to intention-to-treat and per-protocol may be referred to as-initiated and as-treated, respectively. In the as-initiated analysis patients are followed from baseline where a certain drug is initiated or not (there is no information on intended treatment) to the end of follow-up with no censoring due to treatment changes. In the as-treated analysis patients are censored if they deviate from the treatment strategy at baseline, typically at treatment stop or change to the other drug. In the unrealistic scenario of perfect adherence, the analyses yield the same result.

In practice, both analyses are susceptible to bias. Non-adherence leads to exposure misclassification and dilution of a possible effect in the as-initiated analysis. In the as-treated analysis, censoring of non-adherent patients can be informative and cause selection bias. For example, if frail patients have a higher risk of stopping treatment the proportion of frail patients will decrease over follow-up. Frailty might be a risk factor and differential between exposed and comparator. As in RCTs, the as-initiated analysis is commonly used to avoid selection bias. In studies with long maximum follow-up, an as-treated analysis can be preferable to avoid extensive misclassification of exposure.

Further, if the studied adverse event is suspected to occur while patients are on active treatment this approach is also preferred.

In studies II and III, TNF-α inhibitor episodes were censored as treatment stop and no use-episodes were censored if any TNF-α inhibitor was initiated. An as-initiated analysis was performed as a sensitivity analysis in study II. In study V and the case example of study IV, we performed as-treated analyses and attempted to mitigate potential selection bias from informative censoring of patients who changed treatment by reweighting observations repeatedly over follow-up. We used IPC weighting where the weight of a certain patient and time interval during follow-up is calculated as the inverse of the conditional probability of *not* being censored in the previous interval.[78] Consequently, patients at risk were assigned weights at each time interval so that they also represented the censored patients; including their distribution of risk factors for the outcome.

IPC weighting is a fairly intuitive method that is conceptually similar to weighting used to adjust for confounding. While baseline confounding adjustment is taken for granted in observational studies, IPC weighting and other similar methods used to adjust for selection bias in an as-treated analyses are underutilized.[121] Adjustment post baseline requires the splitting of follow-up discrete time intervals with time-updated covariate status that make the analysis more computationally intensive. However, given the issues with both as-initiated and as-treated analyses it makes sense to estimate both effects, adjust for potential informative censoring in the latter, and to apply variable time windows to give a comprehensive picture of the true potentially adverse effect of a drug.

5.2.7 Data mining with scan statistics

In study III, we performed a data mining analysis to detect signals of adverse events of TNF-α inhibitors. We used newly developed PS matched tree-based scan statistics and tree-temporal scan statistics in a self-controlled analysis. A key feature of these methods is the a priori unrestrictive designs. The investigated potential adverse events were not specified: all diagnoses in the ICD-10 tree were considered (except diagnoses that could not be caused by a drug, e.g. congenital malformations) and there was no restriction in terms of diagnosis granularity. The self-controlled analysis was additionally

unrestricted in terms of the risk windows that were analyzed. We tested a global hypothesis: is there any potential adverse event with an elevated risk in the exposed, as opposed to no elevated risk in any of the potential events. Hence, a large number of potential adverse events were screened, while adjusting for multiple testing to correctly estimate p-values.

We adjusted for basic confounders in the PS matched analysis. Confounding adjustment is commonly left out of signal detection studies due to lack of data on covariate status. The confounding adjustment in the PS matched analysis was static, i.e. we adjusted for the same covariates in the same matched cohort in relation to all events. In the self-controlled analysis, we adjusted for time-fixed factors through the study design. The selection of potential confounders in a data mining study of more than 1000 potential events is not straight-forward. We adjusted for general confounders that could influence the risk of many types of events to increase the relevance of the results and avoid generating spurious signals. However, the aim of the analysis was not to infer causality between exposure and events. New signals generated by the analysis needed to be evaluated by researchers with subject-matter knowledge and possibly investigated further with traditional pharmacoepidemiologic drug safety designs based on other data sources. In a recent study,[86] three confounder selection strategies were evaluated in data mining of adverse events of four drugs using an active comparator design: general covariates, confounders selected with a data-adaptive approach, and covariates tailored to the drug pair studied that were selected by the investigator. It was concluded that confounder selection had little impact on the identified signals and there was a tradeoff between power and extensive confounding adjustment.

A key challenge in adverse event screening is the handling of dependent events. There is deterministic dependence between cuts at different levels in the diagnosis tree, i.e. an event automatically generates events in ancestor nodes. This dependence is accounted for in the estimation of p-values by using the same structure in the data simulated under the null hypothesis. Handling repeated and potentially dependent events within patients is less straight-forward. One approach is to not include more than one event per patient in the analysis by, for example, choosing either the event that occurs first, the rarest, or that which is the most serious event.[81] However, this omission of data leads to decreased power and might result in true signals going undetected. There are also difficulties in prioritizing between events, which contradicts the generally

unrestrictive nature of this type of analysis. In study III, all events were included in the analysis and we accounted for dependence between events within ICD-10 chapters by regarding these events as clusters that were randomized together either to the exposed or the comparator in the simulation under the null hypothesis. Potential clusters of adverse events signals need to be evaluated based on medical and clinical experience.

5.3 ETHICAL CONSIDERATIONS

Drug safety research in children poses many ethical considerations. RCTs represent the gold standard for evidence generation on safety and efficacy. However, the enrollment of children in RCTs of drugs that have yet not been thoroughly tested involves risks. It poses an ethical dilemma: evidence from clinical trials is needed to improve treatment of sick children, but generating that evidence would put sick children at risk. Children are a vulnerable patient group and consent is commonly given on a child's behalf by their parents. Not surprisingly, clinical trials in severe diseases where there are few or no treatment alternatives can more easily enroll pediatric patients.

Not enrolling pediatric patients in RCTs is also an ethical position. This effectively passes the decision of using a drug on to clinicians who potentially will use it off-label based on extrapolated evidence from adults, which certainly also involves risks. It also means that some efficacious and safe drugs are withheld from the use in children. The perception of potential harm of RCTs versus off-label use can be deceiving. Use in a large number of patients in clinical practice based on many individual treatment decisions dispersed over time can seem less harmful than conducting a single RCT that includes a small number of patients, although the former involves more potential harm overall. Regardless of whether RCTs are conducted, a positive consequence of use in clinical practice is that it generates observational data, which can contribute to pharmacoepidemiologic analyses and safer use in the future.

Conducting observational drug safety studies, such as the analyses of this project, are not void of ethical concerns. The primary potential harm to patients comes from the mishandling of their personal medical data; such as violation of privacy and unethical use of data against their interest. Scandinavian health registers are population based and include all persons who receive general-access healthcare, unless they actively opt

out of the registration of their personal data. Patients whose data is used in observational studies are often unaware of this aspect of the health care system; consent is not required and not actively sought. Register data is anonymized by register holders before being distributed for research purposes. However, there is still a risk that individuals are identified based on the detailed information that the register contain. The risk of reidentification increases with more sophisticated and detailed data. For example, if date of birth, sex, municipality, income of both parents and disease history of a person are known, that information can narrow down a gross cohort of 5.3 million unique children in Denmark and Sweden to very few, especially for individuals residing in less populated areas. Furthermore, algorithms for reidentification are becoming increasingly advanced, which increases the threat.[122]

Besides the violation of patients' privacy, reidentification can enable unethical use of personal data. Both government and private organizations could have incentives to use the data against patient interest, e.g. insurance companies could discriminate against clients with preexisting conditions, employers could discriminate against current or potential employees, and legal authorities could use the data within investigations. The potential harm of unethical use of patient-level medical data today and in the future is complex and difficult to review. Nevertheless, the level of detail on each individual and the lack of aggregation, which enables potential reidentification, are key to performing high-quality drug safety analyses; in particular, to achieve robust confounding control. To be able to track individuals in terms of drug use, medical procedures, diagnoses, and timing of health care contacts is crucial to establish causal relationships between drug use and adverse events.

These legitimate interests at stake need to be weighed against one another. On the one hand, there is the potential benefit of this project: new safety evidence for treatments in common and serious diseases in children that can support clinical decision making and potentially lead to better patient outcomes in the future. On the other hand, there is the need to protect patients against potential unethical use of their data and violation of patient privacy. Preventive measures have been taken to mitigate the potential harm, such as data storing on secure servers, granting access only to researchers who perform data analysis, and not publishing results for individual patients, irrespectively of detail level. Given the potential benefit of these studies, the actions to protect the data, and the

relatively low risk of a data breach and reidentification, we think the potential good clearly outweighs the potential harm of this thesis project.

5.4 POINTS OF PERSPECTIVE

In this section we elaborate on how the generation of high-quality pediatric-specific drug safety evidence can be improved in the future.

5.4.1 Data sourcing

Scarce data puts a fundamental restriction on drug safety analyses. It can be a barrier in studying rare events, having necessary precision in our estimates, being able to study effect modification between subgroups of patients, and performing robust confounding control, both through the design and the statistical analysis. However, small sample size is common and expected when studying a subgroup of patients, such as children, where disease and drug use prevalence is lower than in the adult population.

When sourcing data there is typically a tradeoff between population size and data granularity: a large cohort that is needed to study a rare event has less detailed data on individual patients. The relevance of drug safety analyses in pediatrics would be vastly improved if this hurdle could be overcome with improved data recording procedures, centralized collection and international collaborations. In the future, multi-national pooled, harmonized cohorts with clinical data, including electronic medical chart data, laboratory test results, and patient-reported outcomes could be generated. Based on such data sources it would be possible to study rare events, even in subgroups of children, such as relevant age strata, while maintaining robust design and confounding control.

5.4.2 Adverse event data mining

Data mining for adverse events based on health registers, as shown in study III, is a promising source of pediatric-specific drug safety information. If data sources are extended and direct reporting to authorities is enabled this could replace the spontaneous reporting systems in the future. With regular, time-updated screening of

health registers where diagnoses, and separately recorded suspected adverse events, are routinely reported, the identification of new signals can be instantaneous and relevant comparator groups can be generated for robust confounding control based on the same data source. This type of real-time post-approval surveillance analyses based on health registers can identify signals of adverse events of drugs used both on and off-label in children earlier and with higher accuracy than previously.

In such a setting, novel methods will be needed to improve statistical efficiency. The PS matched tree-based scan statistics approach applied in study III is fairly restrictive, which can reduce the usefulness in pediatrics. As noted above (section 5.2.5), PS matching can reduce power through exclusion of study drug users without a match. Further restriction comes from censoring to harmonize follow-up within matching clusters and in combination with an ACNU design where study drug initiators who have previously used the comparator are excluded.[80,123]

The scope for performing signal detection against a comparator with scan statistics outside of the PS matching framework might be more suitable in pediatrics and requires further investigation. Ideally, a more general framework based on repeated inclusion of comparator patients and PS weighting for confounding control would be useful to increase efficiency. Naturally, such an approach would increase the level of dependency between events (repeated eligibility of individual patients and weighting in the pseudo cohort) which needs to be considered in the simulation of data under the null hypothesis. Furthermore, to allow variable follow-up between observations the timing of events (including clusters of events) in relation to baseline needs to be considered. With methods that address these obstacles the opportunity of adverse event data mining in pediatrics would be even more promising in the future.

5.4.3 Best practices in pediatric pharmacoepidemiology

To ensure usefulness of drug safety data from the observational setting and to facilitate comparison and aggregation of results from different studies robust design and statistical methods are key. In terms of design, sensible definitions of eligibility and exposure that are strictly applied throughout the cohort and study period (e.g. with sequential cohorts) are crucial. As described above (section 5.2.2), many types of common biases can be avoided and study results can be clearly interpreted by relying on these principles. Another vital design feature is active comparators, which can

mitigate information bias and confounding by indication. As discussed in section 5.2.3, the less restrictive prevalent new user designs are useful in pediatrics to maximize generalizability and efficiency and yet gain the benefits of confounding control from an active comparator design.

Confounding control in pediatrics offers some particular opportunities. Low age and short disease and treatment history means that complete data on patients since disease onset or even since birth is available for a large proportion of patients in national registers. The possibility to characterize patients at the initiation of a drug based on their entire history, based on large sets of proxy factors, can potentially improve confounding adjustments and needs to be explored in the future.

As described in section 5.2.5, confounding control can also be improved with flexible, data-adaptive PS modeling and with empirically identified potential confounders. Given the challenge of confounding by indication and the complexities of secondary data sources, applying data adaptive methods is as viable as traditional methods for covariate selection. In many cases, these approaches can be applied in parallel and evaluated based on their strengths and limitations. Further, doubly robust methods, such as targeted maximum likelihood estimation, where both the treatment and outcome are modeled to reduce bias and increase efficiency, are promising in pediatric pharmacoepidemiology where statistical precision can be low. However, the small sample properties of these methods need to be explored further.

Finally, the as-treated analysis is often the most relevant analysis from a drug safety perspective, since a potential adverse effect can be diluted and not detected in an as-initiated analysis. Nonetheless, as pointed out in section 5.2.6, this analysis can be susceptible to informative censoring. Methods for time-updated adjustment in an as-treated analysis, e.g. IPC weighting, are generally underutilized and would add robustness to drug safety analyses in pediatrics.

6 CONCLUSIONS

The overall aim of this thesis was to develop new, relevant, and pediatric-specific safety evidence for treatments in chronic inflammatory diseases. Based on data from Scandinavian national health care registers, we found that use of azathioprine was associated with a 6-fold increased risk of acute pancreatitis in pediatric IBD. When investigating the use of TNF-α inhibitors and the risk of serious infection, we found a two-fold increased risk in JIA patients, but no increased risk in patients with pediatric IBD. Through data mining we screened the registers for new signals of adverse events of TNF-α inhibitors in both pediatric IBD and JIA patients. We identified two signals; none of which were deemed relevant for further investigation. Finally, we investigated the differences between common designs in pharmacoepidemiology and provided guidance on key factors that need to be considered when choosing a comparator in an observational drug safety study.

This thesis was driven by the need for relevant safety data when drugs are prescribed to children. Pediatric patients are vastly understudied when it comes to drug safety, both historically and today. The shortage of data applies to most pediatric indications and is by no means limited to chronic inflammatory diseases. Hopefully this work will inspire more research in the future to narrow the evidence gap.

7 POPULAR SCIENCE SUMMARY

There is a key question that physicians need to ask before prescribing drugs to children: Is this drug safe? Physicians rely on information from large trials where drugs are tested in controlled settings. Although the testing of drugs has progressed tremendously during the previous century and is highly regulated to protect patients, the information that comes from trials rarely applies to children. Consequently, when prescribing to children, physicians need to rely on what they know about risks to adults, which may not be relevant because children and adults often react differently to treatments. The overall aim of this thesis was to bridge this gap using data from Scandinavian clinical practice to investigate the safety of some drugs that are commonly used in children with chronic inflammatory diseases. Our data covered 5.3 million children in Denmark and Sweden; including 21,000 patients with these diseases.

We did studies on specific safety concerns where we need more information, i.e. drugs that may potentially cause certain adverse events. First, we looked at azathioprine, which is a common drug that suppresses the body's immune system, and if it increased the risk of acute pancreatitis in children with inflammatory bowel disease. Acute pancreatitis is the sudden and often painful inflammation of the pancreas. When looking at children in Denmark and Sweden who used azathioprine, we found that the drug increased the risk of acute pancreatitis by six times. Among the children who used azathioprine, 1.2% experienced this event in the first 90 days of starting the drug. Second, we studied modern and highly efficacious biologic treatments to see if they increased the risk of serious infections, defined as infections where the patient needs to be hospitalized. In two separate studies based on Danish patients, we found that the use of biologic treatments increased the risk of serious infection by two times in children with rheumatic disease; whereas it did not increase the risk in children with inflammatory bowel disease.

We also did a study of biologic treatments in children, where we searched the data for any new signals of adverse events that were previously unknown, rather than looking at a particular adverse event. We searched among more than 1000 diagnoses and found increased risks in two of them, skin complications and adjustment disorders; neither of which were considered relevant for further investigation.

Finally, we addressed a critical methodological issue when studying safety of drugs based on data from clinical practice: the choice of comparator group, i.e. the patients that we compare to the treated patients. We relied on concepts from the world of clinical trials and concluded that multiple factors influence the selection of comparator group, e.g. if there are patients who use similar drugs that can be used as comparator, if the potential risk of the drug is expected to vary between patient groups, and the size of our patient sample.

In summary, we investigated concerns related to the risks of certain drugs in children with chronic inflammatory diseases. We showed that information that can support physicians when prescribing to children can be derived from data from clinical practice in Scandinavia. Given the dearth of this type of data in children hopefully this work will inspire more studies in the future.

8 POPULÄRVETENSKAPLIG SAMMANFATTNING

När läkare förskriver mediciner till barn behöver de ställa frågan: Är läkemedlet säkert? Läkare använder information från kliniska prövningar där mediciner testas på ett kontrollerat sätt och även om dessa är strikt reglerade för att skydda patienter gäller informationen från prövningar sällan barn. Det innebär att läkare som behandlar barn behöver förlita sig på riskinformation som är känd från vuxna, vilken inte nödvändigtvis är relevant då barn och vuxna reagerar olika på många mediciner. Syftet med den här avhandlingen var att utifrån data från klinisk praxis i Skandinavien undersöka risker med läkemedel som är vanliga hos barn med kroniska inflammatoriska sjukdomar. Våra data täckte 5.3 miljoner barn i Danmark och Sverige; varav 21,000 var patienter med dessa sjukdomar.

Vi genomförde studier på specifika läkemedelsrisker; mediciner som eventuellt orsakar vissa biverkningar där vi behöver mer information. I en studie undersökte vi om användning av azatioprin, en vanlig medicin som hämmar immunförsvaret, ökar risken för akut pankreatit hos barn med inflammatorisk tarm. Akut pankreatit är en plötslig och ofta smärtsam inflammation av bukspottkörteln. När vi tittade på patienter i Danmark och Sverige fann vi att användning av azatioprin ökade risken för akut pankreatit med sex gånger. Bland barn som använde azatioprin fick 1,2% denna biverkning under de första 90 dagarna efter behandlingsstart. I de andra studierna undersökte vi om användningen av biologiska läkemedel, som tillhör senaste generationen av behandlingar och är mycket effektiva, ökade risken för allvarlig infektion (som kräver att patienten blir inlagd på sjukhus). I två separata studier tittade vi på danska patienter och fann att biologiska läkemedel ökade risken för allvarlig infektion med två gånger hos barn med reumatisk sjukdom; medan risken inte ökade hos barn med inflammatorisk tarm.

Vi gjorde också en studie på biologiska läkemedel hos barn där vi sökte efter signaler på nya biverkningar som tidigare inte var kända. Vi sökte bland fler än 1000 diagnoser och hittade signaler på ökad risk i två av dem: hudsjukdomar och anpassningssvårigheter. Ingen av dessa ansågs relevant för en fördjupad undersökning.

Slutligen så undersökte vi en kritisk fråga om metodval i studier av biverkningar utifrån data från kliniska praxis: valet av jämförelsegrupp. Vi använde koncept från utformningen av kliniska prövningar och drog slutsatsen att valet av jämförelsegrupp

beror på en rad faktorer, t ex om det finns patienter som använder liknande läkemedel som kan användas som jämförelsegrupp, om risken för biverkning tros variera mellan olika patientgrupper, och hur stort antal patienter som kan analyseras.

Sammanfattningsvis undersökte vi risker för biverkningar av mediciner hos barn med kroniska inflammatoriska sjukdomar. Vi visade att information om risker som kan användas vid förskrivning av läkemedel till barn kan tas fram utifrån data från klinisk praxis i Skandinavien. På grund av den fortsatt stora bristen på denna typ av information hoppas vi att detta arbete ska inspirera till fler studier i framtiden.

9 ACKNOWLEDGEMENTS

Carrying out this project would not have been possible without the tireless support, strong commitment and vigorous encouragement by many people. Some of them are mentioned below. Thank you!

Björn Pasternak, my main supervisor. Thank you for opening the door and making this work more rewarding and addictive than I had ever imagined. Thanks for being the encouraging, patient, and skilled person you are: it has been a blast and it changed my professional life.

Henrik Svanström, my co-supervisor. Thanks for all the guidance and support over the years, for rewarding discussions, and for showing me the ropes: these analyses would have been in a sad state without you.

Jonas F. Ludvigsson, my co-supervisor. Thank you for spreading your contagious energy and work ethic to this project, for encouragements, for minuscule email response times, and for sharing your expertise in pediatrics.

Martin Kulldorff, co-author and scientific role model. Thank you for invaluable guidance, for generously offering your time and connections, and for sharing your philosophy on the symbiosis between technical and non-technical epidemiologists.

Johan Askling, research group leader. Thank you for taking me on, for creating such a vibrant research environment, and for all the support and good advice over the years.

I wish to thank the co-authors of this work for their important contributions; *Ola Olén*: Thanks for rewarding collaborations, your committed feedback, and for sharing your expertise in pediatric IBD. *Mads Melbye*: Thanks for the support and for facilitating a binational collaboration.

Being part of a great pharmacoepi team I have worked on many studies outside of this thesis project during the last four years. I have had the pleasure to work with many brilliant researchers and I have learned tremendously from these studies. Many thanks to *Peter Ueda*, *Donna Wang*, *Elisabeth Dahlqwist, Arvid Engström, Martin Neovius, Olof Stephansson* and *Benjamin Flam*.

Andrei Barbulescu, my PhD comrade. Thanks for never-ending discussions and two-man strong workshops. I would have been stuck without you. *Thomas Frisell*, thank you for introducing me to a paper that turned out to influence a great deal of this work.

I wish to thank the current and former members of the Askling research group for making this such a joyful place to work, for support and rewarding discussions, for insightful feedback on ideas, drafts and analyses at all the steps of the way.

I am grateful to all the brilliant and dedicated fellow doctoral students at KEP for sharing this journey with me, for early mornings and late evenings, and for mutual support: Thank you!

Gudrun Bergman, my mentor. Thank you for sharing your experiences, for your support, and for giving me some useful perspectives.

Many thanks to *Aaron Levine* and *Kelsi Smith* for great comments and feedback on the thesis summary.

A special thanks to *Unga Lovande*, this fine congregation of young professionals.

Emelie, Juli, Mamma, Rolf – tack för allt stöd och för att ni finns.

Ana, my love and consiglieri. *Ella*, min skatt. Hvala ti za vso potrpežljivost na tej poti. Rad te imam, bolj kot vse ostalo.

10 REFERENCES

1. American Academy of Pediatrics. Committee on Drugs. Guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. *Pediatrics.* 1977;60(1):91-101.
2. Paine MF. Therapeutic disasters that hastened safety testing of new drugs. *Clin Pharmacol Ther.* 2017;101(4):430-434.
3. Clark LT, Watkins L, Pina IL, et al. Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Curr Probl Cardiol.* 2019;44(5):148-172.
4. Steinbrook R. Testing medications in children. *The New England Journal of Medicine.* 2002;347(18).
5. Sachs AN, Avant D, Lee CS, Rodriguez W, Murphy MD. Pediatric Information in Drug Product Labeling. *JAMA.* 2012;307(18):1914-1915.
6. Steinbrook R. Testing mediations in children. *The New England Journal of Medicine.* 2002;347(18).
7. Field MJ, Behrman RE. Ethical Conduct of Clinical Research Involving Children. In: Field MJ, Behrman RE, eds. Washington (DC)2004.
8. Council of Canadian Academies. Improving Medicines for Children in Canada. 2014;Ottawa (ON): The Expert Panel on Therapeutic Products for Infants, Children, and Youth, Council of Canadian Academies.
9. Olsson J, Kimland E, Pettersson S, Odlind V. Paediatric drug use with focus on off-label prescriptions in Swedish outpatient care--a nationwide study. *Acta Paediatr.* 2011;100(9):1272-1275.
10. Kimland E, Odlind V. Off-label drug use in pediatric patients. *Clin Pharmacol Ther.* 2012;91(5):796-801.
11. Turner MA, Catapano M, Hirschfeld S, Giaquinto C, Global Research in P. Paediatric drug development: the impact of evolving regulations. *Adv Drug Deliv Rev.* 2014;73:2-13.
12. Bourgeois FT, Kesselheim AS. Promoting Pediatric Drug Research and Labeling — Outcomes of Legislation. *N Engl J.* 2019;381(9):875-881.
13. Crockett LK, Okoli GN, Neilson CJ, Rabbani R, Abou-Setta AM, Klassen TP. Publication of Randomized Clinical Trials in Pediatric Research: A Follow-up Study. *JAMA Netw Open.* 2018;1(1):e180156.
14. Luo X, Doherty J, Cappelleri JC, Frush K. Role of pharmacoepidemiology in evaluating prescription drug safety in pediatrics. *Curr Med Res Opin.* 2007;23(11):2607-2615.
15. McMahon AW, Dal Pan G. Assessing Drug Safety in Children — The Role of Real-World Data. *New England Journal of Medicine.* 2018;378(23):2155-2157.
16. Osokogu OU, Dukanovic J, Ferrajolo C, et al. Pharmacoepidemiological safety studies in children: a systematic review. *Pharmacoepidemiol Drug Saf.* 2016;25(8):861-870.

17. Lasky T, Artaman A, Czaja AS, et al. Current needs in pediatric pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2016;25(6):738-742.

18. Everhov AH, Halfvarson J, Myrelid P, et al. Incidence and Treatment of Patients Diagnosed With Inflammatory Bowel Diseases at 60 Years or Older in Sweden. *Gastroenterology.* 2018;154(3):518-528 e515.

19. Benchimol EI, Fortinsky KJ, Gozdyra P, Van den Heuvel M, Van Limbergen J, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflamm Bowel Dis.* 2011;17(1):423-439.

20. Kolho KL, Ainamo A. Progress in the treatment and outcome of pediatric inflammatory bowel disease patients. *Expert Rev Clin Immunol.* 2016;12(12):1337-1345.

21. Ruemmele FM, Turner D. Differences in the management of pediatric and adult onset ulcerative colitis--lessons from the joint ECCO and ESPGHAN consensus guidelines for the management of pediatric ulcerative colitis. *J Crohns Colitis.* 2014;8(1):1-4.

22. Ardizzone S, Maconi G, Russo A, Imbesi V, Colombo E, Bianchi Porro G. Randomised controlled trial of azathioprine and 5-aminosalicylic acid for treatment of steroid dependent ulcerative colitis. *Gut.* 2006;55(1):47-53.

23. Markowitz J, Grancher K, Kohn N, Lesser M, Daum F. A multicenter trial of 6-mercaptopurine and prednisone in children with newly diagnosed Crohn's disease. *Gastroenterology.* 2000;119(4):895-902.

24. Chande N, Patton PH, Tsoulis DJ, Thomas BS, MacDonald JK. Azathioprine or 6-mercaptopurine for maintenance of remission in Crohn's disease. *Cochrane Database Syst Rev.* 2015(10):CD000067.

25. Ruemmele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *J Crohns Colitis.* 2014;8(10):1179-1207.

26. Petty RE, Southwood TR, Manners P, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol.* 2004;31(2):390–392.

27. Thierry S, Fautrel B, Lemelle I, Guillemin F. Prevalence and incidence of juvenile idiopathic arthritis: a systematic review. *Joint Bone Spine.* 2014;81(2):112-117.

28. Beukelman T, Patkar NM, Saag KG, et al. 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: initiation and safety monitoring of therapeutic agents for the treatment of arthritis and systemic features. *Arthritis Care Res (Hoboken).* 2011;63(4):465-482.

29. Stoll ML, Cron RQ. Treatment of juvenile idiopathic arthritis: a revolution in care. *Pediatr Rheumatol Online J.* 2014;12:13.

30. Giancane G, Consolaro A, Lanni S, Davi S, Schiappapietra B, Ravelli A. Juvenile Idiopathic Arthritis: Diagnosis and Treatment. *Rheumatol Ther.* 2016;3(2):187-207.

31. Giannini EH, Brewer EJ, N K. Methotrexate in resistant juvenile rheumatoid arthritis. Results of the USA-USSR double-blind, placebo-controlled trial. The Pediatric Rheumatology Collaborative Study Group and The Cooperative Children's Study Group. *N Engl J Med.* 1992;326(16):1043–1049.

32. Hyams J, Crandall W, Kugathasan S, et al. Induction and maintenance infliximab therapy for the treatment of moderate-to-severe Crohn's disease in children. *Gastroenterology.* 2007;132(3):863-873; quiz 1165-1166.

33. Ruemmele FM, Lachaux A, Cezard JP, et al. Efficacy of infliximab in pediatric Crohn's disease: a randomized multicenter open-label trial comparing scheduled to on demand maintenance therapy. *Inflamm Bowel Dis.* 2009;15(3):388-394.

34. Hyams J, Damaraju L, Blank M, et al. Induction and maintenance therapy with infliximab for children with moderate to severe ulcerative colitis. *Clin Gastroenterol Hepatol.* 2012;10(4):391-399 e391.

35. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health.* 2011;11:450.

36. Ruperto N, Lovell DJ, Cuttica R, et al. A randomized, placebo-controlled trial of infliximab plus methotrexate for the treatment of polyarticular-course juvenile rheumatoid arthritis. *Arthritis Rheum.* 2007;56(9):3096-3106.

37. Cosnes J, Bourrier A, Laharie D, et al. Early administration of azathioprine vs conventional management of Crohn's Disease: a randomized controlled trial. *Gastroenterology.* 2013;145(4):758-765 e752; quiz e714-755.

38. Panes J, Lopez-Sanroman A, Bermejo F, et al. Early azathioprine therapy is no more effective than placebo for newly diagnosed Crohn's disease. *Gastroenterology.* 2013;145(4):766-774 e761.

39. Timmer A, Patton PH, Chande N, McDonald JW, MacDonald JK. Azathioprine and 6-mercaptopurine for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev.* 2016(5):CD000478.

40. Teich N, Mohl W, Bokemeyer B, et al. Azathioprine-induced Acute Pancreatitis in Patients with Inflammatory Bowel Diseases--A Prospective Study on Incidence and Severity. *J Crohns Colitis.* 2016;10(1):61-68.

41. Hyams JS, Lerer T, Mack D, et al. Outcome following thiopurine use in children with ulcerative colitis: a prospective multicenter registry study. *Am J Gastroenterol.* 2011;106(5):981-987.

42. Jaspers GJ, Verkade HJ, Escher JC, Taminiau JAJM, Rings EHHM. Azathioprine Maintains First Remission in Newly Diagnosed Pediatric Crohn's Disease. *Inflamm Bowel Dis.* 2006;12(9):831-836.

43. Hindorf U, Lindqvist M, Hildebrand H, Fagerberg U, Almer S. Adverse events leading to modification of therapy in a large cohort of patients with inflammatory bowel disease. *Aliment Pharmacol Ther.* 2006;24(2):331-342.

44. Riello L, Talbotec C, Garnier-Lengline H, et al. Tolerance and efficacy of azathioprine in pediatric Crohn's disease. *Inflamm Bowel Dis.* 2011;17(10):2138-2143.

45. Singh JA, Wells GA, Christensen R, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. *Cochrane Database Syst Rev.* 2011(2):CD008794.

46. Singh JA, Cameron C, Noorbaloochi S, et al. Risk of serious infection in biological treatment of patients with rheumatoid arthritis: a systematic review and meta-analysis. *The Lancet.* 2015;386(9990):258-265.

47. Bonovas S, Fiorino G, Allocca M, et al. Biologic Therapies and Risk of Infection and Malignancy in Patients With Inflammatory Bowel Disease: A Systematic Review and Network Meta-analysis. *Clin Gastroenterol Hepatol.* 2016;14(10):1385-1397 e1310.

48. Yun H, Xie F, Delzell E, et al. Comparative Risk of Hospitalized Infection Associated With Biologic Agents in Rheumatoid Arthritis Patients Enrolled in Medicare. *Arthritis Rheumatol.* 2016;68(1):56-66.

49. Lichtenstein GR, Feagan BG, Cohen RD, et al. Serious infection and mortality in patients with Crohn's disease: more than 5 years of follow-up in the TREAT registry. *Am J Gastroenterol.* 2012;107(9):1409-1422.

50. Kirchgesner J, Lemaitre M, Carrat F, Zureik M, Carbonnel F, Dray-Spira R. Risk of Serious and Opportunistic Infections Associated With Treatment of Inflammatory Bowel Diseases. *Gastroenterology.* 2018;155(2):337-346 e310.

51. Lee WJ, Lee TA, Calip GS, Suda KJ, Briars L, Schumock GT. Risk of Serious Bacterial Infection Associated With Tumor Necrosis Factor-Alpha Inhibitors in Children and Young Adults With Inflammatory Bowel Disease. *Inflamm Bowel Dis.* 2018;24(4):883-891.

52. Askling J, Fored CM, Brandt L, et al. Time-dependent increase in risk of hospitalisation with infection among Swedish RA patients treated with TNF antagonists. *Ann Rheum Dis.* 2007;66(10):1339-1344.

53. Nyboe Andersen N, Pasternak B, Friis-Moller N, Andersson M, Jess T. Association between tumour necrosis factor-alpha inhibitors and risk of serious infections in people with inflammatory bowel disease: nationwide Danish cohort study. *BMJ.* 2015;350:h2809.

54. Grijalva CG, Chen L, Delzell E, et al. Initiation of tumor necrosis factor-alpha antagonists and the risk of hospitalization for infection in patients with autoimmune diseases. *JAMA.* 2011;306(21):2331-2339.

55. Klotsche J, Niewerth M, Haas JP, et al. Long-term safety of etanercept and adalimumab compared to methotrexate in patients with juvenile idiopathic arthritis (JIA). *Ann Rheum Dis.* 2016;75(5):855-861.

56. Davies R, Southwood TR, Kearsley-Fleet L, et al. Medically Significant Infections Are Increased in Patients With Juvenile Idiopathic Arthritis Treated With Etanercept: Results From the British Society for Paediatric and Adolescent Rheumatology Etanercept Cohort Study. *Arthritis Rheumatol.* 2015;67(9):2487-2494.

57. Nagy A, Matrai P, Hegyi P, et al. The effects of TNF-alpha inhibitor therapy on the incidence of infection in JIA children: a meta-analysis. *Pediatr Rheumatol Online J.* 2019;17(1):4.

58. Dulai PS, Thompson KD, Blunt HB, Dubinsky MC, Siegel CA. Risks of serious infection or lymphoma with anti-tumor necrosis factor therapy for pediatric inflammatory bowel disease: a systematic review. *Clin Gastroenterol Hepatol.* 2014;12(9):1443-1451; quiz e1488-1449.

59. Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol.* 2016;31(2):125-136.

60. Pedersen CB, Gotzsche H, Moller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Dan Med Bull.* 2006;53(4):441-449.

61. Schmidt M, Pedersen L, Sorensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol.* 2014;29(8):541-549.

62. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol.* 2009;24(11):659-667.

63. Sundhetsstyrelsen. Fokusrapport - Overvågning af bivirkninger hos børn med autoimmune sygdomme i biologisk behandling. 2012.

64. Larsen L, Jensen MD, Larsen MD, et al. The Danish National Registry for Biological Therapy in Inflammatory Bowel Disease. *Clin Epidemiol.* 2016;8:607-612.

65. Ekbom A. The Swedish Multi-generation Register. *Methods Mol Biol.* 2011;675:215-220.

66. Andersen TF, Madsen M, Jorgensen J, Mellemkjoer L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull.* 1999;46(3):263-268.

67. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449-490.

68. Wettermark B, Hammar N, Fored CM, et al. The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf.* 2007;16(7):726-735.

69. Wallerstedt SM, Wettermark B, Hoffmann M. The First Decade with the Swedish Prescribed Drug Register – A Systematic Review of the Output in the Scientific Literature. *Basic Clin Pharmacol Toxicol.* 2016;119(5):464-469.

70. Kildemoes HW, Sorensen HT, Hallas J. The Danish National Prescription Registry. *Scand J Public Health.* 2011;39(7 Suppl):38-41.

71. Ueda P, Svanstrom H, Melbye M, et al. Sodium glucose cotransporter 2 inhibitors and risk of serious adverse events: nationwide register based cohort study. *BMJ.* 2018;363:k4365.

72. Svanstrom H, Lund M, Melbye M, Pasternak B. Use of Proton Pump Inhibitors and the Risk of Acute Kidney Injury Among Patients with Rheumatoid Arthritis: Cohort Study. *Drug Saf.* 2018.

73. Ray WA, Murray KT, Hall K, Arbogast PG, Stein M. Azithromycin and the Risk of Cardiovascular Death. *N Engl J Med.* 2012;366(20):1881-1890.

74. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf.* 2017;26(4):459-468.

75. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764.

76. Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-75.

77. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res.* 2011;46(3):399-424.

78. Robins JM, Hernán MA, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology.* 2000;11(5):550-560.

79. Hernán MA, Brumback B, Robins JM. Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. *Epidemiology.* 2000;11(5):561-570.

80. Wang SV, Maro JC, Baro E, et al. Data Mining for Adverse Drug Events With a Propensity Score-matched Tree-based Scan Statistic. *Epidemiology.* 2018;29(6):895-903.

81. Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf.* 2013;22(5):517-523.

82. Floyd, A. Risk of acute pancreatitis in users of azathioprine: a population-based case–control study. *The American Journal of Gastroenterology.* 2003;98(6):1305-1308.

83. Bermejo F, Lopez-Sanroman A, Taxonera C, et al. Acute pancreatitis in inflammatory bowel disease, with special reference to azathioprine-induced pancreatitis. *Aliment Pharmacol Ther.* 2008;28(5):623-628.

84. Chaparro M, Ordas I, Cabre E, et al. Safety of thiopurine therapy in inflammatory bowel disease: long-term follow-up study of 3931 patients. *Inflamm Bowel Dis.* 2013;19(7):1404-1410.

85. Aloi M, D'Arcangelo G, Bramuzzo M, et al. Effect of Early Versus Late Azathioprine Therapy in Pediatric Ulcerative Colitis. *Inflamm Bowel Dis.* 2016;22(7):1647-1654.

86. D'Arcy M, Sturmer T, Lund JL. The importance and implications of comparator selection in pharmacoepidemiologic research. *Curr Epidemiol Rep.* 2018;5(3):272-283.

87. Lund JL, Richardson DB, Sturmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep.* 2015;2(4):221-228.

88. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf.* 2010;19(8):858-868.

89. Ardizzone S, Maconi G, Sampietro GM, et al. Azathioprine and mesalamine for prevention of relapse after conservative surgery for Crohn's disease. *Gastroenterology.* 2004;127(3):730-740.

90. Pedersen JE, Ängquist LH, Jensen CB, Kjærgaard JS, Jess T, Allin KH. Risk of pancreatitis in patients with inflammatory bowel disease - a meta-analysis. *Danish medical journal.* 2020;67(3).

91. Ledder O, Lemberg DA, Day AS. Thiopurine-induced pancreatitis in inflammatory bowel diseases. *Expert Rev Gastroenterol Hepatol.* 2015;9(4):399-403.

92. Moon W, Loftus EV, Jr. Review article: recent advances in pharmacogenetics and pharmacokinetics for safe and effective thiopurine therapy in inflammatory bowel disease. *Aliment Pharmacol Ther.* 2016;43(8):863-883.

93. Horneff G. Biologic-associated infections in pediatric rheumatology. *Curr Rheumatol Rep.* 2015;17(11):66.

94. Dabritz J, Gerner P, Enninger A, Classen M, Radke M. Inflammatory Bowel Disease in Childhood and Adolescence. *Dtsch Arztebl Int.* 2017;114(19):331-338.

95. Church PC, Hyams J, Ruemmele F, de Ridder L, Turner D, Griffiths AM. The Continental Divide: Anti-TNF Use in Pediatric IBD Is Different in North America Compared to Other Parts of the World. *Can J Gastroenterol Hepatol.* 2018;2018:3190548.

96. van Hoeve K, Vermeire S. Thiopurines in Pediatric Inflammatory Bowel Disease: Current and Future Place. *Paediatr Drugs.* 2020;22(5):449-461.

97. Moon JS. Clinical Aspects and Treatments for Pediatric Inflammatory Bowel Diseases. *Pediatr Gastroenterol Hepatol Nutr.* 2019;22(1):50-56.

98. Ray WA. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. *Am J Epidemiol.* 2003;158(9):915–920.

99. Danaei G, Tavakkoli M, Hernan MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. *Am J Epidemiol.* 2012;175(4):250-262.

100. Garcia-Albeniz X, Hsu J, Hernan MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol.* 2017.

101. Beukelman T, Xie F, Baddley JW, et al. The risk of hospitalized infection following initiation of biologic agents versus methotrexate in the treatment of juvenile idiopathic arthritis. *Arthritis Res Ther.* 2016;18(1):210.

102. Curtis JR, Danila MI, Chen L. Risk of Cardiovascular Outcomes among Psoriasis Patients Treated with Biologics and Other Systemic Agents. *J Psoriasis Psoriatic Arthritis.* 2016;1(3):128-137.

103. Hicks BM, Filion KB, Yin H, Sakr L, Udell JA, Azoulay L. Angiotensin converting enzyme inhibitors and risk of lung cancer: population based cohort study. *BMJ.* 2018;363:k4209.

104. Hicks BM, Yin H, Yu OHY, Pollak MN, Platt RW, Azoulay L. Glucagon-like peptide-1 analogues and risk of breast cancer in women with type 2 diabetes: population

based cohort study using the UK Clinical Practice Research Datalink. *BMJ.* 2016;355:i5340.

105. Kosiborod M, Cavender MA, Fu AZ, et al. Lower Risk of Heart Failure and Death in Patients Initiated on Sodium-Glucose Cotransporter-2 Inhibitors Versus Other Glucose-Lowering Drugs: The CVD-REAL Study (Comparative Effectiveness of Cardiovascular Outcomes in New Users of Sodium-Glucose Cotransporter-2 Inhibitors). *Circulation.* 2017;136(3):249-259.

106. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol.* 2010;171(6):674-677; discussion 678-681.

107. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.

108. Cepeda MS. Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders. *American Journal of Epidemiology.* 2003;158(3):280-287.

109. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med.* 2015;34(28):3661-3679.

110. Gruber S, Logan RW, Jarrin I, Monge S, Hernan MA. Ensemble learning of inverse probability weights for marginal structural modeling in large observational datasets. *Stat Med.* 2015;34(1):106-117.

111. Blakely T, Lynch J, Simons K, Bentley R, Rose S. Reflection on modern methods: when worlds collide-prediction, machine learning and causal inference. *Int J Epidemiol.* 2021;49(6):2058-2064.

112. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology. *Am J Epidemiol.* 2002;155(2):176-184.

113. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* 2009;20(4):512-522.

114. VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol.* 2019;34(3):211-219.

115. Franklin JM, Eddings W, Glynn RJ, Schneeweiss S. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. *Am J Epidemiol.* 2015;182(7):651-659.

116. Wyss R, Schneeweiss S, van der Laan M, Lendle SD, Ju C, Franklin JM. Using Super Learner Prediction Modeling to Improve High-dimensional Propensity Score Estimation. *Epidemiology.* 2018;29(1):96-106.

117. Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M bias in epidemiologic studies: a simulation study. *Am J Epidemiol.* 2012;176(10):938-948.

118. Schuler MS, Rose S. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *Am J Epidemiol.* 2017;185(1):65-73.

119. King G, Nielsen R. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis.* 2019;27(4):435-454.

120. Ripollone JE, Huybrechts KF, Rothman KJ, Ferguson RE, Franklin JM. Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology. *Am J Epidemiol.* 2018;187(9):1951-1961.

121. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ, Jr. Selection Bias Due to Loss to Follow Up in Cohort Studies. *Epidemiology.* 2016;27(1):91-97.

122. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. *J Am Med Inform Assoc.* 2010;17(2):169-177.

123. Wang SV, Maro JC, Gagne JJ, et al. A General Propensity Score for Signal Identification using Tree-Based Scan Statistics. *Am J Epidemiol.* 2021.