From the Department of Medicine, Solna
Clinical Epidemiology Division
Karolinska Institutet, Stockholm, Sweden

# PHARMACOEPIDEMIOLOGICAL STUDIES OF RHEUMATOID ARTHRITIS – Methodological Considerations and Applications

Andrei Barbulescu


Karolinska Institutet

Stockholm 2022

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN 978-91-8016-724-6
Cover illustration: by Andrei Barbulescu

# Pharmacoepidemiological Studies of Rheumatoid Arthritis – Methodological Considerations and Applications

## THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Andrei Barbulescu**

The thesis will be defended in public at Stockholm, Solna, Karolinska University Hospital, Eugeniavägen 27, Norrbacka S2:01, Rehabsalen, on the 2nd of September 2022 at 9:00

*Principal Supervisor:*
Docent Thomas Frisell
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

*Co-supervisor(s):*
Professor Johan Askling
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

Dr Bénédicte Delcoigne
Karolinska Institutet
Department of Medicine, Solna
Division of Clinical Epidemiology

*Opponent:*
Assistant professor Maurizio Sessa
University of Copenhagen
Department of Drug Design and Pharmacology

*Examination Board:*
Docent Anna Johansson
Karolinska Institutet
Department of Medical Epidemiology and Biostatistics

Adjunct Professor Meliha C Kapetanovic
Lund University
Department of Clinical Sciences
Division of Rheumatology

Professor Maria Feychting
Karolinska Institutet
Institute of Environmental Medicine

To Nicoleta and my family

# POPULAR SCIENCE SUMMARY OF THE THESIS

The four articles which make up this thesis describe studies about harmful and beneficial effects of various treatments for rheumatoid arthritis. Rheumatoid arthritis is a chronic disease of the immune system, which affects close to one out of a hundred persons in Sweden, and mainly manifests itself by gradually destroying the patient's joints. To prevent irreversible joint damage and loss of function, early treatment with disease modifying anti-rheumatic drugs (DMARDs) is recommended. The treatment starts with conventional synthetic DMARDs, and patients who do not respond to these, switch to biological DMARDs or novel synthetic DMARDs called JAK inhibitors. It may take some time before the effects of DMARDs appear, but pain and inflammation can be relieved with glucocorticoids and non-steroidal anti-inflammatory drugs in the meantime.

In the first study, we compared the risk of perforation of the intestine between different DMARDs, motivated by a concern that one of the newer biological DMARDs, tocilizumab, might increase this risk. We found that the proportion of patients who experienced intestinal perforations, among those treated with tocilizumab, was roughly double that in other DMARD treatment groups. The risk was low (around 4 intestinal perforations in 1000 patients treated with tocilizumab per year), but, since the consequences of intestinal perforations can be serious (bacteria could spill out from the gut, into the blood stream, causing sepsis), patients should be carefully evaluated before initiating treatment with tocilizumab, and controlled during treatment.

In the second study, we compared the effectiveness of JAK inhibitors, to that of biological DMARDs. We mainly had data about one of the JAK inhibitors, called baricitinib, which is preferred in Sweden, and we found that it is at least as effective as biological DMARDs. JAK inhibitors are also more convenient to administer, since they are pills that can be swallowed, as opposed to biological DMARDs, which need to be injected. On the other hand, results from a new randomized controlled trial showed that JAK inhibitors may increase the risk of cancer and cardiovascular disease in older patients, compared to some biological DMARDs. Thus, despite similar effectiveness and more convenient administration of JAK inhibitors, biological DMARDs with better known risk profile may be the preferred initial choice, reserving JAK inhibitors for patients who responded poorly to these biological DMARDs.

In the third study, we tried to mimic the design of a randomized controlled trial in a study using data from real clinical practice, called an observational study, and we compared the results of the observational study with those of the trial. The purpose of this comparison was to test if our observational study could provide correct results, since the results of randomized controlled trials are generally accepted as the true effects of treatments. The main advantage of trials is that the studied treatments are given at random to trial participants, making the groups of participants who receive different treatments similar on average. In real clinical practice, treatments are given considering the patient's characteristics. For example, patients who end up receiving a new treatment, perceived as more efficient, may be the ones that are sicker

compared to patients receiving the conventional treatment. Since sicker patients may have worse outcomes regardless of what treatment they receive, comparisons between patients receiving the new and the conventional treatment in real clinical practice are unfair. Statistical methods are used to correct this unfairness (bias) in observational studies, but these are not as effective as randomization. Nonetheless, some authors argue that, besides randomization, trials have other virtues that could benefit observational studies. Indeed, despite the lack of randomization, once we designed our observational study to mimic the target trial, we obtained very similar results to those of the trial. Therefore, our findings indicate that observational studies designed to resemble trials could produce trustworthy results about the effects of medical therapies.

Finally, in the fourth study we assessed the risk of serious infections associated with the use of glucocorticoids. Glucocorticoids are synthetic drugs related to a hormone produced by our own bodies called cortisol which, among other effects, inhibits the immune system, thus controlling inflammation and pain. Drugs that inhibit the immune system may also increase the risk of infection, but how much glucocorticoids increase this risk, and how this depends on the dose and length of treatment is still debated by rheumatologists. The treatment with glucocorticoids for rheumatoid arthritis is very dynamic, the drug being started at a higher dose, which is reduced over time to a stop, and restarted if the disease activity cannot be properly controlled with other drugs. Such a changing treatment, in response to a changing disease activity and other factors, is complicated to study, and may be responsible for the uncertainty surrounding the effects of glucocorticoids. We used a new technique for analyzing the effects of such changing treatments, and we observed that the risk of serious infections increased with the dose and the duration of glucocorticoid use. However, the risk increase was small for low doses (less than 10 mg prednisone daily) used for up to one year, compared to no use.

In summary, our results suggest that: tocilizumab increases the risk of intestinal perforations, thus patients should be carefully evaluated before initiating treatment with tocilizumab, and those who initiate this treatment should be closely monitored during therapy; JAK inhibitors are an effective option for treating rheumatoid arthritis but, in light of new information about their safety, they may be reserved for patients who have failed more established biological DMARDs; and the risk of serious infections increases gradually with the dose and duration of glucocorticoids use, thus reducing the durations of treatment and the dose to the minimum effective is advisable. Finally, designing observational studies to mirror target trials may improve their ability to correctly determine the effects of drug treatments.

# ABSTRACT

*Study I – Biological disease-modifying anti-rheumatic drugs (DMARDs) and the risk of gastro-intestinal perforations*

Study I was motivated by previous signals of an increased risk of lower gastro-intestinal perforations among rheumatoid arthritis (RA) patients treated with tocilizumab.

The primary aim of study I was to compare the incidence of lower gastro-intestinal perforations between RA patients initiating the biological DMARDs tocilizumab, abatacept, rituximab and tumor necrosis factor inhibitors (TNFi). Secondary comparisons with bionaïve RA patients and general population controls were made.

We designed a cohort study which included RA patients identified in the Swedish National Patients Register (NPR) and we identified biological DMARD treatments in the Swedish Rheumatology Quality Register (SRQ). General population controls, matched by sex, age and location to biological DMARD treated RA patients, were available. The main outcome was hospitalization or death due to lower gastro-intestinal perforations identified using a prespecified list of International Classification of Disease, 10$^{th}$ revision (ICD-10) codes in the NPR and the Swedish Causes of Death Register.

In line with previous studies, we observed an increased risk of lower gastro-intestinal perforations among patients treated with tocilizumab compared to patients treated with TNFi (hazard ratio of 2.2, 95% confidence interval 1.3 to 3.8), and there was no evidence of an increased risk among patients treated with abatacept or rituximab. Also, compared to general population controls, only RA patients treated with tocilizumab had an increased risk of lower gastro-intestinal perforations after adjustment for sex, age and baseline glucocorticoids use.

The absolute risk of lower gastro-intestinal perforations was low even under treatment with tocilizumab (~4 events /1000 person-years), but considering the potential for serious complications, the presence of additional risk factors, such as older age and use of glucocorticoids, should be evaluated before deciding to initiate tocilizumab, and patients should be monitored for diverticulitis and lower gastro-intestinal perforations during treatment.

*Study II – Comparative effectiveness of baricitinib, tofacitinib and biological DMARDs in RA*

Study II was motivated by a lack of evidence for the relative effectiveness of the Janus Kinase inhibitor (JAKi) baricitinib compared to non-TNFi biological DMARDs.

The aim of study II was to compare the effectiveness of the JAKis baricitinib and tofacitinib with that of biological DMARDs.

Study II was a cohort study which included RA patients who initiated baricitinib, tofacitinib, abatacept, interleukin-6 inhibitors (IL-6i), rituximab, and TNFi, as identified in SRQ. In the primary analysis, these patients were followed for one year from treatment initiation, at the end

of which the proportions of treatment responders were evaluated using to the following measures: EULAR disease activity score assessed on 28 joints (DAS28) good response, health assessment questionnaire disability index (HAQ-DI) improvement > 0.2 units compared to baseline, and clinical disease activity index (CDAI) remission. Patients who discontinued treatment before one year were classified as "non-responders". Improvements at three-months compared to baseline in DAS28, HAQ-DI, and CDAI, as well as drug retention over follow-up were also assessed.

After confounding adjustment, one-year treatment response proportions were consistently higher on baricitinib compared to TNFi, even though the differences were small. Comparisons with non-TNFi biological DMARDs also favored baricitinib, but not consistently. There was no evidence that response proportions on tofacitinib were different from those on baricitinib or biological DMARDs, but the sample of tofacitinib treated patients was small, limiting precision. Drug retention was significantly higher on baricitinib compared to alternatives, and the magnitude of three-months improvements followed a similar pattern to one-year treatment responses.

In conclusion, our results show that baricitinib and tofacitinib are at least as effective as biological DMARDs for treating RA.

*Study III – Emulation of the SWEFOT trial in observational data*

Study III was motivated by a lack of confidence in the comparative effectiveness evidence generated by observational studies, which could prove a valuable complement to randomized controlled trial (RCT) comparative efficacy evidence. It has been suggested that designing observational studies to mimic RCTs may reduce bias. A sensible approach for testing the validity of trial emulations in observational data is to emulate an existing trial protocol, and then compare the results.

Therefore, the aim of study III was to emulate the protocol of the Swedish Farmacotherapy (SWEFOT) pragmatic trial in an observational study including a non-overlapping sample of participants coming from the same source population (Swedish RA patients) and to compare the results. In SWEFOT, methotrexate (MTX) insufficient responders were randomized to receive additional infliximab or sulfasalazine (SSZ) + hydroxychloroquine (HCQ). In the observational study, we included RA patients initiating infliximab (N = 313) or SSZ + HCQ (N = 196) after MTX, identified using data from SRQ and the Prescribed Drugs Register, and mimicking the SWEFOT eligibility criteria. The primary outcome was the proportion of EULAR DAS28 good responders at 9 months, classifying patients who discontinued treatment as non-responders.

The proportions of responders in the observational emulation were comparable to those in SWEFOT: 39% (vs. 39% in SWEFOT) for infliximab and 28% (vs. 25%) for SSZ + HCQ. The crude observed response ratio was 1.39 (95% confidence interval: 1.04 to 1.86), increasing to

1.48 (95% confidence interval 0.98 to 2.24) after confounding adjustment, compared to 1.59 (95% confidence interval 1.10 to 2.30) in SWEFOT.

Thus, by designing our observational study to emulate SWEFOT, we could closely replicate the trial results, favoring infliximab over SSZ + HCQ combination therapy at 9 months.

*Study IV – Glucocorticoid exposure and the risk of serious infections in RA*

Study IV was motivated by the apparent disagreement between RCT and observational study results regarding the risk of serious infections associated with the use of glucocorticoids in RA, revealed in a meta-analysis. We hypothesized that the conflicting results could be explained by improper representations of a time-varying exposure and improper confounding adjustment, in previous observational studies.

Thus, in study IV we aimed to contrast the incidence of hospitalization for infections (serious infections) between different oral glucocorticoid time-varying dose histories over three years, adjusting for time-varying confounding and selection bias using inverse probability weighting (IPW).

We identified 9639 patients newly diagnosed with RA in SRQ and followed them for three years after their first rheumatology visit. To allow the exposure to vary over time, and to adjust for time-varying confounding, each participant's follow-up was divided into 90-day periods. The average daily-dose of dispensed oral prednisone was calculated in each period, categorizing it into "no use", "low" (≤ 10mg/day) and "high" (> 10mg/day) doses. Time-varying confounders were measured before each exposure period. The incidence of serious infections over follow-up was modelled by pooled logistic regression, allowing separate effects for different periods of the exposure history.

An increased incidence rate of serious infections was associated with higher glucocorticoid doses and the association was stronger for more recent compared to past exposure. Compared to no glucocorticoids, exposure to low doses during the first year added 1.8 serious infection cases per 100 patients (95% confidence interval 0.8 to 2.8) at three years, while exposure to high doses added 4.1 (95% confidence interval 2.5 to 5.8) cases.

Hence, our results broadly agree with previous observational studies showing a dose dependent increased risk of infections associated with (recent) use of oral glucocorticoids.

# LIST OF SCIENTIFIC PAPERS

I. **Gastrointestinal perforations in patients with rheumatoid arthritis treated with biological disease-modifying antirheumatic drugs in Sweden: A nationwide cohort study**
*Barbulescu A, Delcoigne B, Askling J, Frisell T – RMD open 2020 (PMID: 32669452)*

II. **Effectiveness of baricitinib and tofacitinib compared with biological disease-modifying antirheumatic in rheumatoid arthritis: Results from a cohort study using nationwide Swedish register data**
*Barbulescu A, Askling J, Chatzidionysiou K, Forsblad-dElia H, Kastbom A, Lindstrom U, Turesson C, Frisell T – Rheumatology 2022 (PMID: 35134119)*

III. **Combined conventional synthetic disease modifying therapy vs. infliximab for rheumatoid arthritis: Emulating a randomized trial in observational data**
*Barbulescu A, Askling J, Saevarsdottir S, Kim SC, Frisell T – Clinical Pharmacology & Therapeutics. 2022 (PMID: 35652244)*

IV. **Glucocorticoid exposure and the risk of serious infection in rheumatoid arthritis: A marginal structural model application on real-world data**
*Barbulescu A, Sjölander A, Delcoigne B, Askling J, Frisell T – Manuscript*

# CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ACR | American College of Rheumatology |
| ATC | Anatomical Therapeutic Chemical Classification System |
| ATE | Average Treatment Effect |
| ATT | Average Treatment effect among the Treated |
| CDAI | Clinical Disease Activity Index |
| CI | Confidence Interval |
| CRP | C-Reactive Protein |
| DAG | Directed Acyclic Graph |
| DAS28 | Disease Activity Score evaluating 28 joints |
| DMARD | Disease Modifying Anti-Rheumatic Drug |
| bDMARD | Biological DMARD |
| csDMARD | Conventional Synthetic DMARD |
| tsDMARD | Targeted Synthetic DMARD |
| ESR | Erythrocyte Sedimentation Factor |
| EU | European Union |
| EULAR | European League Against Rheumatism (recently renamed European Alliance of Associations for Rheumatology) |
| FCS | Fully Conditional Specification (multiple imputation method) |
| GC | Glucocorticoids |
| GDPR | European personal data protection legislation |
| GI | Gastrointestinal |
| HAQ-DI | Health Assessment Questionnaire Disability Index |
| HR | Hazard Ratio |
| ICD | International statistical Classification of Diseases and related health problems |
| IL-6 | Interleukin 6 |
| IOSW | Inverse Odds of Sampling Weight |
| IPW | Inverse Probability Weight |
| IPCW | Inverse Probability of Censoring Weight |
| IPTW | Inverse Probability of Treatment Weight |

| | |
|---|---|
| JAKi | Janus Kinase inhibitor |
| MI | Multipe Imputation |
| MICE | Multiple Imputation using Chained Equations |
| MSM | Marginal Structural Model |
| MTX | Methotrexate |
| NOMESCO | Nordic Medico-Statistical Committee |
| NPR | Swedish National Patients Register |
| NSAID | Non-Steroidal Anti-Inflammatory Drug |
| OR | Odds Ratio |
| PDR | Swedish Prescribed Drugs Register |
| PPI | Proton Pump Inhibitor |
| RA | Rheumatoid Arthritis |
| RCT | Randomized Clinical Trial |
| SDAI | Simplified Disease Activity Index |
| SMR | Standardized Mortality Ratio |
| SRQ | Swedish Rheumatology Quality register |
| TNFi | Tumor Necrosis Factor inhibitor |
| UK | United Kingdom of Great Britain and Northern Ireland |
| US | United States of America |
| WHO | World Health Organization |

# 1 INTRODUCTION

## 1.1 RHEUMATOID ARTHRITIS

Rheumatoid arthritis (RA) is a chronic inflammatory disease primarily characterized by progressive joint damage with consequential reduced functional capacity (1,2), potentially accompanied by systemic manifestations (3). The estimated adult onset RA prevalence in Sweden is approximately 0.6% to 0.8% of the population (4).

In most cases, RA does not remit spontaneously (5), and requires early initiation of, usually, long-term treatment with the goal of slowing down disease progression towards remission (or at least low disease activity) to prevent loss of function (*treat-to-target* strategy) (6,7). To support this targeted treatment strategy, the disease activity and the functional ability of the patient needs to be quantified. Several composite measures that combine various clinical and laboratory parameters have been developed for this (8).

1) The Disease Activity Score (DAS) using 28-joint counts (DAS28) is calculated as a weighted sum of swollen joint counts (out of 28), tender joint counts (out of 28), an inflammation biomarker (either the C-reactive protein (CRP) or the erythrocyte sedimentation factor (ESR)), and a global assessment of the patient's disease activity and health state provided by the patient (measured on a visual analogue scale from 0 to 100) (9). DAS28 takes values between 0 and 9.4, and has validated thresholds for remission (< 2.6) and low disease activity (< 3.2) (10). However, because tender joint counts and inflammation markers receive higher weights, DAS28 defined remission states allow too many swollen joints, and treatments with strong effects on inflammation markers (such as Interleukin-6 (IL-6) receptor blockers and Janus Kinase inhibitors (JAKi)) have a disproportionate inflouence on DAS28 (11).
2) The Clinical Disease Activity Index (CDAI) is a simple sum of the swollen and tender joint counts (out of 28), and global assessments of the patient's disease activity provided by both patient and physician (measured on a visual analogue scale from 0 to 100) (8,9). Contrary to DAS28, the CDAI components are not weighted differently (all have weights of 1). Remission is defined as a CDAI ≤ 2.8, and it is considered that CDAI discriminates remission states better than DAS28 (11). Also, CDAI is a purely clinical index, which excludes laboratory values, making it more accessible for routine checks.
3) The Health Assessment Questionnaire (HAQ) is a standardized instrument meant to evaluate the self-reported global health state and functionality of patients. The HAQ has several components that address different dimensions of the patient's status, one of them being the Disability Index (HAQ-DI). The HAQ-DI contains 20 questions about several different daily motor activities that the patient can answer on a scale from 1 (no disability) to 3 (completely disabled) (12).

These composite indices are frequently employed as outcomes measures in RA comparative effectiveness studies, and, because they guide treatment choice, they are used as bias adjustment covariates in observational studies. As such, we have used them in all our studies.

## 1.2 THE PHARMACOTHERAPY OF RHEUMATOID ARTHRITIS

The RA pharmacotherapy landscape keeps evolving, with new agents and drugs classes having emerged in the last 30 years.

Traditionally, RA has been treated symptomatically with non-steroidal anti-inflammatory drugs (NSAIDs), other analgesics, and glucocorticoids, to subdue pain and inflammation. However, with the possible exception of glucocorticoids (13), these drugs do not control disease activity in order to prevent joint damage (14).

Medicines that, in addition to symptoms control, also retard joint damage and improve function are designated *Disease-Modifying Anti-Rheumatic Drugs* (DMARDs) and are essential for the treat-to-target strategy (15). Existing DMARDs are either synthetic molecules or peptides produced by biotechnology (called *biologics* and abbreviated bDMARDs). *Conventional synthetic* DMARDs (csDMARDs) entered clinical use via empirical, serendipitous discovery of their effect rather than by aiming at specific biological targets. Most commonly used csDMARDs are methotrexate (MTX), sulfasalazine (SSZ), antimalarials (such as hydroxychloroquine (HCQ)), leflunomide and gold salts. A newer class of synthetic molecules designed to interact with specific, known biological targets (even if this interaction elicits effects on multiple immune pathways) are called *targeted synthetic* DMARDs (tsDMARDs). All tsDMARDs currently used in the treatment of RA are Janus Kinase inhibitors (JAKi) (e.g. tofacitinib and baricitinib which are the subject of Study II) (16). Small synthetic molecules present some advantages over bDMARDs – they are not degraded during digestion, thus can be administered orally (17), and they do not elicit anti-drug antibodies (18). bDMARDs are mainly monoclonal antibodies which bind with high specificity to soluble or cell surface proteins (e.g. cytokines or cytokine receptors) (19). They play an important role in the modern treatment of RA, four of them also having biosimilar alternatives (i.e. molecules very similar to a reference bDMARD) (20). Their effects were investigated in the first three of our studies. Tumor Necrosis Factor inhibitors (TNFi) were the first bDMARDs approved for RA in the late 1990s. Five TNFi molecules are currently available, four monoclonal antibodies (infliximab, adalimumab, certolizumab pegol and golimumab) and a fusion protein (etanercept). They have slightly different mechanisms of action and modes of administration, but they all bind to and block TNF-α (21). In early 2000s, several bDMARDs with other targets than TNF-α were approved for rheumatoid arthritis. Abatacept is a fusion protein that binds to CD80/86 co-stimulatory proteins on antigen-presenting cells, blocking their interaction with the CD28 protein located on T-cells (22). As a result, T-cell activation and consequent cytokine production is hindered. Additionally, abatacept may also have direct effects on some antigen-presenting cells (23). Rituximab is a chimeric monoclonal antibody that targets the CD20 protein expressed on the surface of certain B-cells, leading to their transient depletion (24). For this reason, rituximab was initially approved for treating B-cell lymphomas, but was later

shown to also be effective in rheumatoid arthritis (25,26). Since B-cell depletion following an initial course consisting of two 1000 mg intravenous rituximab infusions, given two weeks apart, is long lasting, the need for re-treatment is evaluated only after a six months pause (27). Tocilizumab and sarilumab are monoclonal antibodies that block the action of IL-6 by binding to its soluble and membrane-bound receptors. IL-6 is a pleiotropic cytokine (meaning that it affects a large number of cells, immune and others), which plays an important role in mediating inflammation, thus in RA (28). Tocilizumab was the first IL-6 antagonist proved efficacious in the treatment of RA (29). Sarilumab has a greater affinity for the IL-6 receptors and a longer half-life compared to tocilizumab, which allows subcutaneous administration once every two weeks, an advantage over tocilizumab, which has to be administered weekly in its subcutaneous form (30). Both tocilizumab and sarilumab showed lower immunogenicity than other bDMARDs and are approved for monotherapy use, in case co-medication with MTX cannot be tolerated (31,32).

Current treatment guidelines state that patients newly diagnosed with RA should start treatment with a csDMARD early in the course of disease. The preferred first choice is oral MTX combined with short-term (3-4 months) glucocorticoids until the effect of MTX manifests (33,34). Patients who are contraindicated or cannot tolerate MTX can use other csDMARDs instead, leflunomide and SSZ being regarded as the most effective alternatives (35). Antimalarials may be considered for patients with low disease activity due to better tolerability (34). Up to 30% of patients will respond insufficiently to the initial treatment with MTX (36). If the initial treatment is inefficient, the European Alliance of Associations for Rheumatology (EULAR) guideline recommends trying another csDMARD or combining csDMARDs for patients without poor disease predictors (i.e. high disease activity, early erosions or presence of autoantibodies), and adding a bDMARD or a tsDMARDs for patients with poor disease predictors (33,37). The American College of Rheumatology (ACR) guideline recommends directly adding a bDMARD or a tsDMARDs to the initial csDMARD (34). Based on few randomized head-to-head comparisons (38,39), the latest treatment guidelines express no preference for any of the bDMARDs or tsDMARDs (33,34). Nonetheless, a ranking is apparent in clinical practice, with TNFi favored as a first line, followed by non-TNFi bDMARDs and later on by tsDMARDs. The general recommendation is that bDMARDs should be combined with a csDMARD, since the csDMARD could decrease immunogenicity against the bDMARD (40), but if a csDMARD cannot be used, there is evidence that IL-6 inhibitor monotherapy is more effective than TNFi monotherapy (41,42).

If remission is not achieved with a first bDMARD, another one could be tried. Thus, many patients switch through several therapies, seeking remission. If sustained remission is achieved, patients may start tapering (end even stop) drugs in the following order: first the glucocorticoids, then the b/tsDMARD, and finally the csDMARD (33). Evidence for sustained remission under different treatment strategies involving dose reduction up to discontinuation was provided by several studies, but how remission is maintained over longer time and how to identify and control relapses as early as possible remains to be understood (43–45).

## 1.3 ETIOLOGICAL OBSERVATIONAL STUDIES

In order to develop treatment guidelines, the available evidence is discussed by panels of experts who agree upon recommendations for clinical practice. Randomized controlled trials (RCTs) are the preferred source of evidence, but they are costly and cannot feasibly answer all questions brought up by prescribers and patients. As stated earlier, neither ACR nor EULAR guidelines express a clear preference for the first b/tsDMARD after MTX insufficient response because there are few head-to-head studies comparing all of these alternatives with each other (41,42,46–50). Hence, "the comparative effectiveness/safety between bDMARDs and tsDMARDs" is on the ACR future research agenda (34).

Etiological observational studies, on the other hand, aim to answer causal question about drug treatments, using data collected outside RCTs, which could complement RCT evidence. Data collected outside RCTs is called observational data, but also real-world data, to emphasizes its origin in real clinical practice as opposed to the tightly controlled settings of RCTs (51,52).

Observational studies can employ primary data (collected specifically for the purpose of the study) or secondary data (collected for other purposes) (53–55). Secondary health care data, for example, is routinely collected to document the interaction between patients and various health-care systems for quality assurance, cost reimbursement or resource allocation planning. Such data can be leveraged by observational studies to answer many clinically important questions at lower cost and timelier than possible in RCTs, thus complementing RCT-generated evidence to inform regulatory and clinical decision making (56–58). All studies included in this thesis used secondary data collected in several linked national Swedish registers (as described in Section 4.1) to answer causal questions about the safety and effectiveness of RA treatments.

Regardless of using primary or secondary data, observational studies are affected by several biases which requires a careful consideration of study design and data analysis, as described in detail in the following sections. Picking the "low hanging fruit" represented by the already collected secondary observational data comes at an additional cost. Since secondary data collection is not tailored to specific studies, not all necessary data may be available, which imposes the use of potentially imprecise proxies, leaving residual bias even in thoroughly conducted studies. Furthermore, data collected routinely is usually not collected in a structured manner, at fixed time points, but rather based on the spontaneous interaction between customer and a service provider (e.g. when the patient visits health-care facilities), thus missing data may occur (see Section 4.2.6) (59).

Despite the limitations of observational data, its use has a long tradition in drug safety monitoring. The collection and analysis of spontaneous adverse event reports from patients treated outside RCTs is essential to pharmacovigilance, since real-world drug use may deviate from the strict protocols of trials, impacting drug safety (60,61). Moreover, RCTs are usually not large enough and long enough to provide sufficient data about rare adverse effects which may start being reported once the drug enters clinical use.

On the other hand, comparative efficacy research has been dominated by RCTs, real-world comparative effectiveness evidence being viewed with skepticism (51,62–64). Nevertheless, in recent years methodologic advances in the field of causal inference brought real-world comparative effectiveness research some support (65). The following section presents a framework for designing and analyzing etiological observational studies, based on causal inference principles, meant to improve the validity of observational studies and their ability to communicate results with the purpose of making them more credible.

## 1.4 EMULATING TARGET TRIALS

The *trial emulation framework* is grounded in the potential outcomes causal paradigm (see Section 4.2.1), but it is non-technical and intuitive (66). According to this framework, an etiological observational study should be designed to mimic the process of conducting an RCT (an existing one or just a theoretical one) which answers the same study question. The purpose of this exercise is to avoid certain biases and to transparently and systematically communicate the decisions and definitions involved in structuring and analyzing the observational data, as described below.

To emulate a trial, one needs to first describe how the research question would be answered in the trial, and then how the most important elements of the trial could be "mimicked" using the available observational data (67). If the observational data is not sufficient to emulate the initial target trial, this is amended, re-evaluating if the new version still answers the question of interest.

In a randomized trial, eligible participants are randomly allocated to initiate two (or several) treatment strategies at baseline, and are followed for a specified period of time during which (or at the end of which) the outcome of interest is measured and compared between the treatment groups. The most important elements of a trial emulation are contained in the previous sentence. How each element could be emulated in observational data is discussed below.

### *Treatment strategy*

Commonly, in the first step, all patients receiving the treatments of interest are identified in the available data, along with dates of treatment initiation. For intention-to-treat analyses, where participants are classified according to their baseline treatment and are assumed to continue the baseline treatment throughout follow-up, it is sufficient to know which treatment each patient initiated and when. On the other hand, for per-protocol analyses, where patients are followed only as long as they respect their assigned treatment protocol, one also has to define protocol violations (such as treatment switches) and identify when during follow-up they took place (66–68). For example, in Study III we allowed patients in the infliximab cohort to switch between infliximab containing products (originator/biosimilars) and also to etanercept, but switching to other DMARDs was considered a protocol violation. Thus, it was not enough to identify patients initiating infliximab, but we also had to identify infliximab discontinuation and switches to other DMARDs during follow-up. Tightly controlled trials specify treatment

protocols in great detail (e.g. dose escalation or dose reduction in case of intolerance for each treatment component). However, most observational studies based on secondary data do not have access to such detailed treatment information and will only be able to emulate pragmatic trials in which treatments "naturally" follow the clinical practice (66,69,70). Another treatment strategy that cannot be emulated in observational studies is the placebo. "No use" could be thought of as equivalent to placebo, but they are not identical (71). Depending on the background disease, "no use" may represent a mixture of treatments, other than the studied "active treatment". In diseases where true "no use" is possible, patients not receiving any treatment may be very different in terms of background disease severity or other characteristics (such as, for example, contraindications for the active treatment) from treated patients and comparisons between "active treatment" and "no use" may be problematic. To reduce the risk of residual confounding, some authors proposed to only compare patients treated with different drugs, known as an *active comparator design* (72,73). Active comparators are most effective in reducing bias when there is no clear preference for one treatment over another, as may be the case for non-TNFi bDMARDs and JAKi after TNFi failure in the treatment of RA (74,75). In this situation, treatment assignment is closer to random, making the patients in different treatment groups more similar on average. Another advantage of comparing active treatments is that the start of treatment for each participant can be clearly identified and used as study baseline (i.e. start of follow-up).

### *Eligibility criteria*

Besides initiating the study treatments, the study population will have to fulfill additional eligibility criteria. In most pharmacoepidemiological studies, the minimum eligibility criterium is having a certain background medical condition (e.g. RA in our studies), for which the studied drugs are used as therapy. In addition, in RCTs, eligible individuals should be able to safely receive all the studied treatments (i.e. have no contraindications) and potentially benefit from them, since each participant has a non-zero probability of random allocation to any treatment. The same principle should be respected in observational studies. Patients with contraindications for the studies treatments should be considered for exclusion if those contraindications represent confounders, in order to have a non-zero probability for each studied treatment in each adjustment covariate pattern. Other characteristics highly correlated with exposure could be considered for exclusion as well (see discussion about positivity in Section 4.2.5 and discussion about adjusting for instrumental variables in Section 4.2.4). Other reasons for restricting the study population in RCTs are improved treatment adherence and population homogeneity. Adherence to the assigned treatment strategy is essential for interpreting intention-to-treat estimates (76) and homogeneity improves the balance of baseline characteristics after randomization and the analysis efficacy by decreasing outcome variance (77). Ethical or practical concerns would further narrow down the study populations in RCTs – e.g. more vulnerable individuals, such as pregnant women or the very young and very old, may be excluded, at least from pre-marketing trials, where not enough safety data is yet available about the active treatment; individuals with impaired mental capacity are excluded because they may not be able to consent to participation. Ethical aspects may be of less concern

for the theoretical target trial to be emulated, which could be more inclusive to reflect the population treated in real clinical practice. Importantly, the eligible population must be clearly described based on characteristics measured before baseline. In an RCT, data collection is prospective, thus at baseline, when eligibility is evaluated, no information is available about the follow-up because it has not yet happened. This is not the case in most observational studies using secondary longitudinal data, where at the moment of eligibility assessment the researcher has access to data about the post-baseline future of each patient. Using such future information when selecting patients to be included in the study could introduce selection bias (see Section 4.2.3) (78). Evoking the target trial reminds the researcher that using follow-up data at eligibility assessment would be impossible if the study was conducted prospectively (79). Finally, a sufficient presence in the study data-base before baseline and sufficient time from baseline to the end of available data are common eligibility criteria to ensure that the necessary data is available for each participant (66).

### *Treatment assignment*

In the target trial, participants would be randomly allocated to treatments, such that the group of patients receiving the active treatment would have similar characteristics, on average, to those receiving the reference treatment (see Section 4.2.1). In clinical practice, patients are usually assigned to different treatments according to their characteristics, which imposes the need to adjust for confounding (see Sections 4.2.4 and 4.2.5).

### *Follow-up*

The follow-up is the time period during which the study outcomes are assessed. The premature end of follow-up via censoring and how to correct for the resulting selection bias are discussed in Sections 4.2.3 and 4.2.5. Here I will focus on the start follow-up, which is an essential design decision. Starting follow-up after treatment initiation (i.e. identifying and analyzing ongoing treatments, also called *prevalent treatments*) could lead to selection bias, since treatments started before the start of follow-up have the opportunity to influence the selection of patients into the study. For example, if eligibility is assigned at the start of follow-up, the ongoing treatment can influence which patients become eligible. Furthermore, even is eligibility is evaluated at treatment start, only patients who survive and continue treatment up to the start of follow-up would be included in the respective treatment cohort (79–81). One simple solution to this problem is called the *new-user design* and it implies starting follow-up at treatment initiation, as it is done in RCTs (82). However, studies of prevalent users are not impossible. One could imagine a target trial where prevalent users are randomized to stopping or continuing their treatment (79). Even when eligibility is assessed at treatment start, some may argue that the start of follow-up should be pushed forward, after a lag time, assuming that outcome events happening soon after treatment start may not be causally related to the initiated treatment (but rather to earlier causes) (83). This could be problematic if treatment can influence (possibly via other events than the outcome) the selection of patients at the start of follow-up (79,84). Also, if outcome events taking place between treatment initiation and the start of follow-up are ignored, the analysis may no longer correspond to a survival analysis which counts the first

occurrence of each event. For example, if the incidence of cancer is the outcome event of interest, ignoring the first month after treatment initiation and not excluding patients who developed cancer during this first month will lead to identifying a mixture of incident and prevalent cancers. On the other hand, if none of the compared treatments can have an effect on developing cancer during the first month after initiation, and incident cancers are counted during this period, then the cumulative incidence curves would be identical between treatments during the first months (being driven by background factors which should be balanced if confounding had been properly adjusted for), diverging later during follow-up, when treatments start affecting the outcome. Less common is starting follow-up before treatment initiation, but it may occur if identifying the treatment takes some time (for example collecting a certain number of prescriptions) and this may introduce immortal time bias (79,81). The problem lies in misattribution of events which happened in the first part of follow-up, while treatment is still being defined, to one of the treatment groups or to neither.

### *Outcome*

In RCT protocols, the outcomes of interest are clearly defined in terms of what quantities should be measured, how and when. As with treatment strategies, observational studies, especially those conducted in secondary data, may not be able to adhere to such strict protocols. The outcome data used in our studies is of two main types. The first type refers to events such as medical diagnoses and treatment discontinuations. Recording the event on a certain date implies that it did not take place on other days (the event may still be recurrent). Thus, the first type of outcome data is essentially recorded continuously. The second type refers to variables which need to be measured explicitly at each time point. For example, DAS28 (used as outcome measure in studies II and III), is measured by rheumatologists on certain dates and is unknown for all the other days. Because in real clinical practice DAS28 measurements are not planned with a study in mind, they are done at various times for different patients. To capture outcome information from as many patients as possible, we used wide time-windows around study end-points, but even doing so it was impossible to completely avoid missing outcome data. Another important difference from RCTs is that outcome assessment is not blinded in observational studies. Nonetheless, it may be argued that extracting the exposure and outcome from independent registers, where each register collects data for other purposes than answering research questions, may reduce the risk of outcome misclassification dependent on exposure status.

### *Causal contrast*

Finally, a causal contrast of interest should be specified, and an analysis plan should describe the statistical methods used to estimate this causal contrast. To specify a causal contrast, one should describe which study participants included at baseline should contribute to the outcome assessment (e.g. everyone who entered the study or just participants who continued baseline treatment up to the end-point), how they should be classified according to exposure status (e.g. would they be classified according to their baseline initiated treatment or according to the treatment status at the end-point), how the outcome would be summarized in each treatment

group (e.g. proportion of participants with outcome events out of all participants in the exposure group at baseline) and how would the summaries be compared (e.g. a ratio or a difference between proportions in different treatment groups). For example, in study III**,** each participant eligible to enter the study at baseline contributed to the analysis with one observation which was classified according to the treatment received at baseline (excluding one patient who died during follow-up). For each observation, we measured the outcome as a binary treatment response variable with values calculated as a function of baseline and end-point DAS28. If a participant had discontinued the protocol treatment before the nine-month end-point, the outcome was imputed to zero (i.e. non-responder). We summarized the outcomes as treatment response proportions within each treatment cohort and contrasted the proportions by calculating their ratio. Because all individuals included at baseline were kept in the analysis and classified according to their baseline treatment status, we described the analysis as intention-to-treat with the modification of classifying participants who interrupted treatment as non-responders. Traditional intention-to-treat analyses may not be informative outside tightly controlled, short-term RCTs where most patients adhere to the baseline treatment until the end-point (68,76). An alternative is the per-protocol analysis which estimates causal contrasts as if all participants followed the protocol treatment strategy. Even in RCTs, where baseline randomization is available, time-varying patients characteristics have to be measured during follow-up to correct for selection bias if patients who violate the protocol during follow-up are censored (66,68).

# 2 BACKGROUND FOR THE INCLUDED STUDIES

## 2.1 STUDY I – BIOLOGICAL DMARDS AND THE RISK OF GASTRO-INTESTINAL PERFORATIONS IN RA

RA patients face a higher burden of gastro-intestinal (GI) complications compared to age and sex matched individuals without RA, from the same population and calendar period (85,86). The proportion of upper and lower GI complications has changed over calendar time, the incidence of the historically dominant upper GI perforations showing a decline (85) while the incidence of lower GI complications is slowly rising (87,88). Ample evidence attributed upper GI toxicity to NSAIDs, which used to be prescribed in high dose and for long periods to contain the symptoms of RA (89–92). The development of NSAIDs with decreased GI toxicity (selective cyclooxygenase-2 inhibitors), the co-administration of proton pump inhibitors (PPI) and the increased early use of DMARDs, which decreases the need for NSAIDs, are all possible explanations for the observed decline in upper GI adverse events (93). While this is reassuring, an increased incidence of lower GI complications is concerning because they require more health-care resources and have higher mortality compared to upper GI complications (94,95). For example, mortality rates as high as 40% have been reported for perforated diverticulitis (96,97).

An increased risk of lower GI injury has been associated with the use of glucocorticoids (98,99), but NSAIDs, including selective cyclooxygenase-2 inhibitors, have also been implicated (100,101). On the other hand, limited available evidence suggests that common csDMARDs may not increase the risk of GI perforations (102–104). Besides, the use of 5-aminosalicylic acid derivatives in the treatment of diverticular disease may even suggest a possible protective role for SSZ (105). While sporadic cases of GI injury in patients treated with TNFi having been reported (106,107), a study of the British Society for Rheumatology Biologics Register Rheumatoid Arthritis (BSRBR-RA) found no difference in the incidence of either upper or lower GI perforations between patients receiving TNFi and those not receiving any bDMARDs (108).

A safety signal for lower GI perforations associated with tocilizumab emerged from RCTs (109,110) and was later confirmed by a few observational studies. A study conducted in the German biologics register (RABBIT) showed an 4.5 times increased lower GI perforations incidence rate among patients treated with tocilizumab compared to those treated with csDMARDs, but no risk increase for abatacept, rituximab or TNFi (111). Two studies compared incidences of GI perforations between TNFi and non-TNFi bDMARDs using US insurance claims data (Medicare, MarketScan) (112,113) and two previously validated International Classification of Disease (ICD) definitions, one with higher specificity and the other with higher sensitivity (114). As expected, the incidence rates of GI perforations were higher when a higher sensitivity definition was used. In the first study, the incidence rate of lower GI perforations for tocilizumab versus TNFi was 4 times higher when using the specific definition and 3 time higher when using the sensitive definition (112). Using the specific definition, the second study reported a 2.5 higher incidence rate of lower GI perforations in the

tocilizumab cohort compared to the TNFi cohort (113). The study did not report increased lower GI perforation incidence rates for rituximab or abatacept versus TNFi.

## 2.2 STUDY II – COMPARATIVE EFFECTIVENESS OF BARICITINIB, TOFACITINIB AND BIOLOGICAL DMARDS IN RA

JAKis are a new addition to the arsenal of targeted disease modifying anti-rheumatic drugs used in RA. As their name suggests, JAKis inhibit the action of Janus Kinases, enzymes (tyrosine kinases) involved in transducing the signal of several cytokines from the membrane receptor to the cell nucleus (115). Each JAKi has selective affinity for certain JAK isoforms, thus inhibiting the action of certain cytokines more than that of others (116). Tofacitinib has higher affinity for isoforms JAK1, JAK2 and JAK3, while baricitinib has higher affinity for JAK1 and JAK2 (117). This pharmacodynamic specificity diminishes at high drug concentrations.

In 2012, the FDA approved the first JAKi, tofacitinib, for the treatment of RA (118), followed by the Swiss medicines authority in 2013 (119) and the Australian medicines authority in 2015 (120). In the European Union (EU), tofacitinib was approved only in 2017 (121), after an initial refusal due to safety concerns in 2013 (122). Tofacitinib was authorized based on a battery of eight phase III RCTs (the ORAL trials). In these studies tofacitinib was proven to be: i) clinically superior to MTX among patients with active RA not previously treated with MTX (but some of whom received other csDMARDs) (123); ii) clinically superior to placebo, when used as monotherapy after (cs/b)DMARD discontinuation (124) and when added over a csDMARD background (mainly MTX) in RA patients who did not respond to previous csDMARD or bDMARD treatments (125,126); iii) non-inferior to adalimumab, when added over an MTX background, with tofacitinib monotherapy yielding a lower response rate compared to both combination treatments (49).

Several observational studies evaluated the effectiveness of tofacitinib, in countries where it had been approved early (119,120,127–129). These studies also provided the first head-to-head comparisons between tofacitinib and other bDMARDs than adalimumab. Their main findings are summarized in Table 2.2.1. Studies that compared CDAI remission and low disease activity found no significant difference between tofacitinib and bDMARDs (119,120,128). Treatment discontinuation due to lack of effect was marginally higher among bDMARDs compared to tofacitinib (119,129), even though the reverse might be true at first line of therapy (129). Moreover, one study showed that tofacitinib may be more frequently discontinued due to adverse events (119). Main limitations were low statistical power (127,128) and lack of clinical data in some studies (127,129). The lack of clinical data precluded direct adjustment for important confounders such as disease activity, and limited the studies outcomes to drug discontinuation or claims-based effectiveness criteria, which cannot be easily interpreted clinically or compared to measurements used in RCTs.

**Table 2.2.1** – Summary of observational studies comparing tofacitinib to other DMARDs

| Author, Year | Country, Data Sources, Period | Treatment Strategies, Number of Participants | Main Findings |
|---|---|---|---|
| Machado 2018 (127) | US MarketScan health insurance claims data, 2011 to 2014 | csDMARDs (n=5399)<br>TNFi (n=13367)<br>Non-TNFi (n=2902)<br>Tofacitinib (n=164) | Similar claims-based effectiveness achieved within 1 year: TNFi vs non-TNFi (RR 0.9 (95% CI: 0.9 to 1.0)) and on tofacitinib vs non-TNFi (RR 0.8 (95% CI: 0.4 to 1.3)). |
| Reed 2019 (128) | US Corrona rheumatology clinical registry, 2012 to 2016 | TNFi mono (n=1889)<br>TNFi + MTX (n=4352)<br>Tofacitinib mono (n=238)<br>Tofacitinib + MTX (n=164) | No differences in CDAI low disease activity or remission at 6 months between: tofacitinib + MTX vs tofacitinib mono (OR 1.1 (95% CI:0.6 to 2.0)); TNFi + MTX vs tofacitinib mono (OR 1.2 (95% CI:0.7 to 2.0)); TNFi + MTX vs tofacitinib + MTX (OR 1.1 (95% CI: 0.7 to 1.9)). |
| Finckh 2020 (119) | SCQM-RA, the Swiss rheumatology clinical registry, 2013 to 2019 | TNFi (n=1847)<br>Non-TNFi (n=1338)<br>Tofacitinib (n=793)<br><br>*non-TNFi bDMARDs: abatacept, tocilizumab, sarilumab | Higher probability of treatment discontinuation: on TNFi vs tofacitinib (HR 1.3 (95% CI: 1.1 to 1.5)) or on Non-TNFi vs tofacitinib (HR 1.1 (95% CI: 1.0 to 1.2)).<br>*Discontinuation due to ineffectiveness was more likely for bDMARDs, while discontinuation due to adverse events was more likely for Tofacitinib.<br><br>No significant differences in CDAI low disease activity proportions at 1 year. |
| Fisher 2020 (129) | Canadian MarketScan health insurance claims data, 2012 to 2015 | Tofacitinib first line (n=1031)<br>Tofacitinib later line (n=1535)<br>bDMARD first line (n=17803)<br>bDMARD later line (n=9849)<br><br>*bDMARDs: TNFi, abatacept, tocilizumab | Higher probability of discontinuation, tofacitinib vs bDMARDs, at first line (HR 1.1 (95% CI: 1.1 to 1.3)).<br>*Discontinuation rates were the lowest on TNFi, while rates on abatacept and tocilizumab were similar to tofacitinib.<br><br>Lower probability of discontinuation on tofacitinib vs bDMARDs, at later line (HR 0.9 (95% CI: 0.8 to 1.0)). |
| Bird 2020 (120) | OPAL data-base (Australian rheumatology clinical register), 2015 to 2018 | TOFA (n=652)<br>bDMARDs (n=2158)<br><br>*bDMARDs: TNFi, abatacept, rituximab, tocilizumab, anakinra | No significant differences, tofacitinib vs bDMARDs, at 18 months: on DAS28 remission (57.8% vs 52.4%), on CDAI remission (30.5% vs 29.0%), on SDAI remission (30.9% vs 29.2%).<br><br>No difference in drug discontinuation. |

CI = confidence interval; HR = hazard ratio; OR = odds ration; RR = risk (proportion) ratio

Baricitinib was the second JAKi to be approved for RA, in 2017 in the EU and in 2018 in the US (118). Baricitinib's approval was based on four phase III RCTs. In these trials, baricitinib was proven: i) clinically superior to MTX in RA patients who have not been treated with DMARDs, with similar efficacy between baricitinib monotherapy and baricitinib in combination with MTX, but improved radiographic progression on baricitinib in combination with MTX (130); ii) clinically superior to placebo when added to a background of csDMARDs in patients who have not responded to previous treatment with csDMARDs or with bDMARDs (131,132); iii) clinically superior to adalimumab when added over a background of MTX in patients with insufficient response to MTX alone (50).

There is little evidence about the effectiveness of baricitinib compared to tofacitinib or bDMARDs in the real clinical practice. Three small studies conducted in Japan compared baricitinib to tofacitinib (133–135). They found that baricitinib was more frequently used as monotherapy compared to tofacitinib (133,134). The proportions of CDAI six-months remissions was higher on baricitinib compared to tofacitinib, but no differences were observed in HAQ-DI (133,134). Drug retention was also higher on baricitinib, with fewer patients discontinuing due to adverse events, compared to tofacitinib (133,134).

To summarize, real-world evidence comparing baricitinib to tofacitinib indicates a potentially superior effectiveness of baricitinib, reflected in higher drug retention and potential clinical superiority. However, if any difference exists between the two JAKis, it is likely minor and the small studies available were not able to provide convincing evidence for it.

## 2.3 STUDY III – EMULATION OF THE SWEFOT TRIAL IN OBSERVATIONAL DATA

RCTs are the gold standard for comparing the efficacy and safety of treatments, and they are the mandatory source of evidence on which the approval of new medical treatments is based (136). However, large RCTs require extensive resources and take a long time to complete, hence authorities do not usually require more than two positive RCTs designated as the basis ("pivotal") for granting marketing authorization approval (137). Consequently, at marketing approval, a new treatment had usually been compared to placebo and to only a few existing alternatives (138,139). Furthermore, the populations included in clinical trials are selected and may not represent all patients receiving the studied treatments in clinical practice (140,141). Prescribers and patients need to balance the relative benefits and risks of all available treatment options in order to make optimal treatment decisions. To inform their decisions, comparative effectiveness evidence needs to continue being generated after the entry of drugs in clinical practice.

Observational studies could complement phase IV RCTs by providing timely post-marketing comparative safety and effectiveness evidence, but they are viewed with skepticism due to their susceptibility to bias. The trial emulation framework, described in Section 1.4, states that conducting etiological observational studies to emulate existing or theoretical target RCTs could help with avoiding some biases. Complex designs and analysis methods are sometimes

necessary for bias correction (see for example Section 4.2.5), but they are not always clearly reported in pre-agreed protocols (142). Hence, a structured and transparent emulation procedure has been previously proposed (143), which entails choosing the target trial, assessing the emulation feasibility, drafting a study protocol and registering the protocol before analyzing any outcome data. The intention is to avoid selecting which analyses to perform and which results to present based on observed exposure-outcome associations (i.e. result manipulation and publication bias) (61,144).

A sensible approach to testing the trial emulation framework is to emulate existing RCTs and to compare results between the target RCT and its emulation (143). This approach has been adopted in some studies, with mixed results (145–149). In Study III we add to this literature by emulating an RCT from the field of rheumatology, using observational data from the Swedish Rheumatology Quality register (SRQ) linked to other national Swedish registers, and comparing the emulation results to those of the trial. Our target trial was Swedish Farmacotherapy (SWEFOT), an open-label RTC nested in SRQ which compared the addition of infliximab over MTX with addition of SSZ and HCQ over MTX, among early RA patients unsuccessfully treated with MTX monotherapy for 3 months. In the primary analysis, patients were evaluated at 12 months after inclusion (i.e. 9 months after randomization), at which point the EULAR good response proportion ratio, comparing infliximab (+ MTX) to SSZ + HCQ (+MTX), was 1.6 (95% confidence interval: 1.1 to 2.3) (150).

Several observational studies compared treatment effectiveness of adding a TNFi (or another bDMARDs) versus adding SSZ + HCQ to the MTX background (151–154). The results are summarized in Table 2.3.1. Briefly, all observational studies indicated higher effectiveness of TNFi + MTX compared to SSZ + HCQ + MTX, despite various outcome definitions, including clinical measurements as well as prescription-based algorithms for treatment persistence.

Two RCTs conducted after SWEFOT did not find SSZ + HCQ added to MTX inferior to adding etanercept to MTX at the one-year end-point (37,155). Several differences from SWEFOT could be pointed out. First of all, in both studies physicians and patients were blinded to the allocated treatment. Second, the study designs and the estimated causal contrasts were different. The main analysis in the first trial was an intention-to-treat analysis, classifying patients according to their baseline allocation regardless of compliance with that treatment (155). In the second trial, patients with insufficient response to the baseline treatment by six months were switched to the opposite arm, but patients were analyzed according to the baseline treatment (37). On the other hand, secondary analyses showed similar results with non-responder imputation (155) and among patients who did not switch (37). Third, there were differences in the included populations. The first trial included a large proportion of patients positive for rheumatoid factor (> 90%) (155). The second trial included patients with a long disease activity (mean of ~5year), and a larger proportion were male (37).

**Table 2.3.1** – Summary of observational studies comparing TNFi + MTX with SSZ + HCQ + MTX

| Author, Year | Country, Data Sources, Period, Population | Treatment Strategies, Number of Participants | Main Findings |
|---|---|---|---|
| Lie 2011 (151) | Norway, NOR-DMARD register, before 2010, MTX insufficient responders with RA duration < 5 years | TNFi + MTX (n=98)<br>csDMARDs + MTX (n=129)<br>* csDMARDs + MTX included 44 patients on SSZ + HCQ + MTX | Lower treatment discontinuation over 2 years for TNFi + MTX compared to several csDMARD + MTX (including for SSZ + HCQ + MTX) (HR 0.4 (95% CI 0.3 to 0.7)).<br>Overall superior clinical responses at 3 and 6 months for TNFi + MTX compared to csDMARD + MTX (e.g. EULAR good response: 36% vs 16% and 40% vs 21% respectively). |
| Sauer 2017 (154) | US, Data about veterans from several linked data-bases, 2006 to 2012, Veterans with diagnosed RA (majority male) | TNFi + MTX (n=3204)<br>SSZ + HCQ + MTX (n=1160) | Higher treatment persistence (on 3 prescription-based definitions) and adherence on TNFi + MTX compared to SSZ + HCQ + MTX. |
| Källmark 2021 (152) | Sweden, SRQ, 2000 to 2012, MTX insufficient responders with DAS28>2.6 at study treatment start | TNFi + MTX (n=1155)<br>SSZ + HCQ + MTX (n=347) | Short- and long-term DAS28 remission more likely achieved under TNFi + MTX compared to SSZ + HCQ + MTX at 1-year (OR 1.8 (95% CI 1.2 to 2.7) and 1.9 (95% CI 1.0 to 3.5)), and at 2-years (OR 1.9 (95% CI 1.2 to 3.1) and 1.6 (95% CI 0.9 to 2.8)). |
| Curtis 2021 (153) | US, Corrona register, 2001 to 2009, bionaïve and bDMARD exposed RA patients who initiated the study treatments | TNFi + MTX (n=3926)<br>SSZ + HCQ + MTX (n=262) | Higher treatment discontinuation in the SSZ + HCQ + MTX cohort compared to the TNFi + MTX cohort (HR 2.2 (95% CI 1.6 to 2.9).<br>Higher CDAI low disease activity proportion in the TNFi + MTX cohort compared to the SSZ + HCQ + MTX cohort in both bionaïve patients (49% vs 33%) and among bDMARD exposed patients (32% vs 27%). |

CI = confidence interval; HR = hazard ratio; OR = odds ration; RR = risk (proportion) ratio

## 2.4 STUDY IV – GLUCOCORTICOIDS AND THE RISK OF SERIOUS INFECTIONS IN RA

Glucocorticoids are a class of therapeutic agents with anti-inflammatory action derived from the endogenous steroid hormone cortisol. In the treatment of RA, most frequently used glucocorticoids are oral prednisone and prednisolone, while methylprednisolone is common for parenteral administration (156). Glucocorticoids were first administered to RA patients by Hench, Kendall and colleagues in the 1950s, for which they were awarded the Nobel prize for medicine (157,158).

Glucocorticoids produce various effects via genomic and non-genomic mechanisms. Their small and lipophilic molecules can cross cell membranes, bind to intra-cellular (cytosolic) receptors and regulate the transcription of a large proportion of the human genome (159). Such genomic effects can occur even at very low doses, have longer latency and are responsible for both anti-inflammatory activity (by inhibiting the expression of pro-inflammatory factors or by activating the transcription of genes expressing anti-inflammatory factors) and for adverse outcomes (mainly by transcription activation) (156,160). Components of the glucocorticoid cytosolic receptor complex can dissociate and produce non-genomic effects as well (e.g. inhibit the release of the proinflammatory arachidonic acid) (161). Glucocorticoids can also inhibit cellular immunity by binding to membrane receptors (related to the cytosolic receptors) on immune cells such as monocytes and T-cells. Finally, at high concentrations, when glucocorticoid receptors are saturated, glucocorticoids can incorporate directly into cellular and mitochondrial membranes, changing their physical properties and modifying cell function (159,161).

The extensive effects of glucocorticoids on immune cells present as: i) decreased numbers of circulating leukocytes; ii) inhibited synthesis and release of pro-inflammatory cytokines; iii) inhibited release of lysosomal enzymes and reactive oxygen species; iv) hindered access of leukocytes to inflammation sites through reduced adhesiveness and permeability of the endothelium (162,163). This reduced ability of the immune system to fight pathogens increases the risk of infection. While the increased risk of infection is generally ascribed to high glucocorticoid doses (164), the effect of low doses, commonly used in RA, and the relation between low-dose treatment duration with and risk of infection is less clear (165).

As stated previously, the preferred source of evidence for causal effect of glucocorticoids on the risk of infection would be RCTs. However, most RCTs are not designed to study safety outcomes. Two systematic reviews of RCTs assessing the effect of glucocorticoids on the risk of infections noted potential selective reporting (only serious infections, only events for which causality could be ascertained or only pre-specified infections) (164,166). This likely led to underestimation of absolute risks, but relative risk estimates should have remained unbiased since underreporting should be independent of treatment, at least in blinded trials. While selective reporting might not introduce bias, it may decrease precision, especially when dealing with rare outcomes. In a meta-analysis of 21 RCTs evaluating the risk of infection associated

with the use of glucocorticoids in RA, an inconclusive pooled risk ratio of 1.0 (95% confidence interval 0.7 to 1.4), covering both harmful and protective effects, was reported (166).

In the same review, a meta-analysis of 42 observational studies yielded a pooled relative risk of 1.7 (95% confidence interval 1.5 to 1.9), indicating an increased risk of infection with glucocorticoids use (166). Many possible reasons for conflicting results between RCTs and observational studies could be suspected. For example, not all included observational studies were primarily designed to study the effect of glucocorticoids on the risk of infection. Instead, glucocorticoid use was one of many covariates included in regression models (possibly a confounder for the main exposure) (167–169). Associations between glucocorticoids use and the risk of infections from such analyses don't necessarily have a causal interpretation since confounding adjustment for the main exposure might not be suitable for glucocorticoids exposure. Another important issue noted in the review concerning observational studies, was the substantial heterogeneity in exposure definitions. This was prompted by the complex pattern of glucocorticoids exposure in clinical practice and by various ways in which such exposure is captured in clinical databases, electronic health records or prescriptions registers. In rheumatology, glucocorticoids are commonly initiated at first diagnosis, are subsequently tapered to discontinuation as slower acting csDMARDs take effect, and are often restarted in response to recurrent or under-controlled inflammation or when switching DMARDs (for this reason they are known as "bridging" therapy) (33).

More recent observational studies continued to define exposure to glucocorticoids in a variety of ways. Exposure and confounding definitions as well as results from recent cohort studies are summarized in Table 2.4.1, and those from case-control studies in Table 2.4.2. In some cohort studies, exposure was considered fixed (decided) at baseline and kept constant during follow-up, being measured using information before baseline (170,171), but also after baseline (172,173). Fixing exposure at baseline has the advantage of allowing the measurement of confounders at one point only (before baseline). However, if exposure changes over the course of follow-up, the exposure value measured around baseline may not accurately represent the true exposure pattern that produced the outcome. If on the other hand, the baseline exposure value is estimated by averaging or cumulating exposure over follow-up, then the baseline covariate measurements may not be sufficient to adjust for confounding. The "per-protocol" strategy employed by George et al. (170), where participants were followed as long as they adhered to the baseline dose, is an appropriate design for studying a baseline-fixed exposure, but in this case time-varying covariates need to be measured to account for selection bias via censoring at dose changes. In other cohort studies, as in most case-control studies, exposure was allowed to change over the course of follow-up. Adjusting for baseline confounders when exposure is updated over the course of follow-up may leave residual unadjusted confounding if patients characteristics, which changed during follow-up, influenced exposure changes after baseline (174). Adjusting for time-updated confounding was common in case-control studies, where exposure and confounding were measured within certain time-windows before each index-date (i.e. outcome event date) (175–177). However, if the exposure window is long enough (e.g. from the start of follow-up to index date) and confounders are measured in

overlapping windows, then exposure could affect the confounders, potentially introducing bias (see Section 4.2.5) (86,177). Regardless of how exposure and confounders were measured, all results pointed to an increasing risk of infection associated with the dose of glucocorticoids used.

**Table 2.4.1** – Summary of recent cohort studies addressing the risk of infection associated with glucocorticoids exposure

| **Author, Year** | **Country, Data Sources, Period, Population** | **GC Exposure Definition, Confounding Adjustment** | **Main Findings** |
|---|---|---|---|
| Fardet 2016 (172) | UK, THIN general practice data-base, 2000 to 2012, patients with various diseases treated with glucocorticoids (including RA) | Exposure defined using prescription data.<br>Exposed – collected GC prescriptions, followed during their first treatment period<br>Not Exposed – did not collect GC prescriptions<br>Confounding covariates measured at baseline, but also during follow-up. | Increased risk of various infections among patients exposed to GC versus not exposed, with HR varying between 2.0 and 5.8. Dose-dependence observed.<br>Other potential risk factors identified: age, diabetes. |
| Roubille 2017 (173) | France, ESPOIR cohort, 2002 and 2005, patients with early RA (< 6 months), naïve to DMARDs and GC | Exposed – any GC prescription during 7 years of follow-up<br>Not Exposed to GC<br>Confounding covariates measured at baseline. | Most GC use occurred during the first 6 months of follow-up at low dose (<5mg/day).<br>19 Severe infections occurred during follow-up: 16 (4%) in the GC exposed cohort and 3 (1%) in the unexposed cohort (no adjusted contrast was presented) |
| Best 2018 (171) | US, MarketScan and Medicare, 2012 to 2013, RA patients | Oral GC exposure accumulated during 2012, categorized according to quartiles. Contrast between each cumulative dose quartile and no GC use estimated.<br>Confounders measured during the baseline year (2012). | Dose response relationship between the risk of hospitalization for opportunistic infections during 2013 and the use of oral GC during 2012, with adjusted ORs between 1.0 and 1.9. |
| George 2020 (170) | US, Medicare and Optum, 2007 to 2015, RA on stable DMARD treatment, started 6 months before baseline | Average daily dose of GC estimated from prescriptions within 90 days before baseline and categorized into: (≤ 5 mg/day); (> 5 mg/day ≤ 10 mg/day); (> 10 mg/day).<br>Follow-up censored at the end of the index DMARD or at a change in GC dose.<br>Confounding covariates measured during the baseline period. | Dose response relationship between the incidence of hospitalization for infection and the dose of GC.<br>The cumulative incidences at 1-year were higher in Medicare (ranged from 8.6% on no GC to 17.7% on > 10 mg/day) than in Optum (ranged from 4.0% to 10.6%). |

**Table 2.4.1** – Summary of recent cohort studies addressing the risk of infection associated with glucocorticoids exposure

| Author, Year | Country, Data Sources, Period, Population | GC Exposure Definition, Confounding Adjustment | Main Findings |
|---|---|---|---|
| Haroui 2015 (178) | Canada, BioTRAC prospective cohort, up to 2011, bionaïve RA patients who initiated infliximab | The dose of GC measured at follow-up visits, once every 6 months, and categorized into: no GC, (≤ 5 mg/day), (> 5 mg/day).<br>Adjustment for time-varying confounding was done in Cox regression models including the time-varying GC dose. | Dose response relationship between the risk of patient/physician reported infections and the dose of GC observed: HR 2.1 for low dose; HR 2.5 for high dose (vs no GC). |
| Schenfeld 2017 (174) | US, MarketScan and Medicare, 2005 to 2014, bionaïve RA patients who initiated TNFi | Time-varying GC daily-dose estimated from prescription data within follow-up episodes. Categorized into: no GC; very low dose (≤ 5 mg/day); low dose (≤ 7.5 mg/day); high dose (> 7.5 mg/day); v) very high dose (> 20mg).<br>Confounding covariates measured at baseline. | Dose response relationship between the incidence of hospitalization for infection and the dose of GC: IRR 1.4 for low-dose GC; IRR 2.8 for high-dose GC (vs no GC). |
| Wu 2019 (175) | UK, linkage between primary health-care data-base CPRD, Hospital Episode Statistics and Mortality register, 1997 to 2017, patients with polymyalgia rheumatica and giant cell arteritis | Daily GC dose derived from prescription data.<br>Several time-varying GC exposure variables defined:<br>Current use (binary); current daily dose, categorized into: (> 0.0–4.9 mg), (5.0–14.9 mg), (15.0–24.9 mg), (≥ 25.0 mg);<br>Cumulated dose since one year before the start of follow-up;<br>Cumulated dose within the last year<br>Time-varying confounding covariates measured at baseline and during follow-up | Current GC use associated with an increased risk of infections (all causes): HR 1.5.<br>Dose response relation-ship with HR increasing from 1.4 to 2.3 over the four current daily dose categories (vs no GC).<br>Cumulated GC dose within the last year associated with an increased risk of infection: HR 1.5 for 1g increase in cumulated dose.<br>No association for the cumulated dose since one year before the start of follow-up. |

GC = glucocorticoids; HR = hazard ratio; IRR=incidence rate ratio; OR = odds ratio

**Table 2.4.2** – Summary of recent case-control studies addressing the risk of infection associated with glucocorticoids exposure

| Author, Year | Country, Data Sources, Period, Population | GC Exposure Definition, Confounding Adjustment | Main Findings |
|---|---|---|---|
| Dixon 2012 (177) | Canada, Québec region administrative data, 1985 to 2003, RA patients > 65 years | The time-updated cumulative dose of oral GC, calculated within a 3-year window before each index date, using different (estimated) weights for exposure in different periods to reflect a differential impact on current risk at index date.<br>Confounding covariates measured prior to index date (i.e. time-varying). | Higher weights estimated for recent GC doses compared to past doses, indicating that recent GC use had a stronger impact on current infection risk.<br>Increased risk of infections with increasing oral GC doses and duration of use, with a small but statistically significant risk increase even at 5 mg /day for one week relative to no use. |
| Widdifield 2013 (176) | Canada, Ontario health administrative data, 1992 to 2010, RA patients older than 65 years | Current GC exposure defined as prescriptions which included the index date. Current GC daily dose categorized into: low (≤ 5 mg), medium (6–9 mg), high (10–19 mg), and very high (≥ 20 mg).<br>Past exposure defined as non-current exposure within one year before index-date (binary).<br>Confounding covariates measured prior to index date (i.e. time-varying). | Current GC exposure associated with an increased risk of hospitalization for infections in a dose-dependent manner, with ORs ranging from 4 for low doses to 7.6 for very high doses (vs no GC).<br>Past GC use independently associated with the current risk of hospitalization for infection with an OR of 2.3. |
| Wilson 2019 (86) | UK, linkage between primary health-care data-base CPRD and Hospital Episode Statistics, 1997 to 2012, RA patients | GC exposure measured between the start of follow-up (RA diagnosis) and index date.<br>Exposure within 180 days before index date considered recent.<br>Confounding covariates measured prior to index date (i.e. time-varying) | An increased risk of hospitalization for infections associated with any GC use before index-date (OR 1.3) and with current GC use before index date (OR 1.5), but not with past GC use (vs no GC).<br>Dose response relationship observed. |

GC = glucocorticoids; OR = odds ratio

# 3 RESEARCH AIMS

The overall aim of this PhD project was to apply principles and methods of causal inference in comparing the safety and effectiveness of various RA therapies.

## 3.1 STUDY I – BIOLOGICAL DMARDS AND THE RISK OF GASTRO-INTESTINAL PERFORATIONS IN RA

In our first study, we aimed to use data from the Swedish RA population to test the signal of an increased risk of GI perforations associated with tocilizumab.

Even though a few observational studies had already explored this association in other populations (German and American), our study was motivated by the possibility to adjust for confounding using a more comprehensive set of variables, including RA severity and comorbid conditions. Moreover, we estimated the incidence of GI perforations in a general population comparator cohort to offer a broader context around our results.

## 3.2 STUDY II – COMPARATIVE EFFECTIVENESS OF BARICITINIB, TOFACITINIB AND BIOLOGICAL DMARDS IN RA

In the second study, we aimed compare the effectiveness of baricitinib to that of tofacitinib, abatacept, rituximab, IL-6 inhibitors and TNFi.

The second study was motivated by the scarcity of evidence about the effectiveness of baricitinib compared to bDMARDs, and the accumulation of a considerable amount of data on the use of baricitinib by Swedish rheumatologists, after its EU approval in 2017, which was available for analysis.

## 3.3 STUDY III – EMULATION OF THE SWEFOT TRIAL IN OBSERVATIONAL DATA

In the third study, we aimed to examine how closely we could emulate the design of the SWEFOT open-label trial in observational data, and then to benchmark the results of the emulation against those of the trial.

Additionally, we aimed to explore how results change when relaxing the trial eligibility criteria, thus including patients who would not have been eligible for the target trial, but who received the study treatment in practice.

The third study was motivated by the mistrust in comparative effectiveness observational studies, and by the possibility to improve the internal validity of observational studies by emulating target trials.

## 3.4 STUDY IV – GLUCOCORTICOIDS AND THE RISK OF SERIOUS INFECTIONS IN RA

In the fourth study, we aimed to compare the risk of serious infections (i.e. hospitalization for infection) between different patterns of time-varying oral glucocorticoids exposure, among early RA patients.

The fourth study was motivated by the discrepancy between RCT and observational results, previously reported in a meta-analysis (166), and by limitations of previous observational studies in modelling the time-varying exposure to glucocorticoids in RA, and properly adjusting for time-varying confounding.

# 4 MATERIALS AND METHODS

## 4.1 DATA SOURCES

All studies included in this thesis relied on individual data from several population-based national Swedish registers (58) linked via the personal identification number allocated to each Swedish resident at birth or immigration (179,180).

The Swedish Rheumatology Quality register (SRQ) was initiated in 1995 to prospectively follow RA patients and it contains the Swedish Biologics Register (ARTIS), which covers more than 95% of all bDMARD utilization in Sweden (181). Hence, SRQ was the basis for identifying RA patients, especially those treated with bDMARDs. This is important since some bDMARD administration (e.g. intravenous) takes place in the hospital, thus it is not covered by the prescribed drugs register (PDR). The SRQ contains information on the date of treatment initiation (more precisely the date of a decision to initiate treatment), and the date and reason for treatment discontinuation, as well as data on disease activity (such as DAS28 and its components, and CDAI), and functional ability (HAQ-DI), recorded at each rheumatology visit (182).

The Swedish PDR started in July 2005 and collects data about dispensed prescriptions from community pharmacies in Sweden (183,184). Each record refers to a dispensed pharmaceutical product and contains information about the item dispensed (active substance, brand name, pharmaceutical formulation, dosage per unit and number of units in the package), the amount dispensed, the date of prescription and the date of dispensation, information about expenditure and reimbursement, a unique patient identifier and prescriber information. Active substances are classified according to the Anatomical Therapeutic Chemical (ATC) classification system (185). The PDR does not include data on over-the-counter dispensations and on in-hospital drug use and contains incomplete data on vaccinations and drugs used in nursing homes. Also, importantly, information about the indication for treatment and dosing schedule is optionally recorded in a free text field which is difficult to use in research.

The Swedish National Patient Register (NPR) was launched in 1964 but covered only inpatient care until 2001 when specialty outpatient hospital care started being reported (186). The most important data provided by the NPR are the visit date (as well as discharge date for hospital stays), and a list of diagnoses and procedures applied to the patient during the visit. The unique patient identifier is also available to link patient data to other sources. Diagnoses are coded using the Swedish versions of the WHO ICD system (187). Procedures are coded using a Swedish versions of the Nordic Medico-Statistical Committee (NOMESCO) Classification of Surgical Procedures. A limitation of the NPR is that it does not have good coverage of primary care visits since it is not mandatory to report them. Also, no data on disease severity is available, thus we relied on the SRQ to obtain this information for RA patients.

The Swedish Cause of Death Register is available in its electronic format since 1952 and provides information on the date and causes of death for each individual (188). Contrary to the

NPR, the Cause of Death Register encodes the causes of death using the international WHO ICD system. The information about the causes of death comes from the medical death certificate compiled by the physician who last saw the dead individual. The causes of death are separated into the underlying cause of death (the disease or injury which initiated the causal chain leading to death) and contributing causes of death.

The Swedish Cancer Register collects data on incident malignancies (and some benign tumors) since 1958. The register receives notifications about newly identified cancers from several sources (e.g. hospitals, outpatient clinics, primary care physicians, pathology and cytology laboratories) but not from the cause of death register, thus missing cancers that were not treated (e.g. in elderly or those with difficult early diagnosis such an pancreatic cancers) (189). Nonetheless, national coverage is high for most malignancies.

Demographic data have been collected from the Total Population Register maintained by Statistics Sweden (190). This register started in 1968 and contains data about: sex, date and country of birth, migration dates, death date and others for all Swedish residents. Data about the patient's education level were obtained from Longitudinal Integrated Database for Health Insurance and Labor Market Studies (LISA) (191).

## 4.2 GENERAL INTRODUCTION TO METODOLOGY

The following sections will introduce some general concepts about the methods used to design and analyze our four studies. It includes general introductions to the potential outcomes framework, the paradigm under which the causal inference methods used in this thesis have been developed, to directed acyclic graphs, which can be used to intuitively reason around bias sources and adjustment methods, and finally to inverse probability weighting, handling of missing data and survival analysis.

### 4.2.1 The potential outcome framework of causal inference

The potential outcomes framework provides a way of defining *causality*, at the individual and at the population level, as a contrast between *potential outcomes*. It is used to describe the conditions under which causal relationships can be identified in scientific studies (192). It is by no means the only framework for causal inference (193), but it is the framework under which most methods used in this thesis have been developed and explained.

An individual potential outcome is the outcome that an individual would develop under a certain exposure level. Only one potential outcome can be observed – the one corresponding to the exposure received. The other potential outcomes are called counterfactual because they are "contrary to the fact" (contrary to what actually happened), thus unobserved. It is interesting to note that potential outcomes are conceived as fixed, a priori properties of the individual, existing before the individual receives any treatment (this is why they are potential).

Analogously, at the population level, one could define expectations of potential outcomes corresponding to each exposure level as the outcome expectations (proportions for binary

outcomes, arithmetic means for continuous outcomes) had all individuals in the study been exposed to each of the exposure levels. An *average treatment effect* in the study population is then defined as a contrast between two such potential outcome expectations, corresponding to the two different exposure levels compared (194).

Defining a causal effect as a contrast between two potential outcome values (or between two potential outcome expectations, at the population level) implies that causal effects are relative. There is no unique causal effect of a treatment, rather an infinite number of causal effects, depending on the reference used for comparison and on the population in which the comparison is made. When talking about the causal effect of a treatment one usually refers to the comparison between exposing individuals to the treatment versus not exposing them, all other things being equal. A causal effect of an active ingredient may be inferred from a comparison against placebo, where control patients receive the same pharmaceutical form or medical act, but missing the active ingredient, while a comparison against patient receiving no treatment would estimate a causal effect of the entire treatment (active ingredient(s) plus pharmaceutical form or medical act etc.).

Ideally, a study would contrast population level potential outcome expectations obtained by simultaneously exposing the entire population to each exposure level, but this is impossible since the same person cannot be observed at the same time under two different exposure levels. In real studies, one group of patients (i.e. cohort) would be observed under one exposure level while another group would be observed under another exposure level. If the two groups were selected at random, as in an RCT, the probability of outcome observed among those exposed to the active treatment should be the same as the probability of outcome had the entire study population been exposed to the active treatment, since the group selected to receive the active treatment would be a random sample (thus representative) of the entire study population. Therefore, the random sample exposed to the active treatment can be used to infer the potential outcome corresponding to the active treatment in the study population. The reference treatment group would also be a random sample. Hence, if the refence group was also to receive the active treatment, the same outcome proportion would be observed as in the active treatment group (in a very large study, with ignorable random error). This property is called *no unmeasured confounding* or *exchangeability* (195), because in the absence of confounding (as in an RCT), the two treatment groups can be exchanged for one another (the group allocated to the active treatment could receive the control treatment and the group allocated to the control treatment could receive the active treatment and the results should be the same). In observational studies, confounding is usually present, but exchangeability is assumed conditional on the set of measured confounders. This is called *conditional exchangeability*, and it essentially means that, conditional on (i.e. stratified on) the measured confounders, there are no other unmeasured confounders. This is an assumption that cannot be tested in the data, but rather relies on the judgement of the researcher about the causal structure of the study.

### 4.2.2 Directed acyclic graphs

Directed acyclic graphs (DAGs) are a useful tool for visualizing how the variables (measured or unmeasured) involved in a study could be connected in a causal structure, and how this structure could give rise to the observed associations. Visualizing possible causal structures behind the data provides intuitive explanations for confounding and selection bias. As such, DAGs will be used throughout the following sections where these biases are introduced.

A DAG consists of *nodes* which represent variables (usually abbreviated by capital letters), and arrows between these nodes (also called *directed edges*), which represent causal relations between variables, with the cause placed at the tail, and the effect placed at the tip of the arrow. For example, in Figure 4.2.2.1.A, L, A and Y are variables (the nodes of the graph), L is a cause of A, and A is an effect of L and a cause of Y. Going along edges, *paths* can be described between nodes. For example, from L to Y there is one direct path (the arrow from L into Y) and an indirect path going through (i.e. intercepted by) A. These are both *directed paths* because in both of them, starting from L, one goes in the direction pointed by the arrows to end up in Y. Thus, L is a cause of Y through both these paths. Causality "flows" through directed paths, meaning that directed paths describe causal associations between variables. The first node in a directed path is sometimes called an *ancestor* and the nodes on the path that are caused by the ancestor are called *descendants*. DAGs are *acyclic* because no directed path can return to (or end in) its origin node, meaning that a variable cannot cause itself. Paths that connect two nodes but contain arrows in opposite directions are called *backdoor paths*. For example, in Figure 4.2.2.1.A, the path starting from A, going to Y, intercepted by L is a backdoor path because it contains arrows in opposite directions. This path does not represent a causal relationship between A and Y but it will induce a statistical association between A and Y. A backdoor path is *blocked* if it contains nodes into which arrows collide. For example, in Figure 4.2.2.1.B, L is such a variable into which arrows coming from A and from Y collide. In this configuration L is called a *collider,* and the path between A and Y intercepted by L is blocked. Consequently, the backdoor path between A and Y that collides in L will not introduce an association between A and Y. However, as detailed in Section 4.2.3, conditioning on L will open the backdoor path and induce an association between A and Y (196,197).

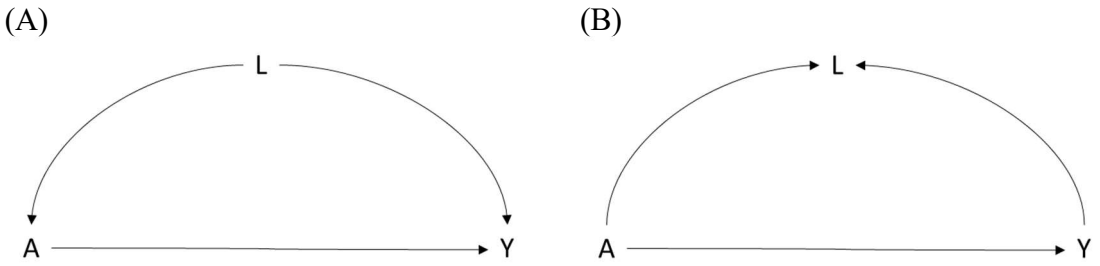


**Figure 4.2.2.1** – Simple DAG structures

DAGs can be thought of as qualitative illustrations of the structure that generated the study data. Looking again at Figure 4.2.2.1.A, one could describe a mathematical function that is used to generate variable Y, and, because according to the DAG, L and A are direct causes of

Y, they will be part of this function (for example, the function could be Y= 5A+2L). The DAG is a qualitative illustration of relationships between variables because the numerical coefficients that describe how A and L cause Y are not given in the DAG. The DAG only tells that "A and L are part of some function that defines Y". It is usually implicitly assumed that besides the variables of interest (endogenous), which are described in the DAG, there are other "external and unknown" (exogenous) causes of each variable (i.e. are part of the variables' function) and it is sometimes useful to add them to the DAG (see Figure 4.2.2.2). These external variables can be viewed as "error terms", which add random variability to the variables in the DAG. Adding error terms makes it more obvious that the causal relationships between variables is non-deterministic (stochastic). For example, the function which defines Y could be $Y=5A+2L+3U_Y$, and this means that for each fixed combination of A and L values, Y will take a variety of values as dictated by different values of $U_Y$ (198).

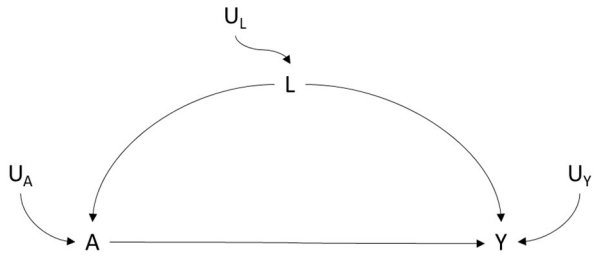


**Figure 4.2.2.2** – DAG structure with error terms

### 4.2.3 Selection bias

In epidemiology, the term "selection bias" has been used to refer to different issues. This section describes selection bias that can be illustrated in a DAG as conditioning on a collider, which opens a backdoor path, which distorts the association between exposure and outcome, such that the association no longer has a causal interpretation. Another mechanism described as selection bias is when the selection of the study population is dependent on effect modifiers (i.e. variables that interact with exposure changing the magnitude of its association with the outcome) which impacts the generalizability of results (199).

The basic collider structure is shown in the DAG in Figure 4.2.2.1.B. In this structure, the variables A and Y are both independent causes of a third variable L, which makes L a collider, and conditioning on L will open the backdoor path A>L<Y, inducing a non-causal association between A and Y, which combines with the existing causal association.

Intuitively, the association between A and Y conditional on L can be explained as follows. Say A, Y and L are all binary variables, and A and Y are the only two independent causes of L, the presence of either cause (A=1 or Y=1) increasing the probability that L=1 (compared to the absence of A and Y). Analyzing only observations with L=1 (i.e. conditioning on L), if A=0 then it is more likely that Y=1, since there is no other cause that could have produced L=1. Thus, conditional on L, knowing the value of A gives information about the value of Y (i.e. they become statistically associated). In this situation, where both A and Y were positively associated with L, conditional on L, A and Y become inversely associated.

One circumstance in which selection bias may occur is when exposure was initiated before baseline (prevalent exposure), thus potentially influencing the study population selection at baseline, and selection is also influenced by outcome risk factors (78,79). Similarly, a pre-baseline variable could act as a common determinant of baseline exposure and of baseline selection into the study. The latter situation is discussed a potential limitation in Study III where we compared the treatment response, measured as a function of DAS28, between patients who initiated infliximab with patients who initiated a combination of SSZ and HCQ. Patients in the SSZ + HCQ cohort were allowed to initiated the two treatments one after the other. The baseline of the study was set at the initiation of infliximab or the addition of the second drug in the combination. Thus, some patients in the SSZ + HCQ cohort were treated with one of these drugs before baseline, which was not allowed in the infliximab cohort, making such "pre-treatment" with SSZ or HCQ deterministic of exposure. Also, this pre-treatment may influence baseline DAS28 which was an inclusion criterium. By conditioning the inclusion in the study on DAS28, the pre-treatment would become associated with unmeasured common causes of baseline DAS28 and the outcome (see Figure 4.2.3.1). Since the pre-treatment was deterministic of exposure we could not adjust for it to close the biasing collider pathway (we could adjust for outcome risk factors, but unmeasured ones may remain).

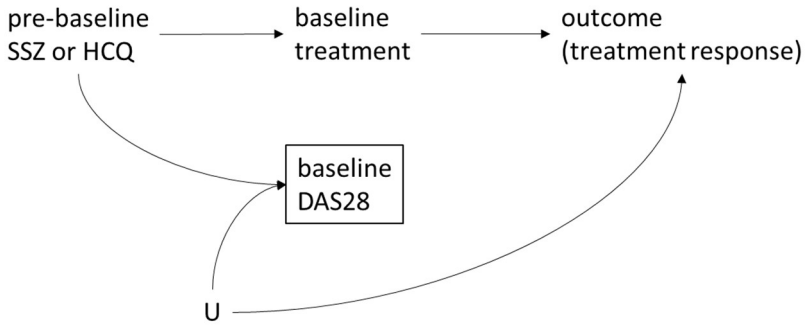


**Figure 4.2.3.1** – Possible selection bias structure in Study III

Informative censoring is another mechanism for selection bias.(200) If censoring is associated with the exposure, and it shares common causes with the outcome (i.e. informative censoring), then a backdoor path between exposure and outcome may be opened by studying only patients who were not censored before the time of outcome measurement. Informative censoring may occur in comparative effectiveness studies when only patients remaining on treatment to the time of outcome measurement are analyzed (see Figure 4.2.3.2). If the study treatment is stopped due to lack of effect (e.g. high disease activity (DAS28) despite treatment), then the treatment is a cause of censoring (via treatment stop). Besides the study treatment, there are other causes of disease activity (for example other treatments or biological characteristics of the patient (U)). Since the study outcome is also some measurement of disease activity at the end of follow-up, the outcome variable will share causes with the censoring variable, resulting in a collider structure that can introduce selection bias when studying only patients who remained on treatment.

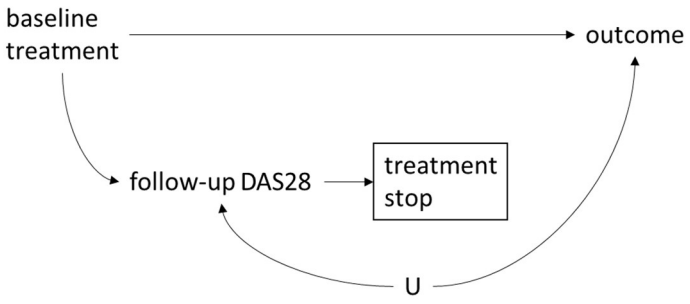


**Figure 4.2.3.2** – Selection bias through informative censoring via treatment discontinuation due to lack of effect (high follow-up disease activity (DAS28)) in an analysis of patients remaining "on treatment". The box around "treatment stop" means conditioning on this variable – i.e. analysing only patients who did not discontinued treatment before outcome assessment

One way to avoid selection bias due to censoring during follow-up (i.e. right censoring) is to avoid selection, that is to analyze the entire sample of patients that initiated the study without excluding anyone. However, for patients who switched treatment during follow-up, the outcome measured at the end of the study may reflect the effect of the latest treatment used instead of reflecting the effect of the study treatment (assigned at baseline). This may overestimate the effect of less effective treatments if they are more likely discontinued and replaced by effective treatments. Several alternative solutions to this problem exist (201). In studies II and III we imputed the treatment response of patients who discontinued treatment before outcome evaluation to negative responses, a method known as "non-responder imputation".

### 4.2.4 Confounding bias and the selection of variables for adjustment

Confounding bias is probably the most discussed source of bias in observational studies. The most basic DAG structure which describes confounding is represented in Figure 4.2.2.1.A, where A is the exposure variable, Y is the outcome variable, and L is a common cause of exposure and outcome. Because in this case L is a direct cause of both A and Y, it means that L is part of the functions that define the values of A and Y, thus determining A and Y to co-vary as L changes. Also, in DAG language, because the path between A and Y, which goes through L, contains arrows in opposite directions, it is a backdoor path, and since there is no collider on it, it is open. Therefore, the observed association between A and Y reflects both the causal relationship between them and the non-causal association via the open backdoor path. The confounding bias refers to the distortion of the causal association by the open backdoor path, via the common cause L in this case. The backdoor path can be blocked by conditioning on L (for example by studying the association between A and Y among observations with fixed values of L – i.e. stratification). The association between A and Y observed conditional on L will reflect the causal relationship between A and Y (198). In fact, while conditioning on

colliders (or descendants of colliders) found on backdoor path opens these paths, conditioning on non-colliders blocks the paths (202).

Other more complex structures can lead to confounding bias, and, in all these situations, open backdoor paths, as well as sufficient sets of variables needed to close these paths, can be identified using DAGs. In Figure 4.2.4.1 the backdoor paths A<$U_1$>L>Y and A<L<$U_2$>Y can both be blocked by adjusting for L, without any need to measure $U_1$ or $U_2$. The minimum set of variables needed to close all backdoor paths between A and Y contains only L.

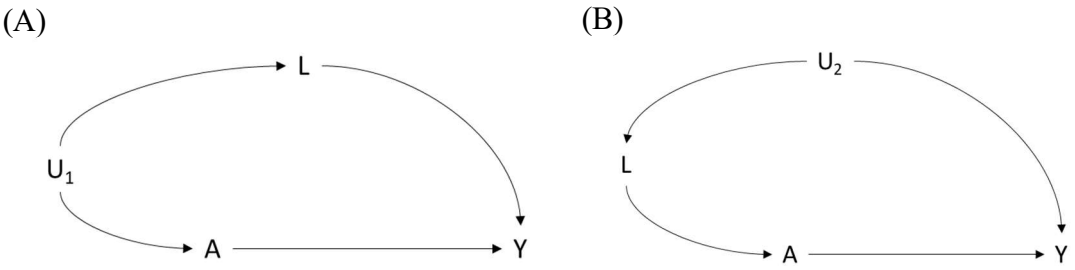


**Figure 4.2.4.1** – Confounding DAG structures, where A is the exposure of interest, Y is the outcome of interest, L is a measured variable and $U_1$, $U_2$ are unmeasured variables

DAGs can also be used to visualize why exchangeability does not hold in the presence of confounding variables. Take for example Figure 4.2.4.1.B. There is an arrow from variable L to exposure allocation (A), which means that exposure A is not allocated randomly but dependent on the value of L. Say that individuals with a higher value of L have a higher probability to be exposed to A. This means that exposed individuals will have, on average, higher levels of L than unexposed individuals. But L is also associated with the outcome Y via the common cause $U_2$. Say that individuals with a higher level of L are more likely to have the outcome Y. This means that individuals allocated to receive the exposure will have a higher probability to experience the outcome Y by virtue of being more likely to have higher values of L, compared to individuals allocated to be unexposed. This happens during exposure allocation, so even if nobody actually received the allocated exposure (or if exposure had no actual effect on the outcome Y – i.e. not arrow from A to Y), one would still observe an association between exposure allocation A and the outcome Y. This association "flows" through the open backdoor path A<L<$U_2$>Y, and makes the exposure cohorts not exchangeable with each other.

Nonetheless, in practice it is rarely possible to specify the complete DAG which generated the study data, and based on that to identify a minimum set of variables require for blocking all backdoor paths between exposure and outcome. Instead of using knowledge about the structure(s) that may have generated the data, some investigators choose to select covariates based on the statistical associations observed in the available data. For example, in one of the studies evaluating the adverse effects of glucocorticoids, the authors first estimated associations between covariates from a preliminary list and the outcome of interest in univariate regression models. In a subsequent step different outcome models were tested, including covariates one by one in a model of the outcome as function of exposure, and keeping only covariates which

altered the exposure effect estimate by >10% (86). Such practices have been criticized by proponents of causal inference methods since the same statistical associations may result from different data structures and adjustment for the same set of variables may reduce bias under certain structures and induce bias under others (203). For example, the same statistical association could be produced by the structure in figure 4.2.2.1.A or the one in 4.2.2.1.B. While in 4.2.2.1.A variable L is common cause of A and Y (confounder) and adjusting for L would reduce bias by closing a backdoor path, in 4.2.2.1.B variable L is a common effect of A and Y (collider) and adjusting for L would introduce bias by opening an otherwise closed backdoor path between A and Y. It may be argued that in order for L to be caused by both A and Y it should be measured after both of these, and measuring variables before baseline (i.e. before exposure assignment) would protect against dangerous adjustments. However, the classical example of the so-called "M-bias" in Figure 4.2.4.2 shows that this is not necessarily the case. In this figure, L is measured before A and Y and is associated with both, which are reasons to believe L is a common cause of A and Y, when in fact it is associated with A and Y via unmeasured common causes $U_1$ and $U_2$. On the other hand, in some situations, adjusting for variables measured after A may close backdoor paths, such as adjusting for L in Figure 4.2.4.1.A, where $U_1$ is not measured.

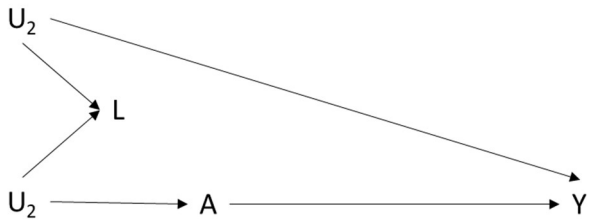


**Figure 4.2.4.2** – M-bias DAG structure where A is the exposure of interest, Y is the outcome of interest, L is a pre-treatment variable considered for adjustment and $U_1$, $U_2$ are unmeasured variables.

Besides introducing bias, the selection of covariates based on changes in the exposure coefficient over models containing different sets of covariates risks publication bias if researchers choose estimates which confirm their expectations, and is complicated by the so called non-collapsibility of certain relative risk measurements, which means that exposure coefficient estimates from models including different sets of covariates are not expected to be the same even in the absence of bias (204,205).

Even when the complete data-generating structure is unknown, DAGs can still be used to guide the rules for covariate selection. A principled covariate selection algorithm has been recently synthesized by VanderWeele (206). It amounts to selecting common causes of exposure and outcome (classical confounders), or proxies of these, as well as causes of outcome that are also associated with the exposure or causes of exposure that are also associated with the outcome. Even though, as mentioned above, some post exposure variables could be useful for

confounding adjustment, these may be influenced by exposure, thus selecting covariates from pre-exposure variables is recommended when the causal structure is not known.

When using methods based on the propensity score, which focuses on first modelling treatment assignment (see Section 4.2.5), one may be tempted to identify and adjust for all measured variables which determined treatment assignment (207). One may also have a better knowledge about variables that cause treatment assignment, since prescribers and patients could be interviewed to find out what drives their treatment decisions. However, adjusting for variables that are causes of exposure without being associated with the outcome through other paths than via exposure (also called instrumental variables or instruments) could amplify bias via unmeasured common causes of exposure and outcome (208,209). Identifying instruments from the data is difficult. The instrumental variable Z in Figure 4.2.4.3 is associated with exposure A (maybe known to be a cause of exposure). Testing the univariate association between Z and the outcome Y also suggests that Z is associated with the outcome. However, a confounder has to be associated with the outcome independently of exposure, so one might test the association between Z and Y among the unexposed. Because A acts as a collider, conditional on A the backdoor path Z>A<U>Y will be open as the directed path Z>A>Y closes. Thus, Z will appear associated with Y independently of A and it seems a good candidate for confounding adjustment based on this algorithm. This suggests that, when selecting covariates for confounding adjustment from variables that have a causal relationship to at least one of exposure and outcome (and are associated with the other via common causes), one should prioritize causes of the outcome and exclude known instruments and even variables strongly associated with exposure which are only weakly associated with the outcome (210,211).

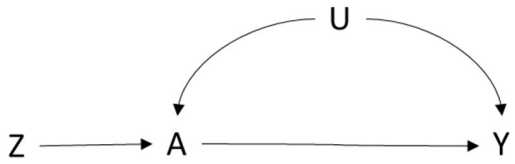


**Figure 4.2.4.3**– DAG structure where Z is an instrumental variable and there is unmeasured confounding (U) between exposure A and outcome Y.

Finally, even though the rules for covariate selection proposed by VanderWeele do not require complete knowledge about the entire data generating processes, they still require some external causal knowledge (e.g. that the covariate causes the exposure or the outcome). However, causal relationships are not universal, and it should be reasoned how causal information obtained from previous studies applies to the local context of one's study. For example, the socio-economic status may determine to a larger degree access to healthcare (and consequently health status) in countries with private healthcare systems compared to settings with universal, state sponsored, healthcare (212).

### 4.2.5 Inverse probability weighting and marginal structural modelling

This section provides some explanation of how IPTW and other related inverse probability schemes work.

IPTW is a propensity-score-based method used for confounding adjustment when estimating marginal structural models (MSM). An MSM is a model of potential outcomes as function of exposure (and optionally effect modifiers), that can be estimated in studies where treatment assignment was randomized, or in observational studies where confounding was adjusted for by IPTW, as described below (213,214). The term *structural* refers to the fact that the outcome model estimates a causal effect of exposure, and the term *marginal* refers to estimating an effect averaged (i.e. marginalized) over covariate patterns (i.e. population strata defined by set values of each measured confounder), since the model usually contains exposure and no other covariates, confounding adjustment being dealt with outside the outcome model by IPTW.

#### *The propensity score*

The *propensity score* was first defined in the context of binary treatments as the probability of being treated, conditional on a (sufficient) set of confounders. In observational studies the propensity score is unknown and needs to be estimated from the data. The propensity score is called a scalar (i.e. one-dimensional) balancing score. It is *one-dimensional* because it reduces a set of many variables necessary for confounding adjustment to a single variable (the propensity score itself), and it is a *balancing score* because it has been proven that, conditional on the propensity score, the treatment becomes statistically independent of the confounders included in the propensity score model (215,216). This is easily shown, since patients with the same propensity score have the same probability of being treated, regardless of their covariate patterns (i.e. combinations of measured characteristics). Thus, within strata of the propensity score, exposed and unexposed individuals have on average the same distribution of measured covariates, and are said to be *balanced* in terms of measured covariate distribution.

Evaluating the achieved covariate balance between treatment cohorts is a useful diagnostic after a propensity-score-based confounding adjustment. One commonly employed balance metric is the *standardized mean difference*. This is calculated for each covariate as the difference between the mean in the active treatment cohort and the mean in the control treatment cohort, divided by the squared mean variance (Equation 4.2.5.1). For binary variables the mean is replaced by the prevalence (proportion), and categorical variables with several categories can be first transformed into several binary dummy variables, applying the binary variable formula for each dummy (217).

Standardized differences can be calculated and compared before and after confounding adjustment as in Figure 4.2.5.1, which summarizes standardized differences calculated within each imputed data-set in Study III. In Figure 4.2.5.1 standardized mean differences for some of the covariates are represented by bars. This is because, due to missing data, standardized differences varied between imputed data-sets, the bars representing the observed range.

**Equation 4.2.5.1** – Standardized mean differences for a binary treatment

Continuous Variable

$$\frac{\mu_{k,active} - \mu_{k,control}}{\sqrt{\frac{s^2_{k,active} + s^2_{k,control}}{2}}}$$

Binary Variable

$$\frac{(p_{k,treatment} - p_{k,control})}{\sqrt{\frac{p_{k,treatment}(1 - p_{k,treatment}) + p_{k,control}(1 - p_{k,control})}{2}}}$$

$\mu_{k,active}$ is the mean of continous covariate $k$ among patients receiving the active treatment while $\mu_{k,control}$ is the mean of the same covariate $k$ but among patients receiving the control treatment; $s$ denotes corresponding variances of $k$ in the active and control treatment; $p$ denotes the proportion of a binary covariate.

(A) Before confounding adjustment (B) After confounding adjustment

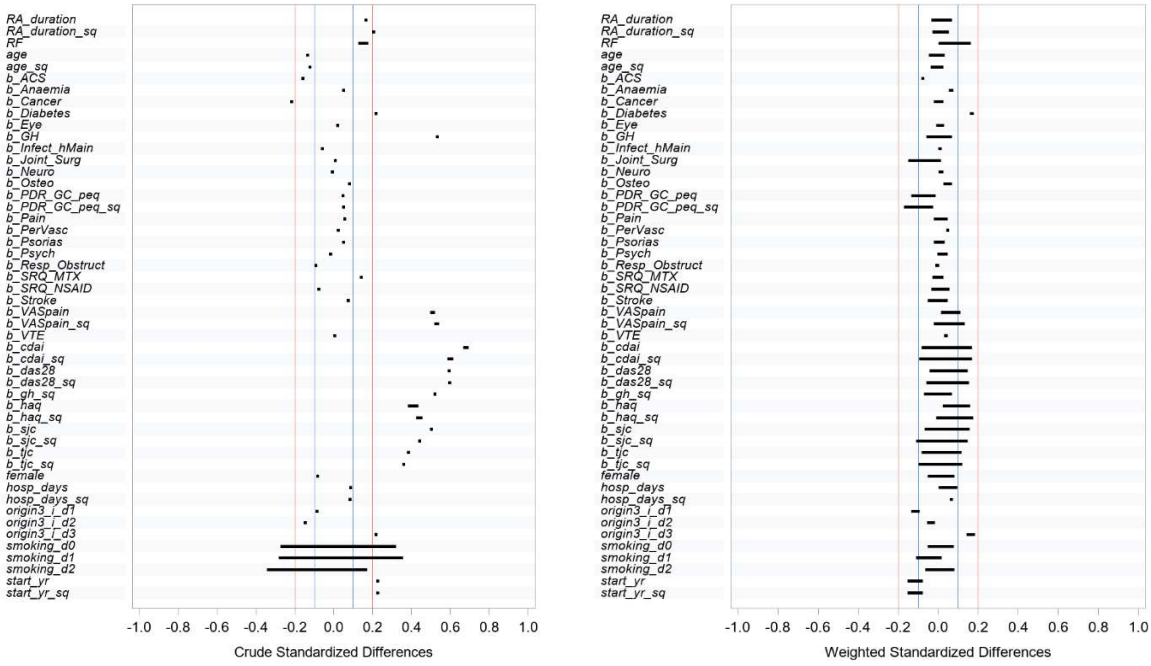


**Figure 4.2.5.1** – Standardized mean differences in Study III, before (A) and after (B) confounding adjustment. The threshlds of 0.1 and 0.2 are commonly employed to delineate balanced from unbalanced data and were marked here by blue and red lines respectively

For treatment variables with more than two levels, when the average treatment effect in the entire population is the target of estimation, the distribution of covariates within each treatment cohort can be compared to the marginal distribution of covariates in the entire study population.

Therefore, standardized differences can be calculated as unweighted/ weighted covariate means (or proportions) in each treatment cohort minus the unweighted overall population mean (or proportion), divided by the unweighted population standard deviation (218). This approach has been used in study I.

Once a suitable set of covariates has been identified, and the propensity score has been estimated, it can be used in various ways to adjust for confounding (217). It can be introduced alongside exposure in the outcome model, replacing all other covariates, thus reducing the model's dimensionality (i.e. the number of coefficients to be estimated). The analysis can be stratified by quantiles of the propensity score distribution. Exposed and unexposed observations can be matched by the propensity score. Finally, the propensity score can be used to calculate inverse probability of treatment weights (IPTW). Inverse probability of treatment weighting has been used as confounding adjustment method in studies I, III and IV for the following reasons:

1) It provides effect estimates marginalized (averaged) over covariate patterns (similar to randomized controlled trials). This was especially important in Study III, where results had to be compared to the SWEFOT trial.
2) It is easily extended to treatments with more than two levels (while this is more complicated for other propensity score methods such as matching). This facilitated the analyses in Studies I and IV, where exposure had more than two levels.
3) It may be generalized to time-varying treatments. This feature of IPTW has been used in Study IV, where histories of time-varying glucocorticoid doses were compared.
4) IPTW is an intuitive method, that can easily be implemented using standard software.
5) Contrary to matching, no observations are discarded, which was important for preserving sample size in Study III.
6) IPTW can be relatively easily combined with multiple imputation which has been used to address missing data in Studies I, II and III.

### *Inverse probability of treatment weighting*

For each observation in the study, the IPTW is the inverse of the probability of receiving the observed treatment level conditional on all variables considered sufficient for confounding adjustment. For a binary treatment, the IPTW for exposed observations is one divided by the propensity score while for unexposed observations it is one divided by one minus the propensity score. For rare treatments, the IPTW denominator (i.e. propensity score) may be small, at least for some patients, leading to high weights. In an alternative IPTW formula, called the *stabilized IPTW (sIPTW),* the numerator of 1 is replaced by the marginal probability of receiving the observed treatment level (i.e. the proportion of observations in the entire study population with the same treatment level) (219). The sIPTW produces weights closer to one, since the marginal probability in the numerator will be closer to the propensity score in the denominator for most observations. The sIPTW are thus preferred and we used them in our studies.

The sIPTW formula is easy to understand if one thinks about the goal re-weighting – to make treatment assignment independent of the measured confounders, that is to equalize the probability of being assigned to a certain treatment over all covariate patterns (220). A simulated numerical example in Figure 4.2.5.2 (based on Hernan and Robins (219)) illustrates how the sIPTW functions. There are 100000 patients, 37% of which have a high baseline disease activity, and 63% having a low disease activity. The proportion of patients receiving the treatment in the entire study population is 43% (19618 exposed in the low disease activity stratum plus 22901 exposed in the high disease activity stratum), and it is more probable to receive the treatment if one has a high disease activity (63% versus 31%). Disease activity is identified as a confounder. Thus, one wants to adjust for disease activity by making the probability of receiving treatment independent of disease activity. Calculating sIPTW for exposed participants, these are 0.43 / 0.31 among participants with low disease activity and 0.43 / 0.63 among participants with high disease activity. After reweighting each of the 31% exposed participants in the low disease activity stratum from a weight of 1 to a weight of 0.43 / 0.31, they will represent 43% of the stratum (0.31 * (0.43 / 0.31)). Similarly, after reweighting each of the 63% exposed participants in the low disease activity stratum, from a weight of 1 to a weight of 0.43 / 0.63, they will represent 43% of the stratum (0.63 * (0.43 / 0.63)). Thus, after reweighting, the probability of being exposed will be the same between the disease activity strata (43%), which is equal to the probability of being exposed in the entire population, and the similarly for the unexposed. Therefore, in the reweighted population, exposure was rendered independent of disease activity.

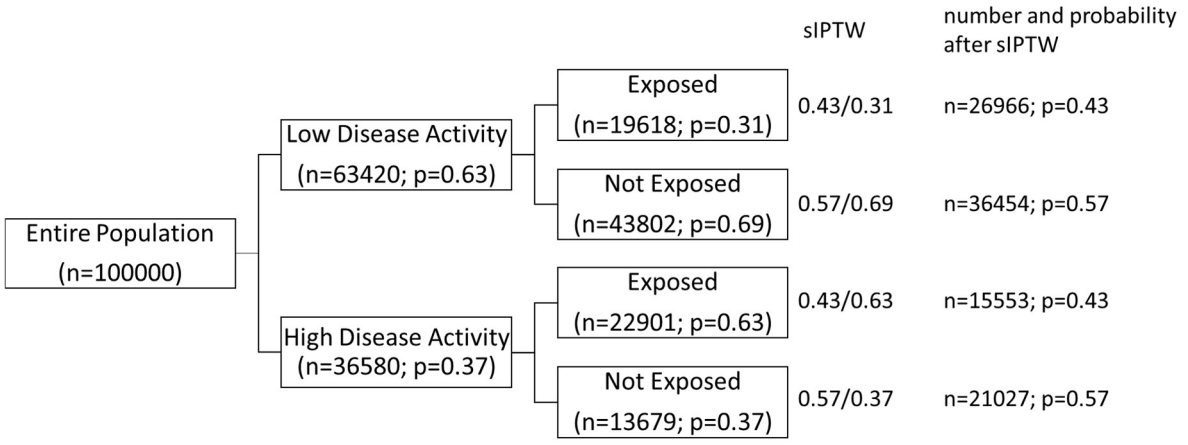


**Figure 4.2.5.2** – Rendering exposure independent of disease activity by stabilized inverse probability of treatment weighting (sIPTW). In the reweighted population the probability of exposure is the same under low vs high disease activity, thus exposure is independent of disease activity.

Even when using sIPTW, it is possible than some patients receive a treatment, used by a large proportion of the study population (high marginal probability in the numerator), but unlikely

for them considering their (measured) characteristics (low propensity score in the denominator), leading to high weights. This is what is called a near-violation of *positivity*. Alongside (conditional) exchangeability (see Section 4.2.1), positivity is one of the assumptions required to recover the potential outcomes from observational data and estimate causal effects in epidemiological studies (221). Positivity and exchangeability were combined under the assumption of *strong ignorability* by Rosenbaum and Rubin (216). Positivity refers to a non-zero probability of receiving each treatment level within each covariate pattern. If, for example, one of the study treatments was contraindicated in individuals with certain characteristics considered for confounding adjustment, then the potential outcome corresponding to the contraindicated treatment cannot be estimated in that part of the study population because there is no data for it (assuming the contraindication is perfectly respected in practice). To estimate the average causal effect in the entire study population, one must be able to estimate the causal effect in each covariate pattern of the study population (222). Near-violations of positivity imply that the probability to receive each treatment is non-zero under all measured covariate patterns, but it is low for some treatment levels under certain covariate patterns. In fact, when probabilities of treatment are estimated parametrically (for example, using logistic regression) they will be bounded away from 0, but they can be very close to 0. This leads to high sIPTW, because individuals who received unlikely treatments despite their characteristics have to be strongly upweighted to recover what would have happened to all individuals with those characteristics, had they all received the unlikely treatment (i.e. the potential outcome corresponding to the unlikely treatment). Using little information entails low precision (i.e. uncertainty in estimation), leading to large standard errors and wide confidence intervals. Weight truncation (for example at the 1$^{st}$ and 99$^{th}$ or 5$^{th}$ and 95$^{th}$ percentile of the observed weights distribution) has been proposed in order to avoid extreme upweighting of observations with rare exposure, but this involves allowing for some residual bias (222,223). We applied weight truncation in studies I, III and IV. Other weighting methods have recently been proposed in order to avoid extreme weights in time-fixed exposure settings (223). One example are *overlap weights*, which are bound between 0 and 1 (224).

It has been suggested that conditional exchangeability is less likely in covariate patterns where some treatments are unlikely. In other words, individuals who receive one treatment level despite a low probability of receiving it, as predicted by their characteristics, are likely a selected group of the respective covariate pattern (thus not representative of it) (225). Perhaps there is an unmeasured confounder which explains why they received an unlikely treatment. Several propensity score based trimming methods have been advanced, which exclude such observations, with the aim of reducing unmeasured confounding (226). In study III we have used the method proposed by Stürmer et al. where, for a binary treatment, observations with a propensity score lower than the 5$^{th}$ percentile among the active treatment or larger than the 95$^{th}$ percentile among the control treatment, were excluded from a sensitivity analysis.

*Adjusting for time-varying confounding*

The use of inverse probability weights has been extended by Robins et al. to the context of a time-varying exposure with time-varying confounding (227). A time-fixed exposure is set at baseline, and assumed to remain constant throughout the study. Because there is only one treatment assignment (at baseline), there is only one opportunity for confounding events to influence this treatment assignment. A time-varying exposure is allowed to change during follow-up. Thus, several treatment decisions are made throughout follow-up that could be influenced by events happening during follow-up, therefore time-varying confounding could ensue. If the events acting as confounders during follow-up are influenced by previous exposure (i.e. there is exposure-confounding feedback), then adjustment for such confounding during follow-up, by conditioning in conventional outcome regression models, could be problematic (228). This is illustrated in the DAG in Figure 4.2.5.3.A, where variables ($L_t$) measured before exposure at each time ($A_t$) act as confounders, thus should be adjusted for in the analysis. When modelling outcome $Y_2$ as a function of exposure history ($A_0$, $A_1$ and $A_2$) and confounders history ($L_0$, $L_1$ and $L_2$), the coefficients corresponding to the exposure variables $A_0$ and $A_1$ will be biased causal effects estimates in the presence of unmeasured common causes of $L_1$, $L_2$ and $Y_2$, because, in this situation, $L_1$ and $L_2$ act as colliders, and conditional on them opens backdoor paths. Moreover, $L_1$ and $L_2$ are mediators for the effects of $A_0$ and $A_1$ respectively, and conditioning on them will block part of the total effect of exposure history. One solution to this problem is to decouple outcome modelling from confounding adjustment, and this can be done via IPTW.

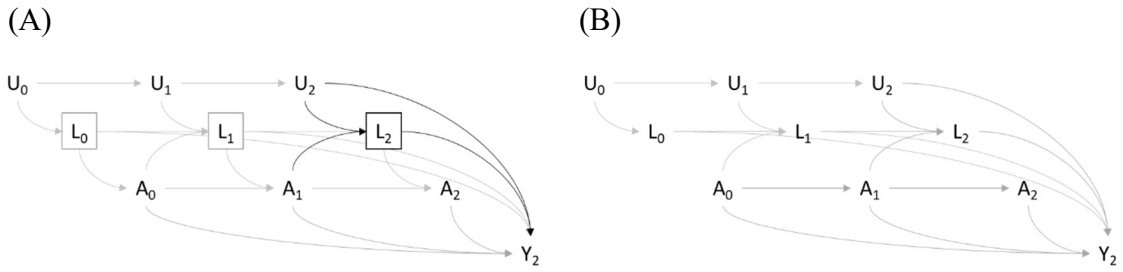


**Figure 4.2.5.3** – DAG structure for time-varying exposure and confounding as measured over three time points. The measurements of exposure at baseline and two subsequent times is designated ($A_0$, $A_1$, $A_2$), confounding measurements at the three time points are designated ($L_0$, $L_1$, $L_2$), and may represent vectors comprising many variables, unmeasured common cause of confounders and outcome are denoted $U_0$, $U_1$, $U_2$, and the outcome measured at the end of the study is $Y_2$. Panel (A) represents the structure before IPTW, and conditioning on time-varying covariates (border around $L_t$). Panel (B) represents the structure after stabilized IPTW, where the arrows between $A_t$ and previous $L_t$ were deleted

To adjust for confounding outside the outcome model, the probability of receiving the observed exposure has to be modelled as a function of past covariates (~propensity score), and used to calculate an IPTW for each observation, as described previously. In the case of a time-varying exposure, the process is similar to that for a time-fixed exposure, but instead of estimating only the probability of baseline exposure, the probability of current exposure at each time has to be

estimated, including past exposure in the model together with other potential confounders. This requires dividing the follow-up of each study participant into repeated observations (as we did in study IV, see Section 4.3.4). One can then use *pooled logistic regression* to model the probability of current exposure over all observations (i.e. times), including a flexible function of time in the model (229).

Reweighting observations by the inverse of the estimated conditional probability of current exposure would only adjust for confounding the current exposure at each time point, by rendering its probability independent of past exposure and confounding history. However, the aim is usually to estimate the joint effect of an entire exposure history (e.g. from baseline up to each follow-up time). To adjust for confounding the entire exposure history at each time, the IPTW for each observation (i.e. the contribution of an individual to a discrete follow-up time) is calculated as the product of inverse conditional probabilities of current exposure over all episodes belonging to the same individual up to and including the current episode (see Equation 4.2.5.2) (228). If the conditional current exposure probability model is correctly specified, then in the weighted population the entire exposure history at each time is rendered independent of previous exposure and confounding which is equivalent to an experiment in which exposure had been assigned randomly not only at baseline but at the start of each follow-up episode.

**Equation 4.2.5.2** – Inverse probability of treatment weight (IPTW) at time $t=T$, for a time-varying exposure $A_t$ and for an individual $i$

$$IPTW_i(T) = \frac{1}{\prod_{t=0}^{t=T} pr(\, A(t) = a_i(t) \mid \bar{A}\,(t-1) = \bar{a_i}(t-1), \bar{L}(t) = \bar{l_i}(t)\,)}$$

The denominator represents the product over time points, from baseline ($t=0$) to the current time ($t=T$), of the probability of the observed current exposure at each time $t$, for individual $i$ ($pr(A(t) = a_i(t))$, conditional on observed confounding ($\bar{L}(t) = \bar{l}_i(t)$) and exposure ($\bar{A}(t-1) = \bar{a_i}(t-1)$) history (the "bar" over A and L denotes a history)

The inefficiency of IPTW is accentuated in the setting of time-varying exposure, since multiplying inverse probabilities over episodes increases the possibility of obtaining extreme weights. Stabilized time-varying IPTW are used instead (227,229). As stated previously, to obtain sIPTW in a time-fixed exposure study, the 1 in the numerator is replaced by the marginal probability of observed exposure. In a time-varying exposure setting, the sIPTW numerator is a similar product of probabilities of current exposure as that in the denominator, the difference being that these probabilities in the numerator are conditional only on the exposure history and not on the confounding history (see Equation 4.2.5.3).

**Equation 4.2.5.3** – Stabilized inverse probability of treatment weight (sIPTW) at time $t=T$, for a time-varying exposure $A_t$ and for an individual $i$

$$sIPTW_i(T) = \frac{\prod_{t=0}^{t=T} pr(\, A(t) = a_i(t) \mid \bar{A}\,(t-1) = \bar{a}_i(t-1)\,)}{\prod_{t=0}^{t=T} pr(\, A(t) = a_i(t) \mid \bar{A}\,(t-1) = \bar{a}_i(t-1), \bar{L}(t) = \bar{l}_i(t)\,)}$$

sIPTWs are expected to vary around a mean of 1, and inspecting their distribution for extreme weights and skewness is a frequently employed quality check.(222) Figure 4.2.5.4 shows the distribution of stabilized inverse probability weights over follow-up time points in study IV. The dispersion increases over time, as more extreme weights are possible due to multiplication over an increasing number of episodes (i.e. observations), but interquartile ranges and means remain close to one. Also, sIPTWs conserve the original population size and the marginal distribution of exposure.

In the data weighted by sIPTW, the exposure at each time is rendered independent of the covariates used in the denominator model, conditionally on previous exposure, but it is not independent of previous exposure (Figure 4.2.5.3.B). The structure implies that the effect of exposure at time 2 ($A_2$) cannot be estimated without bias unless previous exposure ($A_1$) is included in the outcome model (because previous exposure acts as a confounder of the effect of exposure at time 2). Nonetheless, MSM are used when the joint effect of an entire exposure history is of interest, thus the entire exposure history would normally be included in the outcome model (i.e. in the MSM), the effect of each component of such history being adjusted by conditioning on previous components.

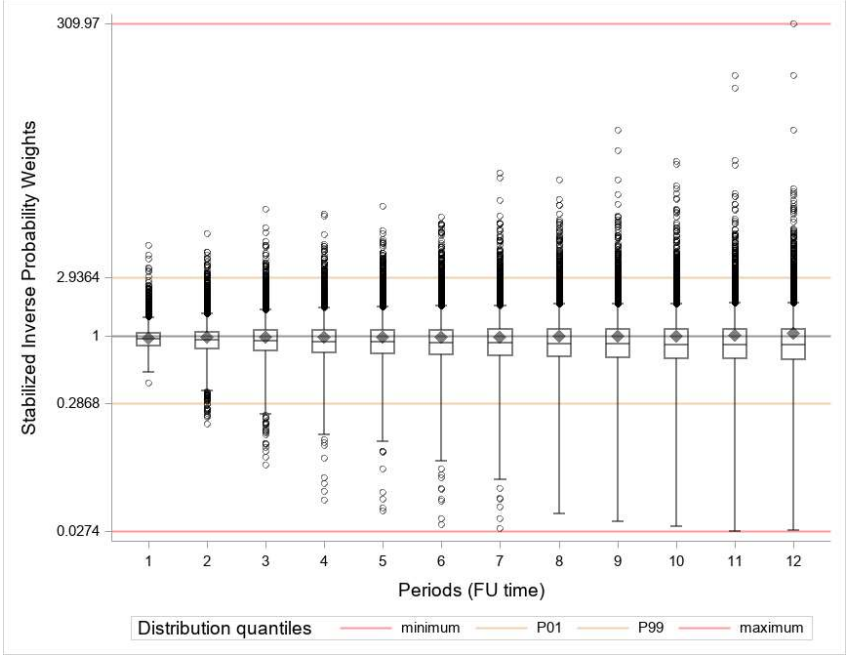


**Figure 4.2.5.4** – Box plots representing the distribution of time-varying stabilized inverse probability weights over the 12 follow-up periods in Study IV.

*Inverse probability of censoring weighting*

Inverse probability weighting can also be used to correct for selection bias via informative censoring, and it has been employed for this purpose in study IV. *Inverse probability of censoring weighting* (IPCW) is analogous to IPTW, but instead of modelling the probability of treatment, the probability of censoring is modelled as a function of covariates. The goal is to estimate the outcome as if nobody had been censored (a counterfactual quantity), using the outcome data from the persons who remained uncensored. If censoring would be statistically independent of the outcome (i.e. non-informative), then the observations which remained uncensored at each time would be a representative sample of the entire study population at risk at that time (i.e. risk-set), and their outcomes could be directly analyzed to infer the probabilities of outcome in the entire risk-set. If censoring is marginally informative, but it can be assumed that within levels of certain measured covariates (i.e. conditional on measured covariates) censoring is non-informative, then the observations not censored could be weighted, proportional to the inverse probability of not being censored, to create a pseudo-population where censoring is marginally non-informative (230). Intuitively, assuming that, conditional on some covariates, the uncensored data is representative of the entire data, then it can be re-weighted to fills in the "empty space" left by censoring.

The formula for calculating time- varying, stabilized IPCW is analogous to that for calculating time-varying sIPTW (see Equation 4.2.5.4), and the conditional probabilities of being censored can also be calculated via pooled binary logistic regression (229). The probability to be uncensored at time *t* is always calculated using observations from the previous time, since observations which make up a risk-set are the observations that have not been censored in the previous risk-set. At baseline no observation has yet been previously censored, thus the IPTCW at baseline is one for all observations. The same covariates used for correcting confounding bias in IPTW can be used for in IPCW models if these are measured risk factors for the outcome, thus potential common causes of outcome and censoring events. However, different adjustment sets can be selected.

**Equation 4.2.5.4** – Stabilized inverse probability of censoring weight (sIPCW) at time $t=T$, for a time-varying censoring variable $C_t$ and for an individual $i$

$$sIPCW_i(T) = \frac{\prod_{t=0}^{t=T} pr(\, C(t) = 0 \mid \bar{A}\,(t-1) = \bar{a}_i(t-1)\,)}{\prod_{t=0}^{t=T} pr(\, C(t) = 0 \mid \bar{A}\,(t-1) = \bar{a}_i(t-1), \bar{L}(t-1) = \bar{l}_i(t-1)}$$

To calculate the final *inverse probability weight* (IPW) which corrects for both selection bias via informative censoring and for confounding bias, IPTW and IPCW calculated for each observation are multiplied (229).

### *The target population*

IPTWs, as described above, standardize the distribution of covariates in each treatment group to that in the entire study population, allowing the analyst to compare the expected outcome had the entire study population been exposed to one treatment versus the expected outcome had the entire study population been exposed to another treatment. A causal effect estimated in the entire study population is called an *average treatment effect* (abbreviated ATE). Thus, by using sIPTW in the studies I, III and IV we estimated ATEs.

Weights targeting other populations of interest have been proposed. For example, *standardize mortality ratio* (SMRs) weights standardize the covariate distribution in the control cohort to that in the treatment cohort. Thus, these weights target the estimation of an *average causal effect among the treated* (abbreviated ATT), which compares the expected outcome observed among individuals exposed to the active treatment with the counterfactual expected outcome had the individuals in the active treatment cohort been exposed to the control treatment instead. The denominator of the SMR weights is the same as that of IPTW, but the numerator is always the probability of receiving the active treatment conditional on measured covariates (i.e. the propensity score in the case of a binary exposure) regardless of the exposure actually received. Thus, for individuals receiving the active treatment, the SMR weights will be equal to 1, and nothing changes in this cohort. The individuals receiving the control treatment will be reweighted to have the same covariate distribution as individuals in the active treatment cohort, thus achieving balance in measured confounders (231).

Using the same logic as SMR weights, *inverse odds of sampling weights* (IOSW) were derived to standardize the covariate distribution of a study to that of an external target population, with the purpose of comparing average treatment effects in the two populations (232). To calculate IOSWs one needs to combine individual data from the study population and the external target population, measuring exposure, outcome and covariates in the same way throughout the two data-sets. Then IOSWs are estimated in the combined data-set, but only observations from the study population are re-weighted and kept in the analysis. The IOSW denominator contains the probability of belonging to the study, conditional on the measured covariates, and the numerator contains the probability of belonging to the external target, conditional on the same covariates. To ensure a balanced distribution of measured covariates between treatment cohorts, the standardization can be conducted separately for each treatment cohort in the study population. Similar assumptions to those required for confounding adjustment are required for standardization to the covariate distribution of an external population. There should be no unmeasured "confounders" of study participation. This means that, conditional on the measured covariates, there are no unmeasured predictors of being included in the study that are also associated with the outcome, or unmeasured causes of the outcome that are associated with being included in the study. Also, there should be study participants, under each exposure level, within all covariate patterns considered for standardization. This is analogous the positivity assumption discussed for confounding adjustment and it ensures the availability of study data

for each covariate pattern of interest in the external target population (233). Standardization by weighting has been applied in a sensitivity analysis of study III, however it was not included in the final published manuscript.

### 4.2.6 Handling missing data and multiple imputation

Missing data refers to the situation where the values of certain variables are absent for some observations in a data-set. The reasons for missing data can be many. For example, the values have not been recorded, the values have been lost or they have been deleted in the process of data management because they were considered incorrect.

#### *Multiple Imputation*

By default, most statistical analysis software analyzes only observations with complete data for all variables included in the analysis, but such analyses may lead to biased estimates as explained below. One flexible and robust method for handling missing data in order to avoid bias is multiple imputation, which we have employed in studies I, II and III (234,235). Multiple imputation can be seen as analogous to inverse probability of censoring weighting in the sense that, under certain assumptions, it uses the observed data to recover the joint distribution of the missing data. There are several versions of multiple imputation, one of them being MICE (Multiple Imputation with Chained Equations, also known as Fully Conditional Specification (FCS)), which is the method we have used. According to MICE, for each variable with missing data, separate conditional models are specified as function of other variables in the data-set (as opposed to specifying a joint distribution for all variables with missing data conditional on observed data) (236). Multiple imputation involves randomly drawing the missing values from their estimated conditional distributions. The parameters defining these conditional distributions are themselves randomly drawn from their estimated distributions. This simulation process is repeated a number of times, leading to several imputed data-sets with slightly different imputed values. Randomly drawing the conditional distribution parameters and then the imputed values is meant introduce uncertainty in the imputation process reflecting the fact that imputed values are estimated (i.e. guessed) rather than known (236–238). Each imputed data-set is analyzed separately, and the results are finally pooled together using a set of equations known as "Rubin's rules" (239). According to these rules, point estimates are averaged over imputations to obtain the pooled point estimate, and the variance around the pooled estimate is obtained by summing up the average within imputation variance with a between imputation variance term (240). The addition of a between imputation variance component is what distinguishes multiple imputation from single imputation. When propensity score methods (including IPTW) are used, the entire analysis, from IPTW estimation to outcome model estimation, is conducted within each imputed data-set, and only the final estimates of interest and their standard errors are pooled (241).

*Missingness generating mechanisms*

To evaluate whether a complete case analysis or a multiply imputed data analysis would provide unbiased results, one should understand the mechanism under which the missing values were generated. It is clearer to visualize data-generating mechanisms using DAGs, and these can be employed for missingness as well (242,243). Traditionally, missingness generating mechanisms were grouped into three classes: missingness completely at random, missingness at random and missingness not at random (239).

1) The DAG in Figure 4.2.6.1 describes a *missingness completely at random* (MCAR) process. It could be a situation when some of the values in variable L were randomly deleted by a computer error. In such a situation, the complete cases (i.e. observations without any missing value) are a random sample of the entire data (there are no arrows into missingness from any variable of interest to the study) and analysing them (i.e. a complete case analysis) would not introduce bias.

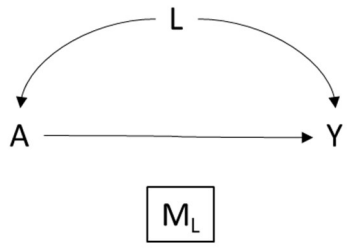


**Figure 4.2.6.1** – DAG structure for a study of the relationship between exposure A and outcome Y, confounded by L and with data missing in L ($M_L$ is the missingness indicator), generated completely at random (MCAR)

2) The DAGs in Figure 4.2.6.2 describes *missingness at random* (MAR) processes, since $M_A$ is marginally associated with A, but it is independent of A conditional on observed data. Multiple imputation requires that missingness for the imputed variables was generated under MCAR or MAR (239). This is required since the imputed values are sampled randomly from the distribution of the imputed variable, conditional on other measured variables. If missingness would not be conditionally random, then the random imputation would be biased. To ensure MAR under the structure in Figure 4.2.6.2.A, one only needs to condition on C. To ensure MAR under the structure in Figure 4.2.6.2.B, one has to impute A conditional on both C and D, since Y has to be introduced in the imputation model to preserve the association between A and Y, and, conditional on Y, the path A>Y<D>$M_A$ is opened, thus needs to be closed by conditioning on D. A complete case analysis would not introduce bias under the DAG in Figure 4.2.6.2.A, since conditional on $M_A$, no backdoor path is opened betwwen A and Y. The same is not true in Figure 4.2.6.2.B where $M_A$ is a collider, and conditional on $M_A$ the backdoor path A<C>$M_A$<D>Y is opened. Analysing the complete-case data conditional on D or on C (as well as ob confounder L) would provide unbiased conditional causal estimates for the association between A and Y. However, the marginal association between A and Y cannot be estimated from complete cases since

D (and C) directly determines missingness, thus the marginl distribution of D (or C) cannot be estimated, and it is needed for marginalizing over D (or C) (242).

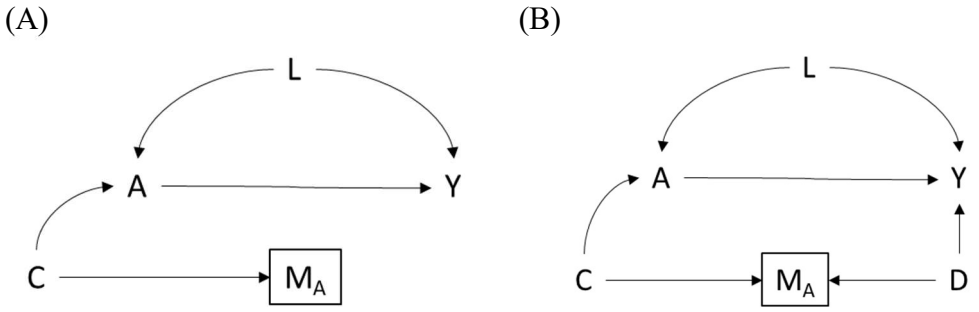


**Figure 4.2.6.2** – DAG structures for a study of the relationship between exposure A and outcome Y, confounded by L, and with data missing in A ($M_A$ is the missingness indicator), generated under MAR

3) The DAGs in Figure 4.2.6.3 describes *missingness not at random* (MNAR) processes. This is the case when missingness depends on unmeasured variables or on the values of variables with missing data themselves. Since in Figure 4.2.6.3.A missingness in A ($M_A$) depends on the value of A itself there is not variable that one could condition on to render $M_A$ independent of A (239). It is commonly believed that under MNAR complete-case analyses are biased, but under the structure depicted in Figure 4.2.6.3.A this would not be the case, since conditioning on $M_A$ does not open any backdoor path between A and Y. On the other hand, multiple imputation would be biased without additional information about the about the location of missingness within the distribution of A. In Figure 4.2.6.3.B, the missing data in Y ($M_Y$) cannot be imputed without bias for the same reason. Moreover, in this setting, the complete case analysis will be biased, since conditional on $M_Y$ one opens backdoor paths between A and Y via any unmeasured cause of Y ($U_Y$). This is a situation in which includeding external causes of variation for the variables in the DAG (such as in Figure 4.2.2.2) can uncover additional biasing paths.

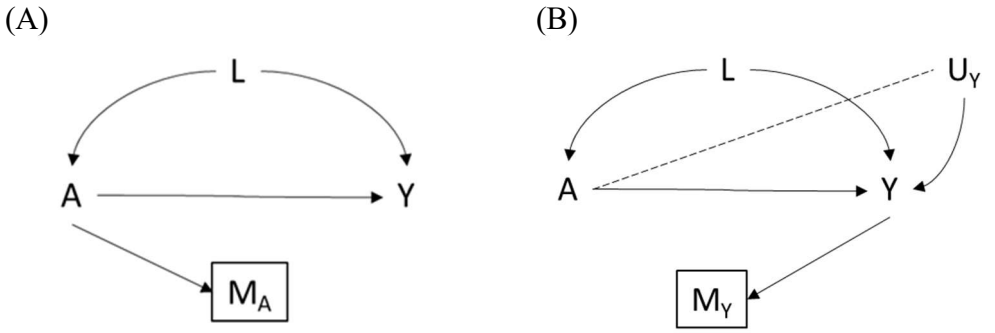


**Figure 4.2.6.3** – DAG structures for a study of the relationship between exposure A and outcome Y, confounded by L and with data missing on A ($M_A$) or in Y ($M_Y$), generated under MNAR

To achieve MAR, it is recommended to include as many variables as possible in imputation models (239). Besides achieving MAR, the set of variables included in the analysis should also be included in the imputation models in order to preserve the correlations between them. It is especially important to include the outcome variable. In survival analysis one proposed solution, which has been used in studies I and II, is to include the binary event indicator together with the Nelson-Aalen cumulative baseline hazard estimate (238). Other variables such as predictors of the imputed variables or of their missingness indicators are also commonly included (240). The intuition behind including as many such variables as possible is to block any potential associations between the imputed variable and their missingness (achieve MAR). However, as discussed in Section 4.2.3, conditional on some variables (colliders) associations may also be opened. Furthermore, large imputation models may have convergence problems in practice.

Besides respecting the MAR assumption, the imputation models must also be correctly specified. Imputation models must be compatible (a.k.a. congenial) with the analysis model. Therefore, when modelling interactions or non-linear associations, these must also be reflected in the imputation process. Nevertheless, correct model imputation is not always straightforward in these cases. For example, when modelling continuous variables using polynomial terms (e.g. square terms), a continuous variable may be first imputed and then transformed, conserving the mathematical relationship between the variable and its square transformation. However, in this case, imputation models would contain only the linear term of the variable, thus the non-linear relationship between variables would not be conserved in the imputations. Alternatively, the variable is first transformed and both variable and its transformation are imputed as separate variables, thus allowing for the transformations to be present in the imputation models. However, in this case, the imputed values and their imputed squares may not always correspond (244,245). The latter method was used in the studies included in this thesis. More complex algorithms, aimed at preserving the relation between variables and their transformations, as well as obtaining unbiased estimates of interest, have been proposed, but their implementation is more time-consuming (246).

### *Missing data in time-varying settings*

While multiple imputation had been mainly explored in studies of time-fixed exposure, its use can be extended to time-varying exposures (247). Nevertheless, depending on the number of observations, variables included in imputation models, proportion of missing data etc. the process of multiple imputation can become very time-consuming. While creating the imputed data-sets may be parallelized, each of these parallel subprocesses still involves iterative model estimation (known as “burn-in iterations”). For example, in study IV, the estimated time for analysis of 20 imputed data sets exceeding one week, due to the large number of observations created by dividing the follow-up of 9639 participants into up to 12 observations per individual, hence accumulating 110150 observations.

Another method for handling the missing data would have been to censor persons-time at the first occurrence of a missing value, using IPCW to correct for selection bias due to censoring.

This was not feasible either since, at the second time-point, more than 50% of participants would have been lost.

Instead, to handle missing data in covariates DAS28 and HAQ in study IV we used two alternative methods. Both methods make strong assumptions about the unobserved values. The first method was *last observation carried forward* (LOCF). This is a common method for imputing missing values in longitudinal studies due to its simplicity. Whenever an observation is missing it is replaced by the latest earlier observation for the respective variable and individual. This method does not incorporate any uncertainty due to guessing the imputed value since it creates one imputed data-set. Instead, it is assumed that, whenever a value is missing, it is because the value of the variable did not change since the previously available measurement. Under this assumption the missing values are not actually missing, they are known. A LOFC analysis is unbiased only if the missingness generating process respects this assumption (247). In order to avoid missing values entirely, eligible patients had to have a measurement of DAS28 and HAQ at the baseline visit. However, it might be unrealistic to assume that treatment decisions several months after baseline were based on baseline DAS28/HAQ, when no updated values were recorded.

The second method, the *missingness patterns* (MP) method, implies that only variables with observed values were used in the treatment decision process (247,248). To analyze the data under this assumption, we only kept DAS28/HAQ measurements within 80 days before to 10 days after the start of each exposure assessment window, and we divided the data-set into four patterns corresponding to the 4 combinations of missingness: (1) both DAS28 and HAQ observed; (2) DAS28 observed and HAQ not observed; (3) DAS28 not observed and HAQ observed; (4) DAS28 and HAQ not observed. We estimated separate IPTW (and IPCW) models in each pattern, and we calculated weight components at each time. Then, we pooled the data back together and we calculated IPW (i.e. products of IPTW and IPCW components over time points, for each individual) as described in Section 4.2.5.

### 4.2.7 Survival analysis

Survival analysis refers to a collection of methods used for analyzing data from studies where participants experience the outcome event at various times during the study's follow-up. Thus, survival analysis may be described as studying the timing of outcome events relative to baseline (for this reason it is also known as "time to event" analysis). The term "survival analysis" comes from the situation where the event of interest is the death of the participant, but the same methods can be applied to a variety of other events. In the studies included in this thesis, events of interest were gastro-intestinal perforations in Study I, treatment discontinuation in Study II, and serious infections in Study IV.

In survival analysis participants are followed only to the first occurrence of an event of interest. This is the only possibility when the event of interest is death. For other events, where the participant could be followed after the occurrence of the first event (for example after an initial non-lethal infection) their observation time (follow-up) is stopped after the initial event. When

other events stop follow-up before the occurrence of an outcome event, it is said that the observation has been *censored*. Censoring can happen for various reasons such as:

1) "administratively", when follow-up reaches the end of the available data
2) "loss to follow-up", when the participant leave the data collection process (for example, data can no longer be collected in Swedish national registers once the individual emigrates)
3) "competing events", for example if they die from another causes before the outcome event takes place.

If the follow-up of an individual is censored before the outcome event, then the exact survival time is not known for that individual, but it is known to be greater than the censoring time (i.e. the outcome may have occurred after the censoring event).

Two risk measurements commonly estimated in survival analysis have been employed in the included studies. The *hazard* is mathematically defined as the limit of the probability of experiencing the outcome event in a time interval, conditional on not having experienced the outcome event previously, divided by the length time interval, as this time interval approaches zero (Equation 4.2.7.1) (249).

**Equation 4.2.7.1** – The hazard function

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$T$ is the time of event and $t$ is the time of measurement

In practice time is measured in discrete units. The unit of measuring time in Swedish registry data is the day and in study IV the time unit used was a 90-day episode. The *discrete time hazard* is the proportion of participants who experienced the outcome event at a certain discrete time (e.g. during the second 90-day follow-up episode in study IV), among those who were still at risk at this time (i.e. have not experienced the outcome event or a censoring event at earlier times) (250). Patients at risk at a specific time are called the *risk-set*. To compare hazards between treatment groups, hazard ratios are commonly estimated from Cox proportional hazards models, as has been done in studies I and II (251). Hazards can increase and decrease over time, since both the denominator (the risk-set at time *t*) and the numerator (the number of outcome events experienced by patients in the risk-set at time *t*) vary over time, and there is no reason for hazard ratios to remain constant. However, in order to provide a single treatment effect estimate for the entire follow-up, the proportional hazards model assumes a constant hazard ratio, which is a weighted average of time-specific hazard ratios. This modelling assumption is similar to not modelling interactions between treatment and other participant characteristics, such as sex. One average treatment effect is presented, even though the treatment effect may vary between males and females.

One limitation of hazard ratios as causal effect estimates is an inherent selection bias (252). This is due to the fact that hazards are estimated in selected populations, more exactly among patients who have not experienced the outcome at previous times. If treatment has an effect on experiencing the outcome, and experiencing the outcome at an earlier time shares causes with experiencing the outcome later, then a backdoor path between exposure and outcome is opened as suggested in Figure 4.2.7.1. In study IV, censoring at the first outcome event has been rendered independent of measured outcome risk-factors by IPCW. However, the inherent selection bias of hazard ratios is very difficult to correct since there would always remain unmeasured risk-factors for the outcome.

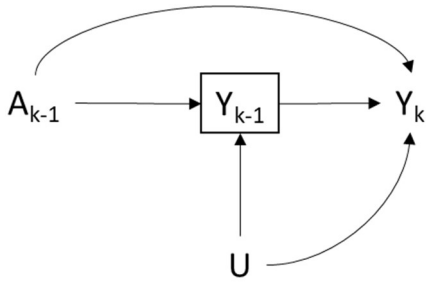


**Figure 4.2.7.1** – Selection bias associated with the hazard ratio

If the target of estimation is the effect of exposure A at time (k-1) on the outcome Y at a later time (k), and this is estimated as a hazard ratio, than participants exposed to A=1 and A=0 at (k-1), who have not experienced the outcome Y (or a censoring event) at (k-1), are compared in terms of outcome probability at time (k). If A has an effect on Y, by conditioning on collider $Y_{k-1}$, the backdoor path $A_{k-1}>Y_{k-}1<U>Y_k$ is opened, adding to the direct causal association, thus producing a biased observed association between $A_{k-1}$ and $Y_k$

A suggested solution to this issue is to estimate another risk measurement – the *cumulative incidence* (250,252). The cumulative incidence is the complement of the cumulative survival probability. The *cumulative survival probability* at a certain time *t* during follow-up is the proportion of study participants who did not develop the outcome event up to (and including) time *t*. Thus, the cumulative incidence at time *t* is 1 - *cumulative survival probability*, which represents the proportion of individuals who experienced the outcome event at any time up to (and including) time *t*. Survival probabilities can be estimated non-parametrically using the *Kaplan-Meier method* (also known as the product limit method) (253). The method starts by discretizing time (into very small pieces) and calculating the probability of surviving each time period conditional on having survived to the start of the period (i.e. being part of the risk-set). These *conditional survival probabilities* are the complements of discrete time hazards. Then, the *cumulative* (or marginal) *survival probability* at a time (discrete period) *t* is the probability of having survived jointly period *t* and all previous periods, and this is calculated as the product of conditional survival probabilities over all these periods. Contrasts of survival probabilities and cumulative incidences between exposure levels are not affected by the same selection bias as hazard ratios since, contrary to the hazard, they are marginal probabilities with respect to

time, always referring to the baseline population (the denominator of these probabilities is the entire population at baseline) and not to a selected risk-set.

In study IV the cumulative incidence over the course of follow-up was calculated in a similar manner to the product limit method described above, but conditional survival probabilities were estimated parametrically using pooled logistic regression to model the probability of experiencing the outcome as a function time and exposure history. This was the MSM estimated in the IPW data-set, where each observation represented the contribution of one person to a discrete time episode (254).

In both the Kaplan-Meier method and the method used in study IV, censoring is accounted for implicitly as censored participants are dropped out of the study and do not participate in future risk-sets. However, the conditional survival at each time is calculated as 1-hazard, seemingly ignoring loss of participants through anything other than the outcome event. In fact, if censoring is assumed independent of the outcome (i.e. non-informative), censored participants would have experienced the outcome with the same probability as those not censored. This is equivalent to estimating a causal effect in a theoretical population in which the censoring did not take place. The interpretation of counterfactual outcomes defined by setting exposure to the study levels and censoring to zero (i.e. no censoring) might be problematic when the censoring event is death by other causes (than the outcome of interest), since intervening to prevent death by other causes in the whole study population might not be realistic (255). If censoring informative, selection bias could occur through a similar mechanism as described in Figure 4.2.7.1. IPCW addresses this issue, rendering censoring independent of measured covariates that are common causes of censoring and the outcome.

## 4.3 STUDY DESIGN AND ANALYSIS

This section will briefly introduce the design and analysis for each of the four included studies. More detailed information can be found in the attached articles.

### 4.3.1 Study I – Biological DMARDs and the risk of gastro-intestinal perforations in RA

Study I was a population-based cohort study primarily designed to compare the incidence of GI perforations between RA patients who initiated TNFi versus non-TNFi bDMARDs, with a special interest in tocilizumab. We included RA patients who have not yet initiated bDMARDs (bionaïve) and matched general population comparators as secondary reference groups.

First, we identified RA patients in NPR using a previously validated definition (256). Next, we identified bDMARD initiations after January 2009 (such that all bDMARDs under study were on the market during the study period) in SRQ/ARTIS. Five general population controls had previously been matched by sex, age and geographical location to bDMARD treated patients.

bDMARD treated patients were followed from the date of treatment initiation up to 90 days after a decision to stop treatment or to switch to another bDMARD. The 90-day extension was employed to ensure that GI perforation events that may have led to a decision to stop treatment

were captured. Consecutive treatment episodes with the same active substance were merged if treatment was restarted within 90 days (270 days for rituximab). One patient could participate with several follow-up episodes if they switched bDMARDs, but only the first episode with each bDMARD was included. Patients were followed as bionaïve from the date when they first fulfilled the RA diagnostic criteria, or from January 2009 if the criteria were fulfilled before this date, up to the date when thy first initiated a bDMARD treatment. General population controls were followed from index-date (which was the date when the matched bDMARD treated patient initiated the first bDMARD) or from January 2009, up to the date when they develop RA. All patients were followed to first outcome event or to death, emigration or the end of available data (31 December 2017), if these came first.

The outcome event was defined as hospitalization with a main or secondary diagnosis of GI perforation (according to a prespecified list of ICD-10 codes) or death due to GI perforation (primary or contributory cause of death). Perforations were classified as belonging to the upper GI tract if located in the esophagus, stomach or duodenum, and otherwise belonging to the lower GI tract. Lower GI perforations made up 85% of all perforations observed during the study period, and since these were the main concern in relation to tocilizumab (109,111), we studied them as the primary outcome, with all (upper + lower) GI perforations assessed in supplementary analyses.

The primary analysis was adjusted, using IPTW, for pre-selected baseline patient characteristics which were considered risk-factors for GI perforation according to previous literature (99,103,257) or to medical knowledge: demographic characteristics (age, sex, education level), comorbid conditions (a history of GI perforations, diverticular disease, intestinal ischemia, inflammatory bowel disease, hospitalized infections, chronic obstructive pulmonary disease as a proxy for smoking, diabetes, cancer), markers of RA severity (number of joint surgeries, CRP, ESR, DAS28, HAQ and RA duration) use of csDMARDs, NSAID and glucocorticoids exposure before baseline. Additionally, we adjusted for line of bDMARD treatment (i.e. how many previous bDMARDs have been used) and for the year of treatment start. Missing covariate data were imputed using multiple imputation (MICE – described in Section 4.2.6) and the imputed data-sets were analyzed using IPTW generalized estimating equations Poisson models to obtain adjusted incidence rates, Cox regression (with robust standard error estimation) to compare hazard rates between treatment groups and the Kaplan-Meier method to estimate adjusted survival curves for each treatment group.

### 4.3.2 Study II – Comparative effectiveness of baricitinib, tofacitinib and biological DMARDs in RA

Study II was a cohort study designed to compare the effectiveness of baricitinib and tofacitinib with that of bDMARDs among RA patients.

All patients with a primary diagnosis of RA, and who initiated bDMARDs or JAKi between January 2017 and November 2019, were included in the study, irrespective of line of therapy. The start date was chosen such that all drugs were available during the study period (first JAKi

approval in EU was in 2017). The end of the study period was chosen such that all participants could be followed at least up to the one-year end-point (unless censored by death or emigration), considering that we had data available until February 2021.

We compared the two JAKis (baricitinib and tofacitinib) with abatacept, rituximab, two IL-6 inhibitors (tocilizumab and sarilumab), and five TNFi (etanercept, adalimumab, infliximab, certolizumab pegol, golimumab). We identified start (i.e. decision to initiate) and end of each treatment episode, as well as reasons for discontinuation, in SRQ. Consecutive treatment episodes on the same active substance were merged if gaps between them were not longer than 90 days (270 for rituximab). Under each treatment level we included all dosing regimens used in practice. One individual could contribute with several observations if they switched treatments (~18% of patients participated to more than 1 treatment cohort).

Follow-up for each observation started at treatment initiation (i.e. baseline) and, in the primary analysis, patients were followed for one year, at which point three binary treatment response measures were evaluated: EULAR DAS-28 good (vs moderate or no) response, defined as a DAS28(ESR) $\leq$ 3.2 units at evaluation and a decrease in DAS28(ESR) > 1.2 units at evaluation versus baseline; HAQ-DI improvement, defined as a decrease in HAQ-DI > 0.2 at evaluation versus baseline; and CDAI remission, defined as a CDAI $\leq$ 2.8 at evaluation. Since in real clinical practice patients are not evaluated at fixed time-points during their treatment courses, we measured the one-year end-pints within a window spanning from 275 to 455 days after baseline using the measurement closest to 365 days or the mean of two measurements if symmetrically situated around this point. Patients who discontinued treatment before one-year were classified as "non-responders". Treatments were considered discontinued at the initiation of another bDMARD or JAKi, if these happened before the recorded discontinuation date for the current treatment. Patients who discontinued treatment due to remission were considered on-treatment until the start of a new bDMARD or JAKi. Patients who discontinued treatment due to pregnancy, death or emigration from Sweden were excluded from the analysis. In secondary analyses we compared drug retention, defined as the proportion of patients remaining on treatment over time, and the change at three months compared to baseline in DAS28(ESR), HAQ-DI and CDAI.

Differences between treatment responses (proportions at one year and changes from baseline at three months) were estimated using generalized linear models with robust standard error estimation. Crude drug retention was estimated and plotted using the Kaplan-Meier estimator and adjusted comparisons were estimated using Cox regression. Multiple imputation was used to account for missing baseline covariates and treatment responses. The set of adjustment covariates consisted of variables assumed to influence treatment response and to be associated with treatment selection. These were measured at baseline and included: demographic characteristics, RA parameters (duration, severity), line of therapy, indicators of previous use of csDMARDs, bDMARDs or JAKi, co-medication with csDMARD, NSAIDs and glucocorticoids, disease history and general health indicators (smoking status and number of days spend in hospital).

### 4.3.3 Study III – Emulation of the SWEFOT trial in observational data

Study III was a cohort study designed to emulate the protocol of SWEFOT. SWEFOT was an open-label RTC, nested in SRQ, which compared addition of infliximab over MTX with addition of SSZ and HCQ over MTX, among early RA patients unsuccessfully treated with MTX for 3 months (150).

To emulate SWEFOT in observational data we first identified patients with a primary diagnosis of RA and with RA debut after July 2005 in SRQ. The start of the recruitment period was chosen to avoid population overlap with SWEFOT and to have PDR data after RA debut for all patients.

We then identified patients who initiated infliximab in SRQ and patients who initiated SSZ or HCQ in PDR. The group of patients who initiated SSZ + HCQ was further narrowed down to patients who initiated both drugs, not necessarily simultaneously, but ensuring that no more than 180 days elapsed between the first dispensation of the first drug and the first dispensation of the second drug, and that there was at least one dispensation of the first drug after the first dispensation for the second drug (i.e. the first drug was not stopped before initiating the second). The baseline date was then identified for each patient as: (1) the first decision to initiate a treatment with infliximab, as identified in SRQ, for patients in the infliximab cohort and (2) the date of the first prescription dispensation for the second drug in the combination in the SSZ + HCQ cohort.

After recruitment, all SWEFOT patients were treated with MTX for three months as part of the trial (150). Therefore, one of the eligibility criteria for our emulation was that patients have been treated with MTX before baseline. However, we extended the time-window of MTX initiation before baseline between 30 and 540 days. Also, we did not require ongoing MTX treatment during the study but we estimated proportions of patients on MTX around baseline, and we used this measurement as adjustment (> 90% of patients were on MTX to baseline). Patients were enrolled in SWEFOT at maximum one year after RA debut. We also relaxed this criterium allowing maximum 540 days between RA debut and the first MTX dispensation. These compromises were made to enlarge the SSZ + HCQ cohort, aiming for at least 200 patients/cohort in our emulation. Next, patients who used any other DMARDs than MTX (and SSZ or HCQ in the combination cohort), between RA debut and baseline, were excluded, since no DMARD use was allowed in SWEFOT before study entry (MTX was administered as part of the trial before randomization to the study treatments). Pre-baseline DMARD treatments were identified in SRQ and PDR. The last inclusion criterium applied was having a DAS28(ESR) larger than 3.2 at baseline. In SWEFOT, only patients with DAS28(ESR) larger than 3.2 after initial treatment with MTX were randomly assigned to additional infliximab or SSZ + HCQ. Additionally, we also excluded patients who, due to late immigration or emigration did not have a complete five-year presence in Swedish registers before baseline, or patients whose baseline date was less than 9 months before the end of data availability, on the first of January 2021. To be included in SWEFOT patients had to be on either no glucocorticoids or a stable dose of maximum 10 mg prednisone equivalents per day. In our

emulation, no restrictions to the use of glucocorticoids before baseline was applied but an average daily dose was estimated using dispensed prescriptions.

The date of treatment discontinuation was identified in the SRQ for infliximab treatments and was estimated as the first gap in the sequence of prescriptions using the dispensation data (PDR) for SSZ and HCQ treatments. To identify gaps in prescription sequences, we first estimated the duration of prescriptions using a daily dose of 2000 mg/day for sulfasalazine and of 200mg/day for hydroxychloroquine. A gap was defined as not collecting a new prescription within 90 days after the estimated end of the previous prescription. As in SWEFOT, treatment discontinuation in the combination cohort was considered when both SSZ and HCQ had been discontinued. In the main analysis, switching from infliximab to etanercept and from SSZ + HCQ to ciclosporin A was allowed as part of the protocol treatment, similarly to SWEFOT. Initiating any other DMARD was considered a protocol treatment discontinuation.

The primary emulated outcome was the proportion of patients achieving a EULAR good (vs moderate or no) response at nine months after baseline, defined as: DAS28(ESR) $\leq$ 3.2 at evaluation and a decrease in DAS28(ESR) larger than 1.2 units at evaluation compared to baseline (150,258). All patients who initiated treatments at baseline, except those who died or emigrated during follow-up, were analyzed at the nine-months end-point. Those who discontinued the protocol treatment were classified as "non-responders".

Missing treatment response as well as baseline covariate values were imputed using multiple imputation (see Section 4.2.6). The average treatment effect (in the entire study population) was estimated as the ratio between the proportion of treatment responders in the infliximab cohort versus the proportion of treatment responders in the SSZ + HCQ cohort. Because treatment is not allocated at random in clinical practice, IPTW was used to adjust for baseline confounding, assuming that conditional on several measured covariates, treatment allocation was independent of the potential outcomes (see Sections 4.2.1, 4.2.4 and 4.2.5). The selected covariates were considered treatment response predictors. The list included: demographic data, year or treatment start and RA duration at treatment start, rheumatoid factor positivity, several disease activity/disability measurements, baseline use of MTX, NSAIDs and glucocorticoids, history of several severe, chronic conditions, and general health indicators such as days spent in hospital and smoking status. The protocol was published before analyzing any association between treatment and outcome (ClinicalTrials.gov – NCT05051137).

### 4.3.4 Study IV – Glucocorticoids and the risk of serious infections in RA

Study IV was a cohort study designed to compare the risk of serious infections between exposure to different patterns of time-varying oral glucocorticoid doses over the course of three years, in early RA patients.

In this study, we included patients with an initial visit recorded in SRQ no later than one year after RA debut and containing disease activity (DAS28(CRP)) and disability (HAQ-DI) information. Starting at this initial SRQ visit, we divided the time of each participant into 90-day episodes, each containing a time indicator. We identified the quantity of oral

glucocorticoids dispensed within each episode using PDR data, and we calculated average daily doses in prednisone equivalents. The daily doses were categorized into: “no glucocorticoids”, “low glucocorticoid doses” (≤ 10 mg/day) and “high glucocorticoid dose” (> 10 mg/day). We selected variables for confounding and selection bias adjustment from among infection risk factors (259–262), and we measured them in time-windows of various lengths before each episode. The baseline fixed and time-varying selected co-variates were: demographic and socio-economic characteristics at baseline, the average daily dose of glucocorticoids used within one year before the first SRQ visit, current DAS28(CRP) and HAQ before each episode, current co-medication, a time-varying history of several comorbid conditions and the number of days spent in hospital before within one year before each episode.

The outcome of interest was serious infections, defined as hospitalization records with a primary diagnosis or infection, according to a pre-specified list of ICD-10 codes. Outcome events were identified in each follow-up episode.

Follow-up started 90 days after the initial SRQ visit, immediately after the window within which baseline exposure was calculated, and continued for a maximum of 3 years (i.e. 1080 days or twelve 90-day episodes) to the first outcome event, death or emigration. Thus, the follow-up time of each participant consisted of a series of 90-day episodes, the exposure value for each such follow-up episode being estimated in the immediately preceding episode, and other covariates being measured before exposure. This design was meant to respect the time order in the causal chain “covariates cause exposure which in turn causes the outcome”, to allow time-updated measurements of exposure and confounding, and the estimation of time-varying inverse probability weights.

To ensure that infections identified within the study period were incident events, patients who experienced infections within 180 days before baseline were excluded. To ensure that all patients had sufficient data before baseline and at least three years of follow-up data (unless censored), the study was restricted to patients with a baseline date within the period January 2007 to February 2018.

We analyzed the data using pooled logistic regression, as previously suggested for MSM.(229) The log-odds of serious infections was modelled over all observations as a function of exposure history and time indicator, represented as a categorical variable, to allow for a time-varying baseline hazard. For rare outcome events (as serious infections are), the odds-ratios estimated from the pooled-logistic regression model closely approximate hazard rations obtained from the corresponding Cox mode. We modelled the exposure history in several ways. Initial models included only the current glucocorticoids dose variable, or current dose plus past doses summarized as counts of periods with low and with high doses. A more flexible model included separate variables for each of the 12 periods of exposure history. In order to be able to estimate separate coefficients for each of the 12 exposure variables over all observations, these had to be complete (i.e. no missing values at any of 12 past periods, at any time). To this end, glucocorticoid exposure before the study was set to zero, adjusting for the actual glucocorticoids use before the study in a separate variable (as illustrated in the supplement of

Study IV). We adjusted for confounding using conventional models and time-varying IPTW to compare the results. Conventional adjustments were done by including adjustment co-variates directly in the outcome model, together with exposure and time. A crude model was followed by adjustment for baseline covariate values only and adjustment for time-updated covariates. For estimating the MSM, time-varying IPTW and IPCW were estimated for each observation (i.e. discrete time) as described in Section 4.2.5 and multiplied to obtained the IPWs that adjust for both time-varying confounding and selection bias via informative censoring. The final IPWs were truncated to the 1st and 99th percentiles of their observed distribution in each period. The IPW adjusted flexible model was used to estimated cumulative incidences of serious infections for different patterns of glucocorticoid dosing over the three-years follow-up. These were calculated as described in Section 4.2.7, and confidence intervals were calculated via non-parametric bootstrapping. As discussed in Section 4.2.6, missing DAS28 and HAQ-DI data were handled using the LOCF approach in the main analysis and the missingness patterns approach in a sensitivity analysis.

## 4.4 ETHICAL CONSIDERATIONS

The observational studies included in this thesis have been conducted using secondary data recorded routinely in several Swedish national registers. These include data about health or socio-economic status that can be traced to living individuals and which, under EU personal data protection legislation (GDPR), are classified as sensitive personal data (263). According to GDPR, sensitive personal data should not be processed without a clear legal basis. The foremost such legal basis is the explicit consent of the subject. However, in Article 9, point 2 of the GDPR, several exemptions to informed consent are laid out, one of them being when "processing is necessary for … scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject." (263,264).

In Sweden, informed consent is generally not required for observational studies using large-scale secondary (register) data. There are several reasons for this. First of all, it would be very expensive to request informed consent from tens of thousands up to millions of people, as included in most population-based research. Secondly, even if requested, the probability of obtaining consent may not be equally distributed over the population (e.g. those with a poor command of the Swedish language may be less likely to consent). Thirdly, even if it is collected prospectively the use of secondary data is retrospective (it may be used years after its collection and for a different purpose than it was collected for). Hence, at the moment of collection, the specific research questions that the data may be used for answering would be unknown, and at the moment of conducting the research it may no longer be possible to contact individuals who left Sweden, are unconscious or who have died (265). As such, requesting consent for individual research projects would cancel two of the main advantages of observational studies over RCTs: lower costs and less restrictive inclusion. Instead of requesting individual consent,

participants are assumed to agree that the state governs the use of register-data for the common good (e.g. increasing the quality and availability of health-care).

The main condition for lawfully conducting research on sensitive personal data with or without individual consent is to ensure safe storage and access to data in order to prevent potential privacy breaches. Therefore, researchers must describe in their ethical applications how the data would be collected, transferred, stored and accessed, for how long it would be stored, who will have access to the data and, if the data is pseudo-anonymized, who would have access to the identification key (265). Pseudo-anonymizing data means replacing direct identifiers (i.e. information which specifically identifies a person) such as name or personal identification number, with a code, the decoding key being hidden away. In Sweden they key is usually kept by the governmental administrator of the register. The key can be used to re-identify individuals in the data and connect additional data to them. Once they key has been destroyed the data may be considered de-identified or anonymized, even though with detailed enough information, individuals with unique combinations of characteristics may still be identifiable (266).

Besides collecting and analyzing personal data, the main reason why individual informed consent is mandatory for RCTs is that participants are exposed to new interventions with largely unknown benefits and risks. This is however not the case for register based observational studies which analyze data from patients treated in routine clinical practice with already approved treatments (267).

However, even if individual informed consent is not necessary, in order to process sensitive personal data, research has to receive approval from an ethical review board according to the Swedish Ethical Review Act (268). Since January 2019, ethical review in Sweden is done centrally by the Swedish Ethical Review Authority (Etikprövningsmyndigheten) (269). Ethical review boards are tasked with assessing whether research is ethically justified. Health research is ethically justified if it has scientific (i.e. internal validity) and social (or clinical) value, that is if it generates knowledge that can inform decision-making in order to improve the individual and/or public health, using limited resources, while ensuring minimal harm to research subjects (270). All studies that are part of this thesis have used pseudo-anonymized secondary data from several linked national Swedish registers and are covered by ethics permit no. 2015/1844-31/2 (amended 2016/1986-32, 2017/2473-32 and 2020-01756) issued by the Regional Ethics Board of Stockholm/Swedish Ethical Review Authority. As argued above, no informed consent was requested from individual participants.

Finally, to ultimately reach its goal of informing the public, research results have to be made publicly available by presentations at various meetings and by publishing results in scientific and popular science articles. Most results generated by our studies have already been presented at scientific/clinical conferences and have been published in open access scientific journals, thus are accessible to the public for no additional cost.

# 5 RESULTS

## 5.1 STUDY I – BIOLOGICAL DMARDS AND THE RISK OF GASTRO-INTESTINAL PERFORATIONS IN RA

Study I (271) included 17594 TNFi, 2527 abatacept, 3522 rituximab and 2377 tofacitinib initiations. Also, 62532 bionaïve RA patients and 76304 general population controls were included as additional reference groups.

Table 5.1.1 presents the comparison of lower GI perforation incidence rates between abatacept, rituximab, tocilizumab and TNFi initiators. We observed higher crude incidence rates of lower GI perforation among abatacept (2.6 events /1000 person-years), rituximab (2.1 events /1000 person-years) and tocilizumab (4.1 events /1000 person-years) initiators compared to TNFi initiators (1.6 events /1000 person-years). After IPTW adjustment for several potential confounders (see Section 4.3.1), the incidence rates for rituximab and abatacept approached that of TNFi, while the incidence rate for tocilizumab remained significantly higher.

**Table 5.1.1** – Lower GI perforations, crude and IPTW-adjusted incidence rates and contrasts between non-TNFi and TNFi bDMARDs

| Cohort | Crude IR (95% CI) | Crude HR (95% CI) | IPTW adj. IR (95% CI) | IPTW adj. HR (95% CI) |
|---|---|---|---|---|
| TNFi | 1.57 (1.21–2.05) | Ref | 1.85 (1.34–2.36) | Ref |
| Abatacept | 2.62 (1.52–4.52) | 1.68 (0.93–3.03) | 1.98 (0.73–3.23) | 1.07 (0.55–2.10) |
| Rituximab | 2.11 (1.39–3.21) | 1.36 (0.82–2.24) | 1.65 (0.84–2.46) | 0.89 (0.50–1.58) |
| Tocilizumab | 4.10 (2.70–6.22) | 2.61 (1.61–4.24) | 4.07 (2.14–6.00) | 2.20 (1.28–3.79) |

CI = confidence interval, HR = hazard ratio, IR = incidence rate; incidence rates are expressed as events /1000 person-years.

Inverse probability of treatment weighted (IPTW) adjustment for: demographic characteristics, year of treatment start, disease history, RA disease activity and disability, comedication with MTX, other conventional DMARDs, selective COX2 inhibitors, NSAIDs, glucocorticoids.

Figure 5.1.1 displays the non-parametric IPTW adjusted survival curves. It can be seen that most events (steps of the curves) were experienced within the first two years of follow-up in all treatment cohorts. The curve corresponding to tocilizumab had the steepest descent, separating from the other curves at the beginning of follow-up. It is interesting to note that in the abatacept cohort no one experienced lower GI perforations after the second year of follow-up.

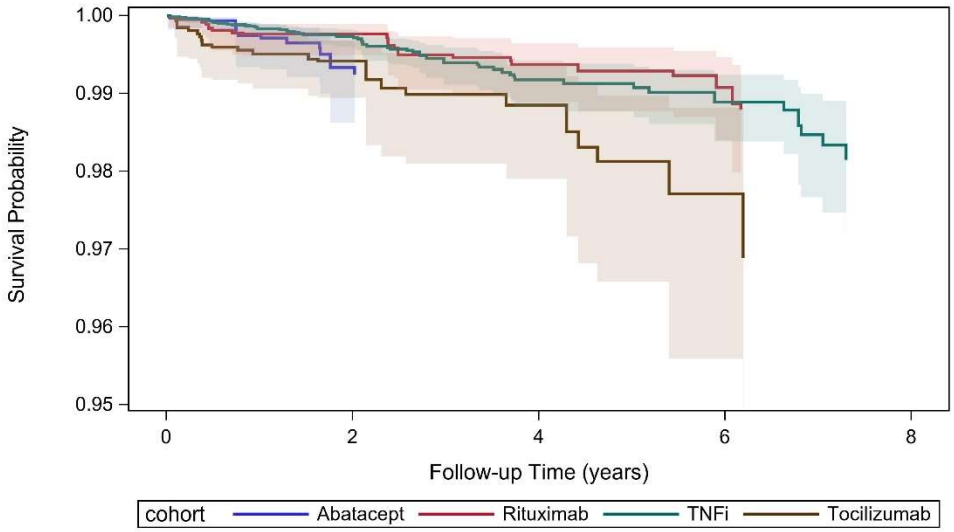


**Figure 5.1.1** – IPTW-adjusted lower GI perforation survival curves. Shaded areas represent 95% confidence limits

After standardization for sex and age, lower Gi perforation incidence rates were reduced among the bionaïve and general population controls compared to bDMARD treated RA patients, but except for tocilizumab, the differences were diminished when also correcting for differenced in glucocorticoids use before baseline (Figure 5.1.2).

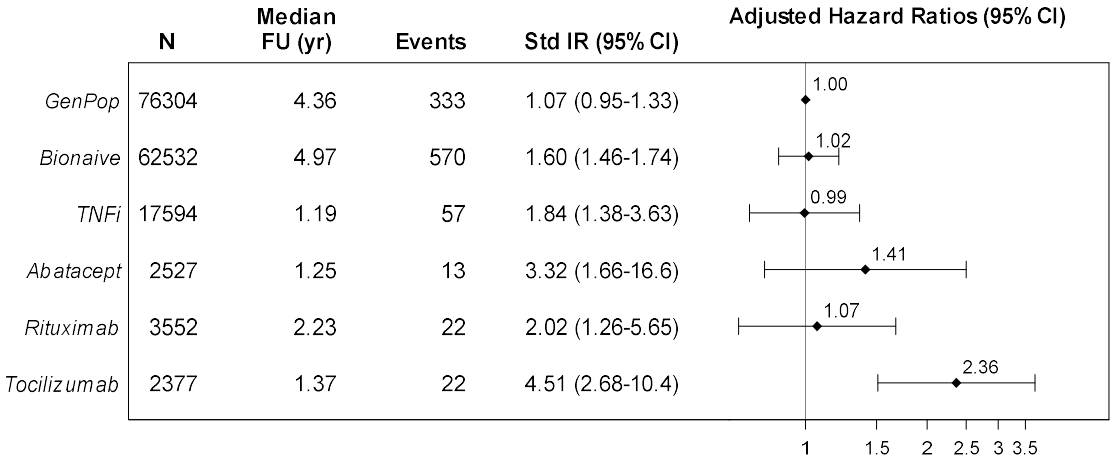


**Figure 5.1.2** – Sex and age standardized lower GI perforation incidence rates and sex, age and glucocorticoid use adjusted comparisons between bDMARD treated and bionaïve RA patients versus general population controls

## 5.2 STUDY II – COMPARATIVE EFFECTIVENESS OF BARICITINIB, TOFACITINIB AND BIOLOGICAL DMARDS IN RA

Study II (272) included 8006 RA patients who contributed with 10772 treatment episodes. Out of the two JAKis, baricitinib was the one mainly used in Sweden (1420 baricitinib treatments initiated versus 316 tofacitinib treatments). Also, tofacitinib was reserved for later treatment (median fifth line compared to median third line for baricitinib).

In the primary analysis of study II, we compared binary treatment responses around one-year after drug initiation measured as: EULAR DAS28(ESR) good (versus no or moderate) response; a more than 0.2 units reduction in HAQ-DI (compared to baseline); and CDAI remission (≤ 2.8 at evaluation). The results of this analysis are presented in Figure 5.2.1.

**Good EULAR response**

| Cohort | N | Crude Proportion | Crude Contrast | Adj. Contrast |
|---|---|---|---|---|
| Abatacept | 1040 | 25.0 (20.7 to 29.3) | -0.2 (-6.1 to 5.7) | -1.3 (-7.0 to 4.5) |
| IL6 inhib. | 838 | 33.9 (28.5 to 39.4) | 8.8 (3.0 to 14.5) | 8.7 (2.9 to 14.4) |
| Rituximab | 1089 | 32.3 (28.1 to 36.4) | 7.1 (1.5 to 12.7) | 3.9 (-1.8 to 9.6) |
| TNF ihib. | 5981 | 29.6 (27.3 to 31.8) | 4.4 (0.4 to 8.4) | -4.3 (-8.7 to 0.1) |
| Tofacitinib | 313 | 21.4 (13.1 to 29.6) | -3.8 (-13.1 to 5.4) | -1.1 (-10.6 to 8.5) |
| Baricitinib | 1403 | 25.2 (21.3 to 29.1) | Ref | Ref |

**HAQ-DI improvement**

| Cohort | N | Crude Proportion | Crude Contrast | Adj. Contrast |
|---|---|---|---|---|
| Abatacept | 1040 | 30.4 (26.6 to 34.1) | -6.4 (-11.3 to -1.5) | -8.6 (-13.5 to -3.8) |
| IL6 inhib. | 838 | 28.0 (23.7 to 32.4) | -8.7 (-14.2 to -3.2) | -12.1 (-17.5 to -6.6) |
| Rituximab | 1089 | 38.6 (34.0 to 43.1) | 1.8 (-4.1 to 7.8) | -2.4 (-8.3 to 3.4) |
| TNF ihib. | 5981 | 32.9 (30.7 to 35.0) | -3.9 (-8.1 to 0.3) | -9.9 (-14.4 to -5.4) |
| Tofacitinib | 313 | 28.2 (18.6 to 37.8) | -8.5 (-19.1 to 2.0) | -7.9 (-18.3 to 2.5) |
| Baricitinib | 1403 | 6.76 (33.36 to 40.17 | Ref | Ref |

**CDAI remission**

| Cohort | N | Crude Proportion | Crude Contrast | Adj. Contrast |
|---|---|---|---|---|
| Abatacept | 1040 | 11.7 (9.0 to 14.5) | -3.2 (-7.4 to 0.9) | -3.8 (-8.0 to 0.5) |
| IL6 inhib. | 838 | 12.6 (9.0 to 16.2) | -2.4 (-6.8 to 2.1) | -2.1 (-6.7 to 2.5) |
| Rituximab | 1089 | 11.1 (7.7 to 14.5) | -3.8 (-8.9 to 1.2) | -5.5 (-10.9 to -0.1) |
| TNF ihib. | 5981 | 16.9 (15.3 to 18.5) | 2.0 (-1.42 to 5.4) | -6.0 (-9.8 to -2.2) |
| Tofacitinib | 313 | 11.2 (5.0 to 17.5) | -3.8 (-11.0 to 3.5) | -0.9 (-8.0 to 6.2) |
| Baricitinib | 1403 | 15.0 (11.7 to 18.3) | Ref | Ref |

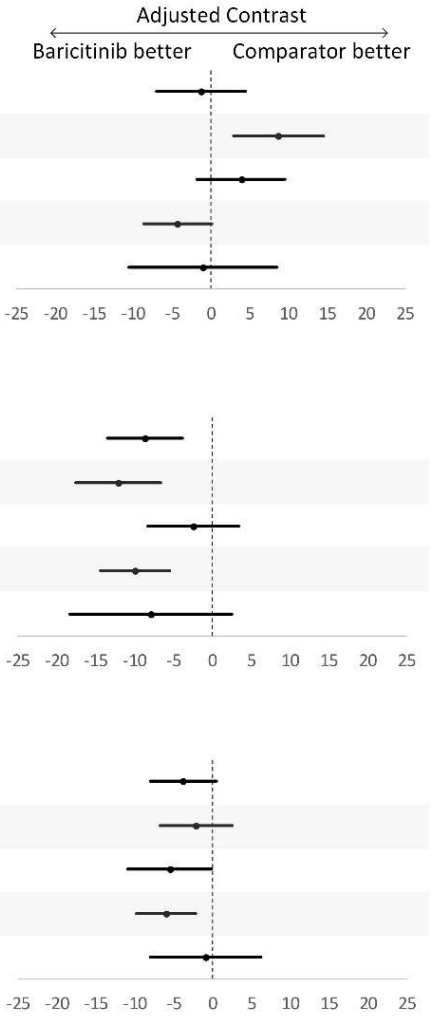


**Figure 5.2.1** – Differences between proportions of good EULAR responders, HAQ-DI improvements and CDAI remissions at one year, comparing tofacitinib and bDMARDs to baricitinib

Barbulescu, Andrei, "Effectiveness of baricitinib and tofacitinib compared with bDMARDs in RA: results from a cohort study using nationwide Swedish register data.", Rheumatology, 2022, Online ahead of print, by permission of Oxford University Press.

Approximately 1% of patients in each treatment cohort died, emigrated or had to stop treatment due to pregnancy within the first year after treatment initiation and these patients were excluded from the analysis. All other patients were kept in the analysis, classifying those who discontinued treatment before one year as "non-responders". The proportions of missing responses for patients who continued treatment ranged from 39% for CDAI remission to 60% for EULAR response and were imputed using multiple imputation.

Crude EULAR good response proportions ranged from 21% on tofacitinib to 34% on IL-6-receptor inhibitors, with baricitinib in-between at 25%. After confounding adjustment, there was no difference left in EULAR good response proportion between tofacitinib and baricitinib and only IL-6-receptor inhibitors retained a significantly higher response proportion compared to the two JAKis. The EULAR good response proportion for TNFi was significantly lower than for baricitinib.

The crude proportions of patients achieving HAQ-DI improvement larger than 0.2 units compared to baseline ranged between 28% on tofacitinib and IL-6-receptor inhibitors and 39% on rituximab and it was 37% for baricitinib. After adjustment for confounding, baricitinib showed the highest proportion of HAQ-DI improvements of all b/tsDMARDs, with statistically significant differences from abatacept, IL-6-receptor inhibitors and TNFi. The proportion of HAQ-DI improvements remained 7.9 percentage points higher for baricitinib compared to tofacitinib after confounding adjustment, but the confidence intervals were wide, covering a null difference, thus there was not enough evidence to support an inferior response on tofacitinib.

The crude proportions of patients achieving CDAI remission ranged from 11% on rituximab and tofacitinib to 17% on TNFi, with baricitinib at 15%. After adjustment for confounding, CDAI remission proportions were equal for baricitinib and tofacitinib and were higher for the two JAKis compared to all bDMARDs, with a statistically significant difference between TNFi and rituximab versus baricitinib.

The results at one year were confirmed by results at three months. Also, the retention of baricitinib was higher than that of all alternatives except rituximab. Considering co-medication with csDMARDs, in our study population JAKi were used more frequently as monotherapy compared to bDMARDs and tofacitinib was used more frequently as monotherapy compared to baricitinib. However, one-year response proportions were lower for tofacitinib in monotherapy compared to tofacitinib combined with csDMARDs, while for baricitinib the proportion of responders were similar in monotherapy or csDMARD combination.

To summarize, our results indicate at least equivalent effectiveness between tofacitinib, baricitinib and bDMARDs with consistently higher treatment responses on baricitinib compared to TNFi.

## 5.3 STUDY III – EMULATION OF THE SWEFOT TRIAL IN OBSERVATIONAL DATA

After applying eligibility criteria, 509 RA patients were included in study III (273) out of 57288 patients with a primary diagnosis of RA identified in SRQ. Of these, 313 initiated treatment with infliximab and 196 initiated treatment with SSZ + HCQ. One patient in the SSZ + HCQ cohort died during follow-up and was excluded from the analysis.

After imputing missing outcome data for 28% of participants in the infliximab cohort, we observed 39% (95% confidence interval 33% to 46%) EULAR good responders. In the SSZ + HCQ cohort, 25% of outcome values were imputed, and we observed 28% (95% confidence interval 22% to 37%) EULAR good responders. This corresponds to a ratio of 1.39 (95% confidence interval 1.04 to 1.86), which rose to 1.48 (95% confidence interval 0.98 to 2.24) after IPTW confounding adjustment, compared to 1.59 (95% confidence interval 1.10 to 2.30) observed in SWEFOT.

There were several differences in population composition between our observational emulation and the target trial. The SWEFOT population was slightly younger and contained a higher proportion of women and of patients of Swedish origin. Patients in the observational emulation had longer RA, but regardless of relaxing this inclusion criterium in the emulation the difference was small (median of 1 year in the observational emulation versus a median of 0.8 years in SWEFOT). RA disease activity was marginally higher in the observational emulation with larger differences between treatment cohorts. Standardizing the composition of some measured covariates in the observational population to that in the target trial increased the response ratio increased to 1.50 (95% confidence interval 0.84 to 2.68) (unpublished sensitivity analysis).

Not allowing treatment switches to etanercept or ciclosporin A in a sensitivity analysis produce similar results to the main analysis. Relaxing the main eligibility criteria one by one, gradually reduced the response ratio all the way to 1.19 (95% confidence interval 0.86 to 1.66) when allowing for use of non-MTX DMARDs before baseline, despite adjusting for the baseline DAS28(ESR), RA duration and previous DMARD use.

## 5.4 STUDY IV – GLUCOCORTICOIDS AND THE RISK OF SERIOUS INFECTIONS IN RA

We identified 9639 RA patients who fulfilled the inclusion criteria for study IV, and they contributed with 110150 90-day follow-up episodes.

Table 5.4.1 shows associations between the history of exposure to glucocorticoids and the current risk of serious infections, using different exposure history summaries (models 1,2, 3) and confounding adjustments (crude, adjustment 1, adjustment 2 and IPW/MSM adjustment). In Model 1, exposure history was modelled as the current dose of glucocorticoids only, comparing high and low doses with no glucocorticoids. Current glucocorticoids exposure, in

both low and high doses, was associated with a significantly higher incidence of serious infections compared to no glucocorticoids and a dose response relationship is discernable. In Model 2, terms were added to represent the past exposure history, since the start of follow-up, alongside current exposure. These were linear terms for sums of previous periods with low glucocorticoid doses and high glucocorticoid dose respectively. Adding past exposure to the regression models reduced associations between current exposure and current risk of serious infections. In Model 3, past exposure was divided into recent past (the year before current exposure) and distant past. While exposure in the recent past year retained an association with the current risk of serious infections, exposure more than one year ago was no longer associated.

**Table 5.4.1** – Different models of the association between glucocorticoids exposure history and the risk of serious infections

| | **Crude HR** | **Adj.1[a] HR** | **Adj.2[b] HR** | **MSM[c] HR** |
|---|---|---|---|---|
| | | (95% Confidence interval) | | |
| ***Model 1*** *– Current period average daily dose of glucocorticoids (3 categories, ref= no use)* | | | | |
| Current per. Low | 1.67 (1.41-1.98) | 1.34 (1.12-1.60) | 1.31 (1.10-1.56) | 1.36 (1.13-1.64) |
| Current per. High | 2.71 (2.19-3.34) | 2.09 (1.66-2.63) | 1.80 (1.43-2.26) | 1.98 (1.54-2.55) |
| ***Model 2*** *– Current period dose (3 cat.) + Linear terms for counts of past periods with low and high dose respectively* | | | | |
| Current per. Low | 1.26 (1.05-1.52) | 1.11 (0.91-1.34) | 1.10 (0.91-1.34) | 1.12 (0.90-1.39) |
| Past Low | 1.12 (1.07-1.17) | 1.09 (1.04-1.14) | 1.09 (1.04-1.14) | 1.09 (1.03-1.15) |
| Current per. High | 1.96 (1.56-2.47) | 1.69 (1.32-2.15) | 1.53 (1.20-1.96) | 1.61 (1.21-2.13) |
| Past High | 1.19 (1.12-1.26) | 1.14 (1.07-1.22) | 1.09 (1.03-1.17) | 1.12 (1.04-1.20) |
| ***Model 3*** *– Current period dose (3 cat.) + Linear terms for the count of periods with low and high doses respectively within each of two past years of follow-up* | | | | |
| Current per. Low | 1.21 (0.99-1.46) | 1.07 (0.88-1.31) | 1.07 (0.88-1.31) | 1.06 (0.85-1.32) |
| Recent yr. Low | 1.19 (1.10-1.30) | 1.14 (1.05-1.24) | 1.14 (1.05-1.25) | 1.16 (1.06-1.27) |
| Past yr. Low | 1.04 (0.94-1.16) | 1.03 (0.92-1.14) | 1.02 (0.92-1.13) | 1.02 (0.91-1.14) |
| Current per. High | 1.80 (1.42-2.27) | 1.57 (1.23-2.01) | 1.45 (1.13-1.86) | 1.47 (1.11-1.97) |
| Recent yr. High | 1.39 (1.24-1.55) | 1.31 (1.16-1.48) | 1.21 (1.08-1.36) | 1.28 (1.12-1.46) |
| Past yr. High | 1.03 (0.90-1.19) | 1.01 (0.87-1.17) | 1.00 (0.86-1.15) | 1.00 (0.84-1.19) |

[a] Conditional adjustment for baseline confounder values (baseline demographics, use of glucocorticoids before baseline, comedication with: TNFi, nonTNFi, csDMARDs, rheumatoid factor positivity, DAS28-CRP and HAQ-DI, disease history, days spend in hospital within the previous year)

[b] Conditional adjustment for the same characteristics measured at baseline, but updating measurements before the exposure assessment window of each follow-up period

[c] Inverse probability weighting (IPW) adjustment for the same time-updated covariates as at (b) above.
HR = hazard ratio

In all exposure models, adjustment for baseline confounder values (Adj. 1) reduced contrasts substantially. Further adjustment for time-updated confounder values (Adj. 2 and MSM) did not have such a strong influence. Comparing adjustment for time-updated confounders in a conventional regression model (Adj. 2) versus in an MSM, the conventional adjustment reduced the contrasts more, but the differences were marginal.

Next, we estimated the cumulative incidence of serious infections, over a follow-up period of three years, under different patterns of glucocorticoids exposure. Since in clinical practice exposure to glucocorticoids is commonly limited to the first year, we started by comparing different doses used during the first year and different treatment durations within the first year.

In Figure 5.4.1, there is a clear dose response relationship, with IPW adjusted cumulative serious infections incidences at three years increasing from 5.3 (95% confidence interval 4.6 to 6.2) infections per 100 persons under no glucocorticoids, to 7.1% (95% confidence interval 6.2 to 8.1) infections per 100 persons under low glucocorticoid doses within the first year, and 9.4% (95% confidence interval 7.9 to 11.1) infections per 100 persons under high glucocorticoid doses within the first year.

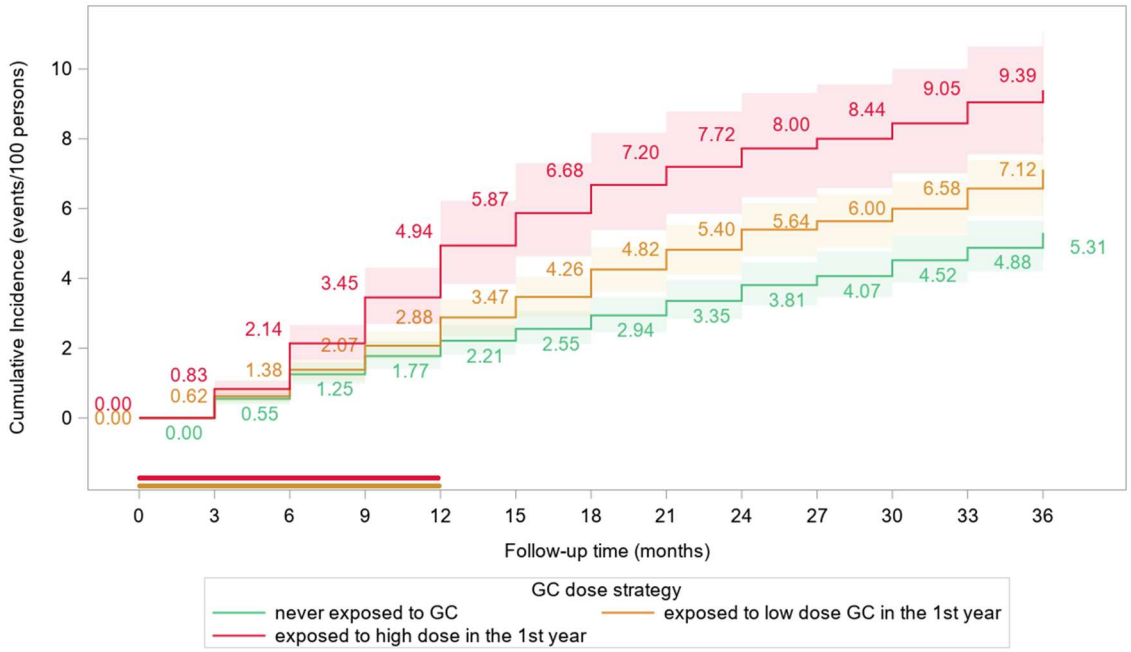


**Figure 5.4.1** – Adjusted cumulative incidence of serious infections under no glucocorticoids, low glucocorticoids doses and high glucocorticoids doses within the first year, followed by no glucocorticoids in the remaining 2 years of follow-up

When comparing exposure to low glucocorticoid doses during the first 3, 6 and 9 months (Figure 5.4.2), the cumulative incidence of serious infection increased with the length of exposure, but the three cumulative incidences at the end of follow-up, of 6.0% (95% confidence interval 5.2 to 6.7) infections per 100 persons, 6.5% (95% confidence interval 5.7 to 7.3) infections per 100 persons, and 6.8% (95% confidence interval 6.0 to 7.8) infections per 100 persons respectively, were not significantly different.

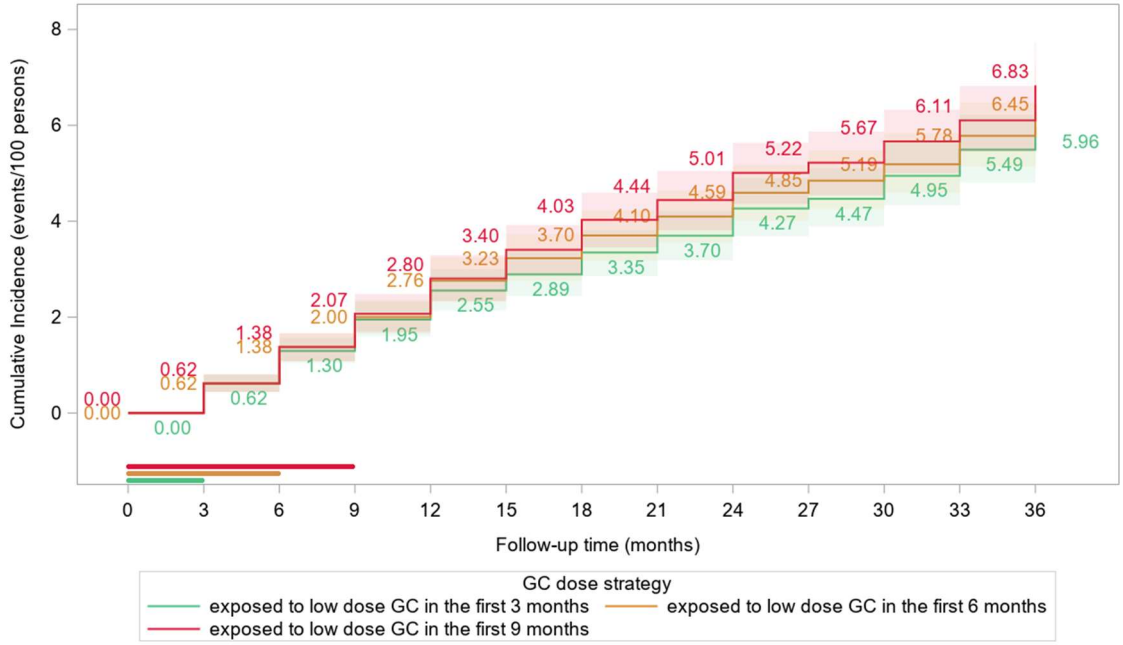


**Figure 5.4.2** – Adjusted cumulative incidence of serious infection under low dose glucocorticoids in the first 3-, 6- or 9-months, followed by no glucocorticoids until 3 years

Lastly, we compared we compared exposure to the same glucocorticoid dose level (low doses), for the same duration, but in different periods during follow-up (Figure 5.4.3). Exposure during the first year of follow-up yielded a cumulative incidence of 7.1% (95% confidence interval 6.2 to 8.1) infections per 100 persons at end of the third year. Exposure during the second year produced a slightly lower incidence of 6.8% (95% confidence interval 5.8 to 8.2) infections per 100 persons. Finally, the cumulative incidence was lowest under exposure during the third year 5.9% (95% confidence interval 5.0 to 6.9) infections per 100 persons. The explanation is as follows. Under all exposure patterns, patients had two years of "no exposure". However, since exposure continues to have an effect (nonetheless diminishing) after having been stopped, patients accumulate more events in period of "no exposure" after "exposed" periods than in periods of "no exposure" before "exposed" periods. Therefore, patients who were exposed during the first year have two years of "no exposure" after the "exposed" period and accumulate more events compared to patients exposed during the third year who had two years of "no exposure" before the "exposed" period.

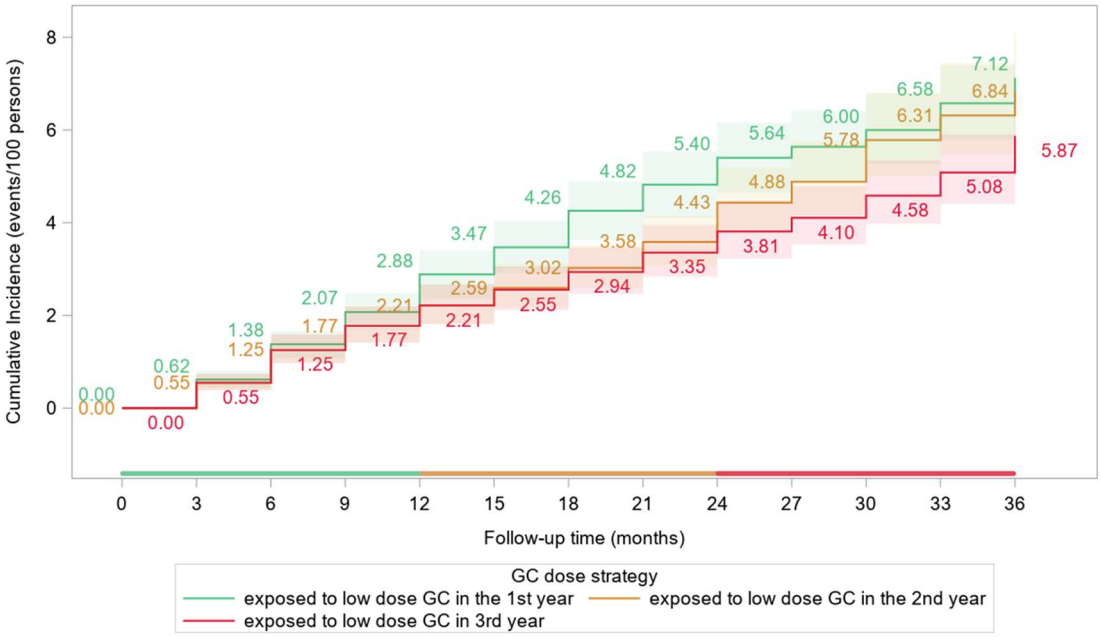


**Figure 5.4.3** – Adjusted cumulative incidence of serious infection under low dose glucocorticoids within the first, second or third year and no glucocorticoids in the remaining 2 years of follow-up

# 6 DISCUSSION

## 6.1 STUDY I – BIOLOGICAL DMARDS AND THE RISK OF GASTRO-INTESTINAL PERFORATIONS IN RA

In study I we compared the incidence of lower GI perforations between RA patients treated with abatacept, rituximab or tocilizumab and RA patients treated with TNFi, and found a higher incidence among patients treated with tocilizumab, thus confirming previous findings.

However, previous studies generally reported lower incidence rates compared to our study, and there was heterogeneity in rate ratios as well. This may be due to chance alone, considering the rare outcome, but differences between studies in outcome reporting and validation may also explain it.

Comparing lower GI perforation incidence rates, Strangfeld et al. reported an incidence of 2.7 lower GI perforations /1000 person-years (111), higher than 4.1 lower GI perforations /1000 person-years in our study. Two American studies reported even lower incidence rates of 1.5 (112) and 1.3 (113) lower GI perforations /1000 person-years for tocilizumab initiators. This may be due higher specificity of outcome ascertainment in the previous studies. In the German study, the reported GI perforations were validated by physicians blinded to the patient's treatment (111). In the two American studies, a validated high specificity outcome definition was used (114). Even though several validation studies have shown high positive predictive values for many conditions identified in the NPR (186), including diverticular disease (274), we have not conducted any specific validation for our GI perforation definition, thus our study may have included false positive GI perforations.

Comparing hazard rations, Strangfeld et al. reported a hazard ratio of 4.5 for lower GI perforations, for tocilizumab vs csDMARDs, and a hazard ratio of 1.0 when comparing TNFi to csDMARDs (111). We found a lower hazard ratio of 2.2 for tocilizumab vs TNFi. The higher hazard ratio in the study by Strangfeld et al. could be due to differential outcome ascertainment between treatment groups. In the German RABBIT registry, patients were prospectively followed by their treating rheumatologist who reported adverse events (including GI perforations) at regular intervals (111). Differential outcome ascertainment between treatment groups is possible if screening for GI complications was more thorough among tocilizumab treated patients, assumed to be at increased risk. In our study, outcome events were extracted from the NPR where data on GI perforations were routinely collected as patients visited hospitals or outpatient clinics, thus independently of exposure assignment by rheumatologists. It could be argued that decoupling treatment assignment from outcome reporting is closer to a blinded outcome assessment, even though the physicians recording and treating the outcome would have access to the patient's treatment history. On the other hand, it can also be argued that severe adverse events such as GI perforations, that may require emergency medical care, are unlikely overlooked regardless of treatment. Our hazard ratio was similar to the hazard ratio of 2.5 reported by Xie et al. (113) but lower than the incidence rate ratio of 4.0 reported by Monemi et al. (112), both studies using US insurance claims data.

A subsequent study, which employed data from three prospective French RA cohorts, confirmed an increased risk of GI perforation among tocilizumab versus rituximab or abatacept initiators (275). Similar to Strangfeld et al. (111), outcome events in the French study were reported by patients and treating physicians, entailing a potential outcome ascertainment bias. Also similar to the German study, a blinded outcome adjudication step was included, probably explaining the lower rates of GI perforation compared to our study (275). Nonetheless, the adjusted hazard ratio comparing the incidence of GI perforations between tocilizumab and the pooled cohort of rituximab and abatacept was 2.9, which is similar to our hazard ratio of 2.2 (considering that adjusted hazard ratios for abatacept and rituximab versus TNFi, which we used as reference, were close to 1 in our study).

Heterogeneity of outcome definitions is not restricted to observational studies. One review found RCTs assessing GI toxicity end-points difficult to compare and advocated the harmonization of definitions (276).

It is difficult to disentangle the etiological contribution of RA disease to GI complications from that of RA treatments, since virtually all RA patients undergo some therapy which could increase their risk of GI complications, and such treatments are not commonly used in the general population. However, comparing general population controls with RA patients, bionaïve or treated with different bDMARDs, and adjusting for sex, age and use of glucocorticoids before baseline, we did not observe a significantly higher incidence rate of GI perforations in other than the tocilizumab treated patients. This suggests that RA itself may not increase the risk independently of the treatment used. Two previous studies comparing the incidence of GI complications between RA patients and non-RA controls have reported an increased incidence among RA patients (85,86). The sex and age adjusted hazard ratio of 2.2 for lower GI perforation reported in the first study would correspond to our observations considering that the study analyzed data before the approval of tocilizumab (85). The other study covers a period after the approval of tocilizumab and reported a hazard ratio of 1.3 among patients without glucocorticoid prescriptions and 1.6 among those with glucocorticoid prescriptions (86). None of these studies divided the RA cohort according to the DMARD treatment received.

Finally, the biological mechanisms through which tocilizumab increases the risk of GI perforation remain incompletely understood, although some clues exist. Firstly, the expression of IL-6 is induced soon after GI injury as it is probably involved in wound healing by modulating angiogenesis and stimulating epithelial proliferation (277–279). Therefore, blocking IL-6 signaling in the early stage of acute injury could impair wound healing. Secondly, mice model experiments showed that IL-6 modulates colonic motility (peristalsis) (280,281). Increased intra-colonic pressure might favor diverticulitis (diverticular inflammation) and subsequent diverticular perforation by pushing the contents of the colon into diverticula (282). Finally, by reducing systemic inflammation it is possible that tocilizumab masks the early symptoms of diverticulitis, delaying diagnosis to the point of perforation. A small RCT found that IL-6 inhibition markedly improved clinical response and reduced the

levels of inflammation markers but produced no changes in lesion healing compared to placebo in active Chron disease patients (283).

## 6.2 STUDY II – COMPARATIVE EFFECTIVENESS OF BARICITINIB, TOFACITINIB AND BIOLOGICAL DMARDS IN RA

In study two we compared the effectiveness of the JAKis baricitinib and tofacitinib with that of bDMARDs, for treating RA, and we observed higher drug retention and at least equivalent treatment response on baricitinib compared to bDMARDs, and no significant differences between tofacitinib and bDMARDs.

The treatment response measures for Study II were chosen based on available data and in order to cover a wide range of disease parameters modifiable by treatment such as disease activity (DAS28, CDAI) and functionality (HAQ-DI). Certain domains, such a radiographic progression, were not covered by the available data, thus could not be used. As mentioned in Section 1.1, DAS28 includes laboratory values (acute phase reactants ESR or CRP) on which IL-6 inhibition has a strong effect (284), which is reflected in our findings. Since signaling via the IL-6 receptor is transduced by JAK isoforms 1 and 2, and both baricitinib and tofacitinib have relatively high affinity for JAK1 (baricitinib also for JAK2), it may be expected that JAKi have strong effect on acute phase reactants (ESR, CRP), similar to IL-6-receptor inhibitors (117,285). Our results showed that the effect on DAS28-ESR were significantly lower for the two JAKis compared to IL-6 receptor inhibitors. This could suggest nuanced modulation of the IL-6 signal-transduction pathway by JAKis compared to direct blockade of the IL-6 receptors achieved by tocilizumab and sarilumab. Potential further evidence in support of this hypothesis comes from lower a lower risk of gastro-intestinal perforations observed for tofacitinib compared to tocilizumab one another study (113).

Crude one-year response proportions were lower on tofacitinib compared to baricitinib on all measures. However, in Sweden baricitinib is the preferred JAKi (mainly due to its lower price), tofacitinib being reserved for patients who failed several other b/tsDMARDs. In consequence, patients initiating tofacitinib may have a less tractable disease compared to those initiating baricitinib, showing poor treatment response. After adjustment for confounding (including line of therapy), there was no evidence of difference in EULAR good response or CDAI remission between the two JAKi.

JAKi are expected to be effective as monotherapy, since they do not require the protection of an added csDMARDs against immunogenicity (i.e. antibodies generated against biological drugs) (40,117). However, we observed higher response proportions for tofacitinib in combination with csDMARDs versus monotherapy, confirming the findings of the ORAL-STRATEGY trial (49). This may suggest a synergic effect. Conversely, one year response rates in our study were similar between baricitinib monotherapy and combination therapy and this is in line with findings from the RA-BEGIN trial (130).

One important limitation of our study was the high proportion of missing outcome values despite measurement of disease activity within wide windows around baseline and end-points.

We used disease activity data from SRQ, which were not collected and recorded for the purpose of the study (secondary data), thus they were not collected at fixed time-points according to the study protocol, but at de facto clinical visits. To avoid potential selection bias from analyzing only observations with available outcome data, we imputed missing observations using multiple imputation. As explained in Section 4.2.6, multiple imputation relies on the assumption that, conditional on a certain set of variables, the missingness in the imputed variable is independent of the actual (unobserved) values of the variable. We imputed conditional on all variables included in the analysis, plus the on-drug indicator, assuming that this set is sufficient to achieve conditional random missingness. It may be reassuring that the results of a complete case analysis, conditional on all baseline covariates that were included in the multiple imputation models, produced similar results to analyzing the multiply imputed data.

The binary one-year treatment response for patients who discontinued baseline treatment before evaluation was classified as a negative response (i.e. non-responder imputation), regardless of the reason for discontinuation (other than remission). The main reported reason was ineffectiveness for all compared treatments, however other reasons were not uncommon, with discontinuation due to adverse events being the second most common. While stopping the treatment because of adverse events may be conceptualized as ineffectiveness, stopping treatment due other reasons may not be. Thus, under non-responder imputation analysis, the true response proportions had all patients continued treatment to evaluation, would be underestimated and the relative effectiveness of treatments may be biased as well (201). To challenge this analysis, we also compared treatment responses only among patients who continued treatment to one-year. This favors treatments with high discontinuation rates, where only patients who remained on treatment (thus for whom treatment was effective) are kept in the analysis. Despite being biased in the opposite direction, this sensitivity analysis supported the conclusions of the main analysis, where baricitinib was at least equivalent to alternative treatments. The results at one year were also supported by the comparison of changes from baseline in 3 months in disease activity and disability, where few treatments had been discontinued and “non-responder” imputation was not applied. An alternative would be to use multiple imputation to impute the treatment response of patients who discontinued treatment. However, at least when treatment discontinuation was due to ineffectiveness the missing response is likely not missing at random. Regardless of conditioning on other variables, missingness likely depends on the value of disease activity, which is part of the response definition. For this reason, we preferred to restrict multiple imputation to imputing missing response among patients on treatment.

As with all treatments, the foreseen benefits should outweigh the risks and for tofacitinib recent evidence suggests a dose response increase in the risk of major cardiovascular outcomes, venous thromboembolism and malignancy, in patients with high risk factors for cardiovascular disease (118,286). Safety warning have been applied to all JAKi and their use should be restricted to patients who did not respond to TNFi, which have more established safety profiles.

## 6.3 STUDY III – EMULATION OF THE SWEFOT TRIAL IN OBSERVATIONAL DATA

According to the trial emulation framework (see Section 1.4), the design of any etiological observational study may benefit from paralleling a theoretical target trial. In study III we explicitly emulated an existing trial with the aim of assessing if the observational emulation would successfully replicate the (assumingly unbiased) trail results. SWEFOT was identified as a good target for emulation for two main reasons. First of all, SWEFOT was an open-label trial, meaning that patients and physicians were not blinded to treatment. Blinding of prescribers and patients is impossible to mimic in secondary observational data. Second, we had access to the register in which the trial was nested, that is SRQ, which contains data on disease activity, essential for assessing eligibility and treatment response the same way as in the trial. The emulation would have been more challenging if we had to measure treatment response by proxy.

Nonetheless, even with access to similar data as collected in the trial, the SWEFOT protocol could not be perfectly emulated. During the feasibility assessment phase, we realized that some compromises had to be done to increase the size of the SSZ + HCQ cohort. In SWEFOT, participants randomized to receive added SSZ + HCQ over the ongoing MTX background were started on SSZ and HCQ simultaneously. In our emulation study we decided to allow sequential initiation of the two drugs, ensuring that the drug initiated first continued to be prescribed after initiation the second. Since the baseline was set at the initiation of the second drug in the combination, the use of the first drug before baseline may have introduced selection bias, as described in Section 4.2.3. Furthermore, immortal time bias may have been introduced by including in SSZ + HCQ cohort only patients who collected at least one prescription for the first drug after baseline (to ensure that the second drug was added over the first instead of a switch). In a sensitivity analysis we used a stricter definition for the SSZ + HCQ cohort in which SSZ and HCQ were only allowed to be initiated simultaneously. The effect estimate under this alternative definition was similar to the main analysis, but this does not necessarily prove that no bias was present under the main definition. It could be the case that the net bias was close to null due to biasing associations in opposite directions cancelling each other out. Nonetheless, it should be noted that 67% of the patients who initiated SSZ and HCQ according to the main definition initiated the drugs simultaneously.

Several eligibility criteria were relaxed compared to SWEFOT. As described in Section 4.3.3, patients were allowed to have initiated MTX longer than one year after RA debut (as long as no other DMARDs were used) and we also allowed more variability in the length of time between MTX initiation and the initiation of the study treatments. Furthermore, no restriction was set on the amount of glucocorticoid used before baseline.

Despite these deviations from the trial protocol, the main results of the emulation were similar to those of SWEFOT, with largely overlapping confidence intervals. Residual bias cannot be

excluded from observational studies, but in this case, broadly restricting the study sample to patients eligible for the target trial and having access to potentially important confounder data such as disease activity, may have reduced bias and we observed similar results to the target trail.

When further relaxing eligibility criteria in sensitivity analyses, the results gradually diverged from the trial results. Allowing longer a longer RA duration at baseline and prior treatment with other DMARDs than MTX reduced response proportions and the contrast between treatments. Reduced responder proportions are expected if patients who have failed previous treatments are included (140). Reduced contrasts may be the result of effect modification which has been observed previously (153) but it may also reflect residual confounding since in our population there were few patients previously treated with non-MTX DMARDs in the SSZ + HCQ cohort, making adjustment for this important confounder difficult (see discussion about positivity in Section 4.2.5).

Other observational studies confirmed the results of SWEFOT (151–154) but two blinded trials did not find TNFi (+ MTX) superior to SSZ + HCQ (+ MTX) (37,155). Other differences might explain the divergent results, such as analyzing data as intention-to-treat despite allowing treatment switches (37), but blinding might have contributed.

Dismissing observational studies as biased when they obtain different results from RCTs may be contra productive when observational studies are compared to RCTs they were not designed to mimic. Using different populations, definitions and analyses would lead to different results even in the absence of bias. Many RCTs employ outcome measures that are not used in routine clinical practice, thus they are not accessible to observational studies. Different study populations, with different compositions of effect modifiers would lead to different average treatment effects. Last but not least, different causal estimands and statistical analysis could lead to different results (e.g. intention-to-treat versus per-protocol analysis).

## 6.4 STUDY IV – GLUCOCORTICOIDS AND THE RISK OF SERIOUS INFECTIONS IN RA

In line with previous observational studies, study IV showed that oral glucocorticoids are associated with an increased risk of serious infections, in a dose dependent manner, and that the strength of the association decreases over time.

Nevertheless, it is difficult to compare the magnitude of absolute and relative risk estimates between studies, mainly due to the diverse ways in which exposure was measured and modelled, as well as due to differences in other study design elements, such as outcome definition or study populations. In a recent cohort study, George et al. reported one-year cumulative serious infections incidence proportions of 5.0%, 6.7%, 9.1% and 11.4% among Medicare RA patients exposed to no glucocorticoids, ≤ 5 mg/day, 5 to 10 mg/day, and >10 mg/day respectively. This cohort consisted of older patients (mean age of 69 years) with a higher prevalence of diabetes (22%) and COPD (13%). In Optum, another US insurance claims database which contains younger (mean age of 58 years) and healthier (diabetes 14% and

COPD 6%) patients, the corresponding cumulative incidence proportions were 2.9%, 3.8%, 6.1% and 9.1% (170). These was still higher than in our study where we estimated 1.8% among patients not exposed to glucocorticoids, 2.9% among patients exposed to ≤ 10 mg/day and 4.9% among patients exposed to > 10 mg/day at one year. However, the prevalence of diabetes in our cohort was around 6.5% and that of obstructive respiratory diseases of 3.5%. A younger RA population with more similar co-morbidity distribution to our study (mean age 56 years and prevalence of diabetes 6%), extracted from the UK CPRD database, was analyzed in a case-control study. The authors reported a serious infections odds ratio of 1.3 when comparing any glucocorticoid use versus no use between study entry and index-date. When dividing the study period into current (180 days before index date) and past, the odds ratios were 1.6 for any current use and 1.0 for any earlier use versus no use (86). Even though the glucocorticoid exposure history was modelled differently these results are comparable to our results. Our results also agree with the observation of Dixon et al. in that the association between glucocorticoids exposure and the risk of infection diminishes over time, approaching zero around one year before the time of risk evaluation (i.e. the index date in Dixon's case-control study) (177).

Compared to several of the previous studies (174,176–178), we observed a reduced contrast between the risk of serious infection under oral glucocorticoids exposure versus no exposure, especially at doses lower than 10 mg/day. The more appropriate adjustment for time-varying confounders via inverse probability weighting does not altogether explain these differences since we obtained similar results after conventional regression adjustment for baseline confounder values only. Two of these previous studies included older individuals (median of over 75 years compared to 63 years in our study), which may be an explanation (176,177).

Recently published results from the GLORIA trial also support an increased risk of infections even for low dose prednisolone (5 mg/day) compared to placebo, among RA patients older than 65 years (287).

The study of time-varying treatment histories requires the measurement of time-varying covariates to be used for bias adjustment. An important limitation of our study was the incomplete longitudinal data on DAS28 and HAQ-DI, which were measured at irregular intervals in clinical practice. The ideal method for dealing with missing data is multiple imputation, as described in Section 4.2.6, but this was not feasible in our large longitudinal data-set. Instead, we used the last observation carried forward (LOCF) approach and, in a sensitivity analysis, the missingness patterns approach. The MSM estimated using the missingness pattern approach was similar to that estimated using LOCF. A simulation study comparing different imputation methods in a setting with incomplete time-varying confounding affected by pervious treatment, showed larger bias for the missingness pattern approach compared to LOCF in the setting where missingness was influenced by the outcome. This is unlikely in our study where data were recorded prospectively (239). Under MCAR or MAR where covariate missingness was influenced only by treatment and other covariates, the bias in the results of the two methods was similar and rather low. Also, LOCF was only

moderately biased if covariates acted as confounders only where observed (i.e. under missingness pattern assumption) but the bias was higher if using the missingness pattern approach to analyze data generated under LOCF assumptions. Thus, the possibility of sparse time-updated covariate data in studies where data was not collected at fixed, regular time-points over follow-up makes the implementation of MSM challenging.

Another limitation, this time related to analyzing time-varying data using MSM specifically, is the statistical inefficiency due to the possibility of obtaining extreme weights due to multiplication over many time-points (as shown in Figure 4.2.5.4). IPTW need to be updated every time when the exposure distribution changes. With very precise exposure information, changes may be observed daily or weekly. In a common dosing scheme employed when glucocorticoids are used as bridging therapy to cover the slow onset of csDMARD effects, glucocorticoids are introduced in a higher dose and then tapered over the course of a few weeks. For example, in the COBRA trial, the dose of glucocorticoid was reduced every week (288). It may not be feasible to divined a long follow-up into daily or weekly observations for each individual, and calculate IPTWs as products over hundreds of observations. Nonetheless, prescription data usually available to observational studies rarely contains such detailed information. As mentioned in Section 4.1, the PDR only contains dosing data in an optional free text field. Such data was missing for many prescriptions and was difficult to use. Instead we used the dispensed quantities of prednisone equivalents to estimate average daily doses over 90-day periods, and we assumed that the daily dose remained constant for 90 days. This allowed a maximum of 12 follow-up episodes per individual. Acknowledging the imprecision of the average dose, we grouped the calculated daily doses in broader dose categories. Categorizing exposure also facilitated the use of logistic regression to estimate the probability of exposure conditional on covariate and exposure history (necessary for calculating the weights).

# 7 CONCLUSIONS, SIGNIFICANCE AND PERSPECTIVES

The projects included in this thesis employed modern methods for etiological observational study design and data analysis to address several gaps in knowledge about the safety and effectiveness of novel RA treatments.

The conclusions of each study as well as some possible directions for future research are presented in the sections below.

## 7.1 STUDY I – BIOLOGICAL DMARDS AND THE RISK OF GASTRO-INTESTINAL PERFORATIONS IN RA

In the first study we identified an increased risk of lower GI perforations for tocilizumab compared to other bDMARDs. The absolute incidence rate remained low (bellow 5 events per 1000 person-years) even for tocilizumab, but the risk is serious considering the potential severe sequela of GI perforation, thus patients for whom tocilizumab treatment is planned should be screened for additional risk factors (such as, history of diverticulitis, co-treatment with glucocorticoids and NSAIDs). The risk of lower GI perforations was higher among RA patients, whether treated with bDMARDs or with csDMARDs, compared to general population controls, but, after adjusting for the use of glucocorticoids before baseline, the risk remained significantly higher only for RA patients treated with tocilizumab.

Despite heterogeneity in incidence rate point estimates, all observational studies published to this moment seem to agree that tocilizumab increases the risk of GI perforation compared to alternative bDMARDs and with csDMARDs.

Sarilumab, a newer IL-6 receptor blocker approved in 2017, showed a lower incidence of GI perforations in an integrated analysis of RCTs compared to tocilizumab (1 event per 1000 person-years vs 3 events per 1000 person-years) (109,289). This may be explained by the exclusion of patients with a history of diverticulitis from sarilumab trials, since the risk of GI perforations was known from tocilizumab, which was not the case in tocilizumab trials. However, it would be interesting for future observational studies to estimate the risk of GI perforations among patients treated with sarilumab in clinical practice, and to compare the risk with that of tocilizumab. If a true difference exists, this should be further pursued in preclinical experiments that could uncover mechanism behind it, and potentially lead to safer IL-6 inhibitors.

A better understanding of the role played by IL-6 in the intestinal homeostasis could further guide the treatment with tocilizumab. It has been suggested that the timing of tocilizumab administration relative to the phase of intestinal injury might be relevant, since IL-6 may be essential in regulating the healing process early after injury (277). Maybe blocking IL-6 receptors later after the resolution of acute diverticulitis episodes could lessen the risk of perforation, hence allowing tocilizumab treatment in patients with a non-recent history of diverticular disease. On the other hand, one of the epidemiological studies discussed in

previous sections suggested that IL-6 blockade may also contribute to inducing diverticulitis, not only diverticular perforation among patients with existing diverticulitis (275).

## 7.2 STUDY II – COMPARATIVE EFFECTIVENESS OF BARICITINIB, TOFACITINIB AND BIOLOGICAL DMARDS IN RA

In the second study, baricitinib was at least as effective as bDMARDs for controlling RA, and potentially more effective than TNFi. Due to the limited use of tofacitinib for the treatment of RA in Sweden, our study was not suitably powered to compare this JAKi to alternative treatments. However, the adjusted results show no evidence of lower effectiveness compared to bDMARDs.

The differences in average changes in disease activity (DAS28, CDAI) and functionality (HAQ-DI) at three-months compared to baseline showed a similar pattern to the differences in treatment response proportions at one year. However, this does not necessarily mean that three-month improvements will predict one-year treatment responses. We have not explored this question in our study but it has been previously addressed for patients treated with bDMARDs (290–292). Based on the results of such studies, current treatment guidelines recommend the evaluation of improvements in disease activity after the first three months of treatment, to predict the likelihood of achieving remission or at least low disease activity (treatment target) by six months (33). Future studies could examine how continuing JAKi treatment after three months compares to switching to another therapy, among patients with different degrees of change in baseline disease activity after the first three months of JAKi therapy. However, considering the existing guidelines, it may be difficult to have observational data on patients who continued JAKi treatment after three months despite not achieving a high enough disease activity improvement by that time.

When prescribing JAKi, their benefits, relative to alternatives, should be weighed against their risks, relative to alternatives. Recent results from a large phase IV (post-marketing) safety trial, the ORAL Surveillance, comparing tofacitinib to a TNFi (adalimumab or etanercept), showed an increased risk of major cardiovascular events and cancer associated with tofacitinib, for patients older than 50 years with an increased risk for cardiovascular events (286). An observational study found no differences in the incidence of cardiovascular events between tofacitinib and TNFi in a wider RA population, but obtained similar results to ORAL Surveillance when narrowing down the population to patients with an increased risk of cardiovascular events (293). The findings of ORAL Surveillance triggered regulatory responses amounting to tofacitinib as well as baricitinib and upadacitinib being recommended only after the failure of initial TNFi.(118) Considering the differential affinity to JAKi isoforms, the safety profiles of distinct JAKis may not be identical, thus future safety studies should confirm the ORAL Surveillance finding for baricitinib and other JAKis, and should further update the safety profiles of these new treatments.

## 7.3 STUDY III – EMULATION OF THE SWEFOT TRIAL IN OBSERVATIONAL DATA

In the third study, we emulated the SWEFOT trial using observational data, and we obtained similar results. Comparing the addition of infliximab versus SSZ + HCQ over a background of MTX, among patients who did not respond to initial MTX therapy, the emulation showed a treatment response ratio of 1.48 versus 1.59 in the target trial, with broadly overlapping confidence intervals. These comparable results encourage the use of the trial emulation framework to generate real-world evidence that could complement RCT evidence.

To complement existing RCT evidence, observational studies have to emulate target trials that have not been or cannot be conducted. However, novel observational results that have not been benchmarked against existing trial results may not be credible even if they were generated under the verified trial emulation framework. One strategy to overcome this confidence barrier could be to start by emulating an existing trial but then to expend the results of the trial by, for example, relaxing eligibility criteria to include patients treated in the real-world but excluded from the trial, by adding additional treatment arms or by extending follow-up beyond that of the trial.

We consider it an advantage of our study to be able to conduct the emulation in the same register that the target trial was nested in (SRQ), which meant not only using the same source population (though in a different calendar period) but also having access to the same outcome measure and to baseline RA disease activity measures. Therefore, pragmatic trials nested in population registers may be ideal reference points to anchor observational studies conducted on the same registers.

To conclude, the emulation trial framework could be one solution to improving the internal validity and transparency in reporting observational studies with the goal of rendering them reliable sources of evidence for regulatory and clinical decision making.

## 7.4 STUDY IV – GLUCOCORTICOIDS AND THE RISK OF SERIOUS INFECTIONS IN RA

Our fourth study confirmed a dose dependent increase in the risk of serious infections associated with the use of glucocorticoids in rheumatoid arthritis, as previously reported in other observational studies. Our results also indicate that, the risk of serious infection is associated not only with the current glucocorticoid dose but also with the cumulated quantity of glucocorticoids received within the most recent year, but not with earlier exposure.

We expanded upon previous observational studies in several ways. First of all, we allowed exposure to vary over time, in order to explore the effects of different dose patterns. Second, we modelled the resulting time-varying treatment history in a flexible way, allowing for independent effects of exposure at different times within the treatment history. Third, we adjusted for time-varying confounding and selection bias using inverse probability weighting, which is preferred over conventional regression adjustment (see Section 4.2.5). Finally, we had

access to a broad range of time-updated individual patient information, including disease and medication history as well as RA disease activity, which we could use for bias adjustments.

Nonetheless, as previously discussed, two important limitation of our data in the context of studying a time-varying exposure are the unstructured follow-up and poor information on the daily glucocorticoid dose. Ideally early RA patients would be included in a prospective cohort study in which they would be evaluated at fixed, regular time-points and their glucocorticoid treatment (and potentially other treatments) would be (re-)adjusted at each evaluation based on recorded patient characteristics. Such a study could be nested in SRQ where RA clinical parameters and treatment adjustments could be recorded for each visit, and the outcome could be obtained via linkage with the NPR, as in our study IV.

To conclude, our findings from study IV support the current recommendations that glucocorticoids could be used to ease pain and inflammation, but for limited time and in the lowest dose effective (33,34). We estimated a relatively low increase in the cumulative incidence of serious infections when comparing the use of glucocorticoids in doses ≤ 10 mg (prednisone) /day for up to one year, compared to no use of glucocorticoids. This could motivate the focus of future research on low doses of glucocorticoids which may have acceptable benefit/risk ratios, at least for some patient groups.

The results of the four studies, combined with existing evidence about the benefits and risks of available treatment options from RCTs and other observational studies, shall contribute to better informing clinical decision making and ultimately improving the life of RA patients.

# 8 ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who have contributed in one way or another to making my PhD education possible and enjoyable.

To my supervisors:

**Thomas Frisell**, my main supervisor. Thank you for teaching me SAS and everything I know about Swedish national register data, for being so involved in all my projects, for countless discussions about stats and epi methods, but most importantly thank you for believing in me all those times when I did not believe in myself and encouraging me to move further.

**Johan Askling**, my co-supervisor. Thank you for teaching me so much about clinical epidemiology and introducing me to the fascinating science (and art) of rheumatology. Besides being a brilliant rheumatologist and epidemiologist, you are also a great boss who maintains an inspiring and welcoming research community at KEP, even during global pandemics.

**Bénédicte Delcoigne**, my co-supervisor. Thank you for sharing your statistical knowledge with me, and for always being so optimistic about my work.

To my mentor, **Rosaria Galanti**. Thank you for all your wise advice and teaching me how to best cope with the research life. You are an inspiration to me.

To **Arvid Sjölander**. Thank you for teaching me the foundations of causal inference in your doctoral course, and further offering me valuable support as a co-author in study IV.

To **Björn Pasternak** and his group. Thank you for inviting me to attend your pharmacoepi meetings. It's been great fun and I have learned a lot from you guys.

To my colleagues at KEP:

**Renata Zelic,** my desk mate and friend**.** I don't even know where to start with expressing my gratitude towards you. Thank you for always questioning things, making me think and be skeptical, for being an example of tenacity and for always finding the time to engage in an interesting discussion. I really think that exchanging ideas with you contributed a great deal to improving my knowledge. Thank you for a thorough proof read of my kappa and, above all, thanks for being a friend and making my life in Sweden less lonely.

**Peter Alping**, PhD comrade and hiking partner. Thanks for making programming fun during the Peter Python Club (sadly interrupted early by the pandemic) and for all the nice days spent hiking through the beautiful Swedish nature and fun afternoons spent playing board games.

**Marina Dehara, Kelsi Smith**, **Matilda Morin**, **Arda Yilal** and all the other PhD/Master students and researchers who I have crossed paths with. Thanks for making me feel at home in KEP.

**Viktor Winzell** and **Laura Pazzagli**, thank you both for all the interesting discussions we had about causal inference in pharmacoepidemiology. Viktor, we should somehow restart causal inference journal club / seminar somehow.

**Anda Gliga**, old friend. Thank you for convincing me to move to Stockholm and helping me navigate my new life here.

I would like to thank my family. You have always encouraged me, and without your support I would not have been able to move to Sweden to study, therefore I would not have started this PhD.

Finally, I would like to thank my girlfriend, **Nicoleta**, who has offered me unconditional love and supported my decision to move to Sweden for the master, and then stay for the PhD, even if this meant spending so much time away from one another. Thank you so much for being patient, kind, selfless and understanding. I love you!

# 9 REFERENCES


1. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Annals of the Rheumatic Diseases. 2003;62(7):611–6.
2. Kvien T. K. Epidemiology and Burden of Illness of Rheumatoid Arthritis. PharmacoEconomics. 2004;22(Supplement 1):1–12.
3. Theander L, Nyhäll-Wåhlin BM, Nilsson JÅ, Willim M, Jacobsson LTH, Petersson IF, et al. Severe Extraarticular Manifestations in a Community-based Cohort of Patients with Rheumatoid Arthritis: Risk Factors and Incidence in Relation to Treatment with Tumor Necrosis Factor Inhibitors. Journal of rheumatology. 2017;44(7):981–7.
4. Neovius M, Simard JF, Askling J. Nationwide prevalence of rheumatoid arthritis and penetration of disease-modifying drugs in Sweden. Annals of the Rheumatic Diseases. 2011;70(4):624–9.
5. Machold KP, Landewé R, Smolen JS, Stamm TA, van der Heijde DM, Verpoort KN, et al. The Stop Arthritis Very Early (SAVE) trial, an international multicentre, randomised, double-blind, placebo-controlled trial on glucocorticoids in very early arthritis. Annals of the Rheumatic Diseases. 2010;69(3):495–502.
6. Kyburz D, Gabay C, Michel BA, Finckh A. The long-term impact of early treatment of rheumatoid arthritis on radiographic progression: a population-based cohort study. Rheumatology (Oxford, England). 2011;50(6):1106–10.
7. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Annals of the Rheumatic Diseases. 2010 Jun 1;69(6):964–75.
8. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 Update of the American College of Rheumatology Recommended Rheumatoid Arthritis Disease Activity Measures. Arthritis Care & Research. 2019;71(12):1540–55.
9. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score With 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score Without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). Arthritis Care & Research. 2011;63(S11):S14–36.
10. Fransen J, Riel PLCM van. The Disease Activity Score and the EULAR Response Criteria. Rheumatic Disease Clinics. 2009 Nov 1;35(4):745–57.
11. Smolen JS, Aletaha D. The assessment of disease activity in rheumatoid arthritis. Clin Exp Rheumatol. 2010 Jun;28(3 Suppl 59):S18-27.
12. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: Dimensions and Practical Applications. Health Qual Life Outcomes. 2003 Jun 9;1:20.
13. Kirwan JR, Bijlsma JW, Boers M, Shea B. Effects of glucocorticoids on radiological progression in rheumatoid arthritis. Cochrane Database of Systematic Reviews [Internet].

2007 [cited 2020 Mar 6];(1). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD006356/abstract

14. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid arthritis. Nature Reviews Disease Primers. 2018 Feb 8;4:18001.
15. Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis: A Review. JAMA. 2018 Oct 2;320(13):1360–72.
16. Reddy V, Cohen S. Role of Janus Kinase inhibitors in rheumatoid arthritis treatment. Current Opinion in Rheumatology. 2021 May;33(3):300–6.
17. Bywall KS, Kihlbom U, Hansson M, Falahee M, Raza K, Baecklund E, et al. Patient preferences on rheumatoid arthritis second-line treatment: a discrete choice experiment of Swedish patients. Arthritis Research & Therapy. 2020 Dec 19;22(1):288.
18. Tanaka Y. The JAK inhibitors: do they bring a paradigm shift for the management of rheumatic diseases? Rheumatology. 2019 Feb 1;58(Supplement_1):i1–3.
19. Mysler E, Caubet M, Lizarraga A. Current and Emerging DMARDs for the Treatment of Rheumatoid Arthritis. Open Access Rheumatol. 2021 Jun 1;13:139–52.
20. Weise M, Bielsky MC, De Smet K, Ehmann F, Ekman N, Narayanan G, et al. Biosimilars—why terminology matters. Nat Biotechnol. 2011 Aug;29(8):690–3.
21. Caporali R, Crepaldi G, Codullo V, Benaglio F, Monti S, Todoerti M, et al. 20 years of experience with tumour necrosis factor inhibitors: what have we learned? Rheumatology. 2018 Oct 1;57(Supplement_7):vii5–10.
22. Maxwell L, Singh JA, Maxwell L. Abatacept for rheumatoid arthritis. Cochrane Database of Systematic Reviews. 2009;2010(1):CD007277-.
23. Bonelli M, Ferner E, Göschl L, Blüml S, Hladik A, Karonitsch T, et al. Abatacept (CTLA-4IG) treatment reduces the migratory capacity of monocytes in patients with rheumatoid arthritis. Arthritis & Rheumatism. 2013;65(3):599–607.
24. Garcia-Montoya L, Villota-Eraso C, Yusof MYM, Vital EM, Emery P. Lessons for rituximab therapy in patients with rheumatoid arthritis. The Lancet Rheumatology. 2020 Aug 1;2(8):e497–509.
25. Edwards JCW, Szczepański L, Szechiński J, Filipowicz-Sosnowska A, Emery P, Close DR, et al. Efficacy of B-Cell–Targeted Therapy with Rituximab in Patients with Rheumatoid Arthritis. New England Journal of Medicine. 2004 Jun 17;350(25):2572–81.
26. Lopez-Olivo MA, Urruela MA, McGahan L, Pollono EN, Suarez-Almazor ME. Rituximab for rheumatoid arthritis. Cochrane Database of Systematic Reviews [Internet]. 2015 [cited 2022 Jun 28];(1). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007356.pub2/abstract?cookiesEnabled
27. European Medicines Agency. MabThera EPAR [Internet]. European Medicines Agency. 2019 [cited 2020 Jan 15]. Available from: https://www.ema.europa.eu/en/documents/product-information/mabthera-epar-product-information_en.pdf
28. Dayer JM, Choy E. Therapeutic targets in rheumatoid arthritis: the interleukin-6 receptor. Rheumatology. 2010 Jan 1;49(1):15–24.
29. Scott LJ. Tocilizumab: A Review in Rheumatoid Arthritis. Drugs. 2017 Nov 1;77(17):1865–79.

30. Lamb YN, Deeks ED. Sarilumab: A Review in Moderate to Severe Rheumatoid Arthritis. Drugs. 2018 Jun 1;78(9):929–40.

31. EMA. RoActemra [Internet]. European Medicines Agency. 2018 [cited 2022 Jun 28]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/roactemra

32. EMA. Kevzara [Internet]. European Medicines Agency. 2018 [cited 2022 Jun 28]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/kevzara

33. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Annals of the Rheumatic Diseases. 2020 Jun 1;79(6):685–99.

34. Fraenkel L, Bathon JM, England BR, St.Clair EW, Arayssi T, Carandang K, et al. 2021 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis & Rheumatology. 2021;73(7):1108–23.

35. Gaujoux-Viala C, Smolen JS, Landewé R, Dougados M, Kvien TK, Mola EM, et al. Current evidence for the management of rheumatoid arthritis with synthetic disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. Annals of the rheumatic diseases. 2010;69(6):1004–9.

36. Pincus T, Yazici Y, Sokka T, Aletaha D, Smolen JS. Methotrexate as the "anchor drug" for the treatment of early rheumatoid arthritis. Clinical and experimental rheumatology. 2003;21(5 Suppl 31):S179-.

37. O'Dell JR, Mikuls TR, Taylor TH, Ahluwalia V, Brophy M, Warren SR, et al. Therapies for Active Rheumatoid Arthritis after Methotrexate Failure. The New England Journal of Medicine. 2013;369(4):307–18.

38. Nam JL, Ramiro S, Gaujoux-Viala C, Takase K, Leon-Garcia M, Emery P, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2013 update of the EULAR recommendations for the management of rheumatoid arthritis. Annals of the rheumatic diseases. 2014;73(3):516–28.

39. Kerschbaumer A, Sepriano A, Smolen JS, Heijde D van der, Dougados M, Vollenhoven R van, et al. Efficacy of pharmacological treatment in rheumatoid arthritis: a systematic literature research informing the 2019 update of the EULAR recommendations for management of rheumatoid arthritis. Annals of the Rheumatic Diseases. 2020 Jun 1;79(6):744–59.

40. Strand V, Goncalves J, Isaacs JD. Immunogenicity of biologic agents in rheumatology. Nat Rev Rheumatol. 2021 Feb;17(2):81–97.

41. Burmester GR, Lin Y, Patel R, van Adelsberg J, Mangan EK, Graham NMH, et al. Efficacy and safety of sarilumab monotherapy versus adalimumab monotherapy for the treatment of patients with active rheumatoid arthritis (MONARCH): a randomised, double-blind, parallel-group phase III trial. Annals of the Rheumatic Diseases. 2017;76(5):840–7.

42. Gabay C., Emery P., Van Vollenhoven R., Dikranian A., Alten R., Pavelka K., et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): A randomised, double-blind, controlled phase 4 trial. Lancet. 2013;381(9877):1541–50.

43. Klarenbeek NB, Güler-Yüksel M, van der Kooij SM, Han KH, Ronday HK, Kerstens PJSM, et al. The impact of four dynamic, goal-steered treatment strategies on the 5-year outcomes of rheumatoid arthritis patients in the BeSt study. Annals of the Rheumatic Diseases. 2011;70(6):1039–46.

44. van Mulligen E, de Jong PHP, Kuijper TM, van der Ven M, Appels C, Bijkerk C, et al. Gradual tapering TNF inhibitors versus conventional synthetic DMARDs after achieving controlled disease in patients with rheumatoid arthritis: first-year results of the randomised controlled TARA study. Annals of the rheumatic diseases. 2019;78(6):746–53.

45. Haschka J, Englbrecht M, Hueber AJ, Manger B, Kleyer A, Reiser M, et al. Relapse rates in patients with rheumatoid arthritis in stable remission tapering or stopping antirheumatic therapy: interim results from the prospective randomised controlled RETRO study. Annals of the Rheumatic Diseases. 2016;75(1):45–51.

46. Smolen JS, Burmester GR, Combe B, Curtis JR, Hall S, Haraoui B, et al. Head-to-head comparison of certolizumab pegol versus adalimumab in rheumatoid arthritis: 2-year efficacy and safety results from the randomised EXXELERATE study. The Lancet. 2016 Dec 3;388(10061):2763–74.

47. Schiff M, Weinblatt ME, Valente R, Heijde D van der, Citera G, Elegbe A, et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis: two-year efficacy and safety findings from AMPLE trial. Annals of the Rheumatic Diseases. 2014 Jan 1;73(1):86–94.

48. Giles JT, Sattar N, Gabriel S, Ridker PM, Gay S, Warne C, et al. Cardiovascular Safety of Tocilizumab Versus Etanercept in Rheumatoid Arthritis: A Randomized Controlled Trial. Arthritis & Rheumatology. 2020;72(1):31–40.

49. Fleischmann R, Mysler E, Hall S, Kivitz AJ, Moots RJ, Luo Z, et al. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. The Lancet. 2017 Jul 29;390(10093):457–68.

50. Taylor PC, Keystone EC, van der Heijde D, Weinblatt ME, del Carmen Morales L, Reyes Gonzaga J, et al. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. The New England journal of medicine. 2017;376(7):652–62.

51. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence — What Is It and What Can It Tell Us? New England Journal of Medicine. 2016 Dec 8;375(23):2293–7.

52. Bolislis WR, Fay M, Kühler TC. Use of Real-world Data for New Drug Applications and Line Extensions. Clinical Therapeutics. 2020 May 1;42(5):926–38.

53. Lauper K, Hyrich KL. How effective are JAK-inhibitors? Perspectives from clinical trials and real-world studies. Expert Review of Clinical Immunology. 2022 Mar 4;18(3):207–20.

54. Stürmer T, Jonsson Funk M, Poole C, Brookhart MA. Nonexperimental Comparative Effectiveness Research Using Linked Healthcare Databases. Epidemiology. 2011 May;22(3):298–301.

55. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? Clinical pharmacology and therapeutics. 2017;102(6):924–33.

56. Jonker CJ, van den Berg HM, Kwa MSG, Hoes AW, Mol PGM. Registries supporting new drug applications. Pharmacoepidemiology and Drug Safety. 2017;26(12):1451–7.

57. ElZarrad MK, Corrigan-Curay J. The US Food and Drug Administration's Real-World Evidence Framework: A Commitment for Engagement and Transparency on Real-World Evidence. Clinical Pharmacology & Therapeutics. 2019;106(1):33–5.

58. Olsen J, Basso O, Sørensen HT. What is a population-based registry? Scand J Public Health. 1999 Jan 1;27(1):78–78.

59. Schneeweiss S. Real-World Evidence of Treatment Effects: The Useful and the Misleading. Clinical Pharmacology & Therapeutics. 2019;106(1):43–4.

60. Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. Clinical Pharmacology & Therapeutics. 2019;106(1):36–9.

61. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. Clinical Pharmacology & Therapeutics. 2019;105(4):867–77.

62. Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. The New England journal of medicine. 2020;382(7):674–8.

63. Vandenbroucke JP, Psaty BM. Benefits and Risks of Drug Treatments: How to Combine the Best Evidence on Benefits With the Best Data About Adverse Effects. JAMA. 2008 Nov 26;300(20):2417–9.

64. Nash P. Real-world Evidence Needs Careful Interpretation. The Journal of Rheumatology. 2021 Jan 1;48(1):1–2.

65. Yuan H, Ali MS, Brouwer ES, Girman CJ, Guo JJ, Lund JL, et al. Real-World Evidence: What It Is and What It Can Tell Us According to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG). Clinical Pharmacology & Therapeutics. 2018;104(2):239–41.

66. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758–64.

67. Zhao SS, Lyu H, Solomon DH, Yoshida K. Improving rheumatoid arthritis comparative effectiveness research through causal inference principles: systematic review using a target trial emulation framework. Annals of the Rheumatic Diseases. 2020 Jul 1;79(7):883–90.

68. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. N Engl J Med. 2017 Oct 5;377(14):1391–8.

69. Schwartz D, Lellouch J. Explanatory and Pragmatic Attitudes in Therapeutical Trials. Journal of Clinical Epidemiology. 2009 May 1;62(5):499–505.

70. Ford I, Norrie J. Pragmatic Trials. New England Journal of Medicine. 2016 Aug 4;375(5):454–63.

71. Hróbjartsson A, Gøtzsche PC. Is the Placebo Powerless? New England Journal of Medicine. 2001 May 24;344(21):1594–602.

72. Lund JL, Richardson DB, Stürmer T. The Active Comparator, New User Study Design in Pharmacoepidemiology: Historical Foundations and Contemporary Application. Current epidemiology reports. 2015;2(4):221–8.

73. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. Nature Reviews Rheumatology. 2015 Jul;11(7):437–41.

74. Walker AM, Patrick AR, Lauer MS, Hornbrook MC, Marin MG, Platt R, et al. A tool for assessing the feasibility of comparative effectiveness research. CER. 2013 Jan 30;3:11–20.

75. Yoshida K, Solomon DH, Haneuse S, Kim SC, Patorno E, Tedeschi SK, et al. A tool for empirical equipoise assessment in multigroup comparative effectiveness research. Pharmacoepidemiology and Drug Safety. 2019;28(7):934–41.

76. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. Clinical Trials. 2012 Feb 1;9(1):48–55.

77. U.S. Food and Drug Administration. Public Workshop: Evaluating Inclusion and Exclusion Criteria In Clinical Trials. Workshop Report. [Internet]. The National Press Club; 2018 [cited 2020 Jul 15]. Available from: https://www.fda.gov/media/134754/download

78. Hernán MA, Hernández-Díaz S, Robins JM. A Structural Approach to Selection Bias. Epidemiology. 2004;15(5):615–25.

79. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of Clinical Epidemiology. 2016;79:70–5.

80. Choi HK, Nguyen US, Niu J, Danaei G, Zhang Y. Selection bias in rheumatic disease research. Nat Rev Rheumatol. 2014 Jul;10(7):403–12.

81. Stürmer T, Wang T, Golightly YM, Keil A, Lund JL, Jonsson Funk M. Methodological considerations when analysing and interpreting real-world data. Rheumatology (Oxford). 2020 Jan 1;59(1):14–25.

82. Ray WA. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. American journal of epidemiology. 2003;158(9):915–20.

83. Solomon DH, Mercer E, Kavanaugh A. Observational studies on the risk of cancer associated with tumor necrosis factor inhibitors in rheumatoid arthritis: A review of their methodologies and results. Arthritis & Rheumatism. 2012;64(1):21–32.

84. Hernán MA. Counterpoint: Epidemiology to Guide Decision-Making: Moving Away From Practice-Free Research. American Journal of Epidemiology. 2015 Nov 15;182(10):834–9.

85. Myasoedova E, Matteson EL, Talley NJ, Crowson CS. Increased Incidence and Impact of Upper and Lower Gastrointestinal Events in Patients with Rheumatoid Arthritis in Olmsted County, Minnesota: A Longitudinal Population-based Study. The Journal of Rheumatology. 2012 Jul 1;39(7):1355–62.

86. Wilson JC, Sarsour K, Gale S, Pethö-Schramm A, Jick SS, Meier CR. Incidence and Risk of Glucocorticoid-Associated Adverse Effects in Patients With Rheumatoid Arthritis. Arthritis Care & Research. 2019;71(4):498–511.

87. Lanas A. A review of the gastrointestinal safety data—a gastroenterologist's perspective. Rheumatology. 2010 May 1;49(suppl_2):ii3–10.

88. Humes DJ, Solaymani–Dodaran M, Fleming KM, Simpson J, Spiller RC, West J. A Population-Based Study of Perforated Diverticular Disease Incidence and Associated Mortality. Gastroenterology. 2009 Apr 1;136(4):1198–205.

89. Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal Toxicity of Nonsteroidal Antiinflammatory Drugs. New England Journal of Medicine. 1999 Jun 17;340(24):1888–99.

90. MacDonald TM, Morant SV, Robinson GC, Shield MJ, McGilchrist MM, Murray FE, et al. Association of upper gastrointestinal toxicity of non-steroidal anti-inflammatory drugs with continued exposure: cohort study. BMJ. 1997 Nov 22;315(7119):1333–7.

91. Lanza LL, Walker AM, Bortnichak EA, Dreyer NA. Peptic Ulcer and Gastrointestinal Hemorrhage Associated With Nonsteroidal Anti-inflammatory Drug Use in Patients Younger Than 65 Years: A Large Health Maintenance Organization Cohort Study. Archives of Internal Medicine. 1995 Jul 10;155(13):1371–7.

92. Hernández-Díaz S, Rodríguez LA. Association between nonsteroidal anti-inflammatory drugs and upper gastrointestinal tract bleeding/perforation: an overview of epidemiologic studies published in the 1990s. Arch Intern Med. 2000 Jul 24;160(14):2093–9.

93. Fries JF, Murtagh KN, Bennett M, Zatarain E, Lingala B, Bruce B. The rise and decline of nonsteroidal antiinflammatory drug–associated gastropathy in rheumatoid arthritis. Arthritis & Rheumatism. 2004;50(8):2433–40.

94. Lanas A, García-Rodríguez LA, Polo-Tomás M, Ponce M, Alonso-Abreu I, Perez-Aisa MA, et al. Time Trends and Impact of Upper and Lower Gastrointestinal Bleeding and Perforation in Clinical Practice. Am J Gastroenterol. 2009 May 5;104(7):1633–41.

95. Laine L, Yang H, Chang SC, Datto C. Trends for Incidence of Hospitalization and Death Due to GI Complications in the United States From 2001 to 2009. Official journal of the American College of Gastroenterology | ACG. 2012 Aug;107(8):1190–5.

96. Jagpal A, Curtis JR. Gastrointestinal Perforations with Biologics in Patients with Rheumatoid Arthritis: Implications for Clinicians. Drug Saf. 2018 Jun 1;41(6):545–53.

97. Morris CR, Harvey IM, Stebbings WSL, Hart AR. Incidence of perforated diverticulitis and risk factors for death in a UK population. Br J Surg. 2008 Jul;95(7):876–81.

98. Mpofu S, Mpofu CMA, Hutchinson D, Maier AE, Dodd SR, Moots RJ. Steroids, non-steroidal anti-inflammatory drugs, and sigmoid diverticular abscess perforation in rheumatic conditions. Annals of the Rheumatic Diseases. 2004 May 1;63(5):588–90.

99. Humes DJ, Fleming KM, Spiller RC, West J. Concurrent drug use and the risk of perforated colonic diverticular disease: a population-based case–control study. Gut. 2011 Feb 1;60(2):219–24.

100. Chang CH, Lin JW, Chen HC, Kuo CW, Shau WY, Lai MS. Non-steroidal anti-inflammatory drugs and risk of lower gastrointestinal adverse events: a nationwide study in Taiwan. Gut. 2011 Oct 1;60(10):1372–8.

101. Sostres C, Gargallo CJ, Lanas A. Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. Arthritis Res Ther. 2013 Jul 24;15(3):S3.

102. Curtis JR, Xie F, Chen L, Spettell C, McMahan RM, Fernandes J, et al. The incidence of gastrointestinal perforations among rheumatoid arthritis patients. Arthritis & Rheumatism. 2011 Feb 1;63(2):346–51.

103. Curtis JR, Lanas A, John A, Johnson DA, Schulman KL. Factors associated with gastrointestinal perforation in a cohort of patients with rheumatoid arthritis. Arthritis Care Res. 2012 Dec 1;64(12):1819–28.

104. Gout T, Östör AJK, Nisar MK. Lower gastrointestinal perforation in rheumatoid arthritis patients treated with conventional DMARDs or tocilizumab: a systematic literature review. Clin Rheumatol. 2011 Nov 1;30(11):1471.

105. Tursi A, Papagrigoriadis S. Review article: the current and evolving treatment of colonic diverticular disease. Alimentary Pharmacology & Therapeutics. 2009;30(6):532–46.

106. Corsi F, Previde P, Colombo F, Cellerino P, Donati M, Trabucchi E. Two cases of intestinal perforation in patients on anti-rheumatic treatment with etanercept. Clin Exp Rheumatol. 2006 Feb;24(1):113.

107. Bykerk VP, Cush J, Winthrop K, Calabrese L, Lortholary O, Longueville M de, et al. Update on the safety profile of certolizumab pegol in rheumatoid arthritis: an integrated analysis from clinical trials. Annals of the Rheumatic Diseases. 2015 Jan 1;74(1):96–103.

108. Závada J, Lunt M, Davies R, Low AS, Mercer LK, Galloway JB, et al. The risk of gastrointestinal perforations in patients with rheumatoid arthritis treated with anti-TNF therapy: results from the BSRBR-RA. Ann Rheum Dis. 2014 Jan;73(1):252–5.

109. Schiff MH, Kremer JM, Jahreis A, Vernon E, Isaacs JD, van Vollenhoven RF. Integrated safety in tocilizumab clinical trials. Arthritis Research & Therapy. 2011 Sep 1;13:R141.

110. Curtis JR, Perez-Gutthann S, Suissa S, Napalkov P, Singh N, Thompson L, et al. Tocilizumab in rheumatoid arthritis: A case study of safety evaluations of a large postmarketing data set from multiple data sources. Seminars in Arthritis and Rheumatism. 2015 Feb 1;44(4):381–8.

111. Strangfeld A., Richter A., Siegmund B., Herzer P., Rockwitz K., Demary W., et al. Risk for lower intestinal perforations in patients with rheumatoid arthritis treated with tocilizumab in comparison to treatment with other biologic or conventional synthetic DMARDs. Ann Rheum Dis. 2017;76(3):504–10.

112. Monemi S, Berber E, Sarsour K, Wang J, Lampl K, Bharucha K, et al. Incidence of Gastrointestinal Perforations in Patients with Rheumatoid Arthritis Treated with Tocilizumab from Clinical Trial, Postmarketing, and Real-World Data Sources. Rheumatol Ther. 2016 Dec 1;3(2):337–52.

113. Xie F, Yun H, Bernatsky S, Curtis JR. Brief Report: Risk of Gastrointestinal Perforation Among Rheumatoid Arthritis Patients Receiving Tofacitinib, Tocilizumab, or Other Biologic Treatments. Arthritis & Rheumatology (Hoboken, NJ). 2016 Nov;68(11):2612–7.

114. Curtis JR, Chen SY, Werther W, John A, Johnson DA. Validation of ICD-9-CM codes to identify gastrointestinal perforation events in administrative claims data among hospitalized rheumatoid arthritis patients. Pharmacoepidemiol Drug Saf. 2011 Nov 1;20(11):1150–8.

115. O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A. The JAK-STAT pathway: impact on human disease and therapeutic intervention. Annu Rev Med. 2015;66:311–28.

116. Lin CM, Cooles FA, Isaacs JD. Basic Mechanisms of JAK Inhibition. Mediterr J Rheumatol. 2020 Jun 11;31(Suppl 1):100–4.

117. Choy EH. Clinical significance of Janus Kinase inhibitor selectivity. Rheumatology. 2019 Jun 1;58(6):953–62.

118. Winthrop KL, Cohen SB. Oral surveillance and JAK inhibitor safety: the theory of relativity. Nat Rev Rheumatol. 2022 Mar 22;1–4.

119. Finckh A, Tellenbach C, Herzog L, Scherer A, Moeller B, Ciurea A, et al. Comparative effectiveness of antitumour necrosis factor agents, biologics with an alternative mode of action and tofacitinib in an observational cohort of patients with rheumatoid arthritis in Switzerland. RMD Open. 2020 May 1;6(1):e001174.

120. Bird P, Littlejohn G, Butcher B, Smith T, da Fonseca Pereira C, Witcombe D, et al. Real-world evaluation of effectiveness, persistence, and usage patterns of tofacitinib in treatment of rheumatoid arthritis in Australia. Clin Rheumatol. 2020 Sep 1;39(9):2545–51.

121. EMA. Xeljanz Authorization [Internet]. European Medicines Agency. 2018 [cited 2022 Apr 14]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/xeljanz

122. EMA. Xeljanz Refusal [Internet]. European Medicines Agency. 2018 [cited 2022 Apr 14]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/xeljanz-0

123. Lee EB, Fleischmann R, Hall S, Wilkinson B, Bradley JD, Gruben D, et al. Tofacitinib versus Methotrexate in Rheumatoid Arthritis. The New England journal of medicine. 2014;370(25):2377–86.

124. Fleischmann R, Kremer J, Cush J, Schulze-Koops H, Connell CA, Bradley JD, et al. Placebo-Controlled Trial of Tofacitinib Monotherapy in Rheumatoid Arthritis. New England Journal of Medicine. 2012 Aug 9;367(6):495–507.

125. Kremer J, Li ZG, Hall S, Fleischmann R, Genovese M, Martin-Mola E, et al. Tofacitinib in Combination With Nonbiologic Disease-Modifying Antirheumatic Drugs in Patients With Active Rheumatoid Arthritis. Ann Intern Med. 2013 Aug 20;159(4):253–61.

126. Heijde D van der, Tanaka Y, Fleischmann R, Keystone E, Kremer J, Zerbini C, et al. Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: Twelve-month data from a twenty-four–month phase III randomized radiographic study. Arthritis & Rheumatism. 2013;65(3):559–70.

127. Machado MA de Á, Moura CS de, Guerra SF, Curtis JR, Abrahamowicz M, Bernatsky S. Effectiveness and safety of tofacitinib in rheumatoid arthritis: a cohort study. Arthritis Research & Therapy. 2018 Mar 23;20(1):60.

128. Reed GW, Gerber RA, Shan Y, Takiya L, Dandreo KJ, Gruben D, et al. Real-World Comparative Effectiveness of Tofacitinib and Tumor Necrosis Factor Inhibitors as Monotherapy and Combination Therapy for Treatment of Rheumatoid Arthritis. Rheumatol Ther. 2019 Dec 1;6(4):573–86.

129. Fisher A, Hudson M, Platt RW, Dormuth CR. Tofacitinib Persistence in Patients with Rheumatoid Arthritis: A Retrospective Cohort Study. The Journal of Rheumatology. 2021 Jan 1;48(1):16–24.

130. Fleischmann R, Schiff M, van der Heijde D, Ramos-Remus C, Spindler A, Stanislav M, et al. Baricitinib, Methotrexate, or Combination in Patients With Rheumatoid Arthritis and No or Limited Prior Disease-Modifying Antirheumatic Drug Treatment. Arthritis & rheumatology. 2017;69(3):506–17.

131. Dougados M, van der Heijde D, Chen YC, Greenwald M, Drescher E, Liu J, et al. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. Annals of the rheumatic diseases. 2017;76(1):88–95.

132. Genovese MC, Kremer J, Zamani O, Ludivico C, Krogulec M, Xie L, et al. Baricitinib in Patients with Refractory Rheumatoid Arthritis. The New England journal of medicine. 2016;374(13):1243–52.

133. Miyazaki Y, Nakano K, Nakayamada S, Kubo S, Inoue Y, Fujino Y, et al. Efficacy and safety of tofacitinib versus baricitinib in patients with rheumatoid arthritis in real clinical practice: analyses with propensity score-based inverse probability of treatment weighting. Annals of the Rheumatic Diseases [Internet]. 2021 Apr 7 [cited 2021 Apr 21]; Available from: https://ard.bmj.com/content/early/2021/04/06/annrheumdis-2020-219699

134. Iwamoto N, Sato S, Kurushima S, Michitsuji T, Nishihata S, Okamoto M, et al. Real-world comparative effectiveness and safety of tofacitinib and baricitinib in patients with rheumatoid arthritis. Arthritis Res Ther. 2021 Jul 23;23(1):197.

135. Ebina K, Hirano T, Maeda Y, Yamamoto W, Hashimoto M, Murata K, et al. Drug retention of sarilumab, baricitinib, and tofacitinib in patients with rheumatoid arthritis: the ANSWER cohort study. Clin Rheumatol [Internet]. 2021 Jan 29 [cited 2021 Apr 20]; Available from: https://doi.org/10.1007/s10067-021-05609-7

136. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical Trial Evidence Supporting FDA Approval of Novel Therapeutic Agents, 2005-2012. JAMA. 2014 Jan 22;311(4):368–77.

137. Lexchin J, Graham J, Herder M, Jefferson T, Lemmens T. Regulators, Pivotal Clinical Trials, and Drug Regulation in the Age of COVID-19. Int J Health Serv. 2021 Jan 1;51(1):5–13.

138. Van Luijn JCF, Gribnau FWJ, Leufkens HGM. Availability of comparative trials for the assessment of new medicines in the European Union at the moment of market authorization. British Journal of Clinical Pharmacology. 2007;63(2):159–62.

139. Goldberg NH, Schneeweiss S, Kowal MK, Gagne JJ. Availability of Comparative Efficacy Data at the Time of Drug Approval in the United States. JAMA. 2011 May 4;305(17):1786–9.

140. Zink A, Strangfeld A, Schneider M, Herzer P, Hierse F, Stoyanova-Scholz M, et al. Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: Comparison of patients according to their eligibility for major randomized clinical trials. Arthritis & Rheumatism. 2006;54(11):3399–407.

141. Vashisht P, Sayles H, Cannella AC, Mikuls TR, Michaud K. Generalizability of Patients With Rheumatoid Arthritis in Biologic Agent Clinical Trials. Arthritis Care & Research. 2016;68(10):1478–88.

142. Eichler HG, Koenig F, Arlett P, Enzmann H, Humphreys A, Pétavy F, et al. Are Novel, Nonrandomized Analytic Methods Fit for Decision Making? The Need for Prospective, Controlled, and Transparent Validation. Clinical Pharmacology & Therapeutics. 2020;107(4):773–9.

143. Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. Clinical Pharmacology & Therapeutics. 2020;107(4):817–26.

144. Franklin JM, Platt R, Dreyer NA, London AJ, Simon GE, Watanabe JH, et al. When Can Nonrandomized Studies Support Valid Inference Regarding Effectiveness or Safety of New Medical Treatments? Clinical Pharmacology & Therapeutics. 2022;111(1):108–15.

145. Evans RN, Harris J, Rogers CA, MacGowan A. Emulating the MERINO randomised control trial using data from an observational cohort and trial of rapid diagnostic (BSI-FOO). PLOS ONE. 2022 May 20;17(5):e0268807.

146. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies. Circulation. 2021 Mar 9;143(10):1002–13.

147. Matthews AA, Szummer K, Dahabreh IJ, Lindahl B, Erlinge D, Feychting M, et al. Comparing Effect Estimates in Randomized Trials and Observational Studies From the Same Population: An Application to Percutaneous Coronary Intervention. Journal of the American Heart Association. 2021 Jun 1;10(11):e020357.

148. Admon AJ, Donnelly JP, Casey JD, Janz DR, Russell DW, Joffe AM, et al. Emulating a Novel Clinical Trial Using Existing Observational Data. Predicting Results of the PreVent Study. Annals ATS. 2019 Aug;16(8):998–1007.

149. Kirchgesner J, Desai RJ, Schneeweiss MC, Beaugerie L, Kim SC, Schneeweiss S. Emulation of a randomized controlled trial in ulcerative colitis with US and French claims data: Infliximab with thiopurines compared to infliximab monotherapy. Pharmacoepidemiology and Drug Safety. 2022;31(2):167–75.

150. van Vollenhoven R, Ernestam S, Geborek P, Petersson I, Cöster L, Waltbrand E, et al. Addition of infliximab compared with addition of sulfasalazine and hydroxychloroquine to methotrexate in patients with early rheumatoid arthritis (Swefot trial): 1-year results of a randomised trial. The Lancet. 2009 Aug 8;374(9688):459–66.

151. Lie E, Heijde D van der, Uhlig T, Mikkelsen K, Kalstad S, Kaufmann C, et al. Treatment strategies in patients with rheumatoid arthritis for whom methotrexate monotherapy has failed: data from the NOR-DMARD register. Annals of the Rheumatic Diseases. 2011 Dec 1;70(12):2103–10.

152. Källmark H, Einarsson JT, Nilsson JÅ, Olofsson T, Saxne T, Geborek P, et al. Sustained Remission in Patients With Rheumatoid Arthritis Receiving Triple Therapy Compared to Biologic Therapy: A Swedish Nationwide Register Study. Arthritis & Rheumatology. 2021;73(7):1135–44.

153. Curtis JR, Palmer JL, Reed GW, Greenberg J, Pappas DA, Harrold LR, et al. Real-World Outcomes Associated With Methotrexate, Sulfasalazine, and Hydroxychloroquine Triple Therapy Versus Tumor Necrosis Factor Inhibitor/Methotrexate Combination Therapy in Patients With Rheumatoid Arthritis. Arthritis Care & Research. 2021;73(8):1114–24.

154. Sauer BC, Teng CC, Tang D, Leng J, Curtis JR, Mikuls TR, et al. Persistence With Conventional Triple Therapy Versus a Tumor Necrosis Factor Inhibitor and Methotrexate in US Veterans With Rheumatoid Arthritis. Arthritis Care & Research. 2017;69(3):313–22.

155. Moreland LW, O'Dell JR, Paulus HE, Curtis JR, Bathon JM, St.Clair EW, et al. A randomized comparative effectiveness study of oral triple therapy versus etanercept plus methotrexate in early aggressive rheumatoid arthritis: The Treatment of Early Aggressive Rheumatoid Arthritis trial. Arthritis & Rheumatism. 2012;64(9):2824–35.

156. Saag K. Systemic glucocorticoids in rheumatology. In: Rheumatology [Internet]. Saint Louis, United States: Elsevier; 2010 [cited 2022 Apr 26]. p. 495–503. Available from: http://ebookcentral.proquest.com/lib/ki/detail.action?docID=1430900

157. Hench PS. The reversibility of certain rheumatic and nonrheumatic conditions by the use of cortisone or of the pituitary adrenocorticotropic hormone. Annals of Internal Medicine. 1952 Jan 1;36(1):1–38.

158. Moreland LW, Russell AS, Paulus HE. Management of rheumatoid arthritis: the historical context. The Journal of Rheumatology. 2001 Jun 1;28(6):1431–52.

159. Buttgereit F, Straub RH, Wehling M, Burmester GR diger. Glucocorticoids in the treatment of rheumatic diseases: An update on the mechanisms of action. Arthritis and rheumatism. 2004;50(11):3408–17.

160. Spies CM, Strehl C, van der Goes MC, Bijlsma JWJ, Buttgereit F. Glucocorticoids. Best practice & research Clinical rheumatology. 2011;25(6):891–900.

161. Stahn C, Buttgereit F. Genomic and nongenomic effects of glucocorticoids. Nature Clinical Practice Rheumatology. 2008 Oct;4(10):525–33.

162. Buttgereit F, Saag KG, Cutolo M, Silva JAP da, Bijlsma JWJ. The molecular basis for the effectiveness, toxicity, and resistance to glucocorticoids: focus on the treatment of rheumatoid arthritis. Scandinavian Journal of Rheumatology. 2005 Feb 1;34(1):14–21.

163. Boumpas DT. Glucocorticoid Therapy for Immune-Mediated Diseases: Basic and Clinical Correlates. Ann Intern Med. 1993 Dec 15;119(12):1198.

164. Stuck AE, Minder CE, Frey FJ. Risk of Infectious Complications in Patients Taking Glucocorticosteroids. Rev Infect Dis. 1989 Nov 1;11(6):954–63.

165. Hwang YG, Saag K. The Safety of Low-Dose Glucocorticoids in Rheumatic Diseases: Results from Observational Studies. NIM. 2015;22(1–2):72–82.

166. Dixon WG, Suissa S, Hudson M. The association between systemic glucocorticoid therapy and the risk of infection in patients with rheumatoid arthritis: systematic review and meta-analyses. Arthritis Research & Therapy. 2011 Aug 31;13(4):R139.

167. Askling J, Fored CM, Brandt L, Baecklund E, Bertilsson L, Feltelius N, et al. Time-dependent increase in risk of hospitalisation with infection among Swedish RA patients treated with TNF antagonists. Annals of the Rheumatic Diseases. 2007 Oct 1;66(10):1339–44.

168. Favalli EG, Desiati F, Atzeni F, Sarzi-Puttini P, Caporali R, Pallavicini FB, et al. Serious infections during anti-TNFα treatment in rheumatoid arthritis patients. Autoimmunity Reviews. 2009 Jan 1;8(3):266–73.

169. Strangfeld A, Eveslage M, Schneider M, Bergerhausen HJ, Klopsch T, Zink A, et al. Treatment benefit or survival of the fittest: what drives the time-dependent decrease in serious infection rates under TNF inhibition and what does this imply for the individual patient? Annals of the Rheumatic Diseases. 2011 Nov 1;70(11):1914–20.

170. George MD, Baker JF, Winthrop K, Hsu JY, Wu Q, Chen L, et al. Risk for Serious Infection With Low-Dose Glucocorticoids in Patients With Rheumatoid Arthritis. Ann Intern Med. 2020 Dec;173(11):870–8.

171. Best JH, Kong AM, Lenhart GM, Sarsour K, Stott-Miller M, Hwang Y. Association Between Glucocorticoid Exposure and Healthcare Expenditures for Potential Glucocorticoid-related Adverse Events in Patients with Rheumatoid Arthritis. The Journal of Rheumatology. 2018 Mar 1;45(3):320–8.

172. Fardet L, Petersen I, Nazareth I. Common Infections in Patients Prescribed Systemic Glucocorticoids in Primary Care: A Population-Based Cohort Study. PLOS Medicine. 2016 May 24;13(5):e1002024.

173. Roubille C, Rincheval N, Dougados M, Flipo RM, Daurès JP, Combe B. Seven-year tolerability profile of glucocorticoids use in early rheumatoid arthritis: data from the ESPOIR cohort. Annals of the Rheumatic Diseases. 2017 Nov 1;76(11):1797–802.

174. Schenfeld J, Iles J, Trivedi M, Accortt NA. Dose relationship between oral glucocorticoids and tumor necrosis factor inhibitors and the risk of hospitalized infectious events among patients with rheumatoid arthritis. Rheumatol Int. 2017 Jul 1;37(7):1075–82.

175. Wu J, Keeley A, Mallen C, Morgan AW, Pujades-Rodriguez M. Incidence of infections associated with oral glucocorticoid dose in people diagnosed with polymyalgia rheumatica or giant cell arteritis: a cohort study in England. Canadian Medical Association journal (CMAJ). 2019;191(25):E680–8.

176. Widdifield J, Bernatsky S, Paterson JM, Gunraj N, Thorne JC, Pope J, et al. Serious infections in a population-based cohort of 86,039 seniors with rheumatoid arthritis. Arthritis Care & Research. 2013;65(3):353–61.

177. Dixon WG, Abrahamowicz M, Beauchamp ME, Ray DW, Bernatsky S, Suissa S, et al. Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: a nested case–control analysis. Annals of the Rheumatic Diseases. 2012 Jul 1;71(7):1128–33.

178. Haraoui B, Jovaisas A, Bensen WG, Faraawi R, Kelsall J, Dixit S, et al. Use of corticosteroids in patients with rheumatoid arthritis treated with infliximab: treatment implications based on a real-world Canadian population. RMD Open. 2015;1(1):e000078–e000078.

179. Laugesen K, Ludvigsson JF, Schmidt M, Gissler M, Valdimarsdottir UA, Lunde A, et al. Nordic Health Registry-Based Research: A Review of Health Care Systems and Key Registries. CLEP. 2021 Jul 19;13:533–54.

180. 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 Nov 1;24(11):659–67.

181. Wadström H, Eriksson JK, Neovius M, Askling J, ARTIS Study Group. How good is the coverage and how accurate are exposure data in the Swedish Biologics Register (ARTIS)? Scand J Rheumatol. 2015;44(1):22–8.

182. Eriksson JK, Askling J, Arkema EV. The Swedish Rheumatology Quality Register: optimisation of rheumatic disease assessments using register-enriched data. Clin Exp Rheumatol. 2014 Oct;32(5 Suppl 85):S-147-149.

183. Wettermark B, Hammar N, MichaelFored C, Leimanis A, Olausson PO, Bergman U, et al. The new Swedish Prescribed Drug Register—Opportunities for pharmacoepidemiological research and experience from the first six months. Pharmacoepidemiology and Drug Safety. 2007;16(7):726–35.

184. 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 & Clinical Pharmacology & Toxicology. 2016;119(5):464–9.

185. Rønning M, McTaggart S. Classification systems for drugs and diseases. In: Drug Utilization Research [Internet]. John Wiley & Sons, Ltd; 2016 [cited 2022 Jul 1]. p. 49–57. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118949740.ch5

186. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011 Dec 1;11(1):450.

187. Klassifikationen ICD-10 [Internet]. Socialstyrelsen. [cited 2022 Jul 1]. Available from: https://www.socialstyrelsen.se/statistik-och-data/klassifikationer-och-koder/icd-10/

188. Brooke HL, Talbäck M, Hörnblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish cause of death register. Eur J Epidemiol. 2017 Sep 1;32(9):765–73.

189. Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, et al. Nordic Cancer Registries – an overview of their procedures and data comparability. Acta Oncologica. 2018 Apr 3;57(4):440–55.

190. Ludvigsson JF, Almqvist C, Bonamy AKE, Ljung R, Michaëlsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol. 2016 Feb 1;31(2):125–36.

191. Ludvigsson JF, Svedberg P, Olén O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. Eur J Epidemiol. 2019 Apr 1;34(4):423–37.

192. Rubin DB. Causal Inference Using Potential Outcomes. Journal of the American Statistical Association. 2005 Mar 1;100(469):322–31.

193. Dawid AP. Causal Inference Without Counterfactuals. Journal of the American Statistical Association. 2000;95(450):407–24.

194. Hernán MA. A definition of causal effect for epidemiological research. Journal of Epidemiology and Community Health. 2004;58(4):265–71.

195. Hernán MA, Robins JM. Randomized experiments: Randomization. In: Causal Inference: What if [Internet]. forthcoming. Boca Raton: Chapman & Hall/CRC; 2020 [cited 2022 May 16]. p. 13–6. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

196. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. Epidemiology. 1999;10(1):37–48.

197. Pearl J, Glymour M, Jewell NP. Preliminaries: Statistical and Causal Models: Graphs. In: Causal Inference in Statistics: A Primer. John Wiley & Sons; 2016. p. 24–6.

198. Pearl J, Glymour M, Jewell NP. Graphical Models and Their Applications: Chains and Forks. In: Causal Inference in Statistics: A Primer. John Wiley & Sons; 2016. p. 35–40.

199. Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias—a review of recent developments. International Journal of Epidemiology. 2018 Oct 1;47(5):1714–22.

200. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection bias due to loss to follow up in cohort studies. Epidemiology. 2016 Jan;27(1):91–7.

201. Mongin D, Lauper K, Finckh A, Frisell T, Courvoisier DS. Accounting for missing data caused by drug cessation in observational comparative effectiveness research: a simulation study. Annals of the Rheumatic Diseases. 2022 May 1;81(5):729–36.

202. Pearl J, Glymour M, Jewell NP. Graphical Models and Their Applications: d-separation. In: Causal Inference in Statistics: A Primer. John Wiley & Sons; 2016. p. 45–8.

203. Hernán MA, Robins JM. Variable selection for causal inference: Variables that induce or amplify bias. In: Causal Inference: What if [Internet]. forthcoming. Boca Raton: Chapman & Hall/CRC; 2020 [cited 2022 May 18]. p. 225–8. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

204. Daniel R, Zhang J, Farewell D. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. Biometrical Journal. 2021;63(3):528–57.

205. Greenland S. Invited Commentary: Variable Selection versus Shrinkage in the Control of Multiple Confounders. American Journal of Epidemiology. 2008 Mar 1;167(5):523–9.

206. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019 Mar 1;34(3):211–9.

207. Rubin DB. Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? Statistics in Medicine. 2009;28(9):1420–3.

208. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of Adjusting for Instrumental Variables on Bias and Precision of Effect Estimates. American Journal of Epidemiology. 2011 Dec 1;174(11):1213–22.

209. Pearl J. Invited Commentary: Understanding Bias Amplification. American Journal of Epidemiology. 2011 Dec 1;174(11):1223–7.

210. Lefebvre G, Delaney JAC, Platt RW. Impact of mis-specification of the treatment model on estimates from a marginal structural model. Statistics in Medicine. 2008;27(18):3629–42.

211. Wyss R, Girman CJ, LoCasale RJ, Alan Brookhart M, Stürmer T. Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. Pharmacoepidemiology and drug safety. 2013;22(1):77–85.

212. Gopalakrishnan C, Gagne JJ, Sarpatwari A, Dejene SZ, Dutcher SK, Levin R, et al. Evaluation of Socioeconomic Status Indicators for Confounding Adjustment in

Observational Studies of Medication Use. Clinical Pharmacology & Therapeutics. 2019;105(6):1513–21.

213. Joffe MM, Have TRT, Feldman HI, Kimmel SE. Model Selection, Confounder Control, and Marginal Structural Models. The American Statistician. 2004 Nov 1;58(4):272–9.

214. Robins JM. Association, Causation, and Marginal Structural Models. Synthese. 1999;121(1/2):151–79.

215. Rubin DB, Imbens GW, editors. Unconfounded Treatment Assignment. In: Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction [Internet]. Cambridge: Cambridge University Press; 2015 [cited 2022 May 19]. p. 257–80. Available from: https://www.cambridge.org/core/books/causal-inference-for-statistics-social-and-biomedical-sciences/unconfounded-treatment-assignment/24426D8710217517D4475D83CB517164

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

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

218. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models. Stat Med. 2013 Aug 30;32(19):3388–414.

219. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. Journal of Epidemiology & Community Health. 2006 Jul 1;60(7):578–86.

220. Pearl J, Glymour M, Jewell NP. Preliminaries: Statistical and Causal Models: Probability and Statistics: Independence. In: Causal Inference in Statistics: A Primer. John Wiley & Sons; 2016. p. 10–1.

221. Hernán MA, Robins JM. Estimation of the causal effects of time-varying exposures. In: Longitudinal data analysis. Boca Raton: Chapman & Hall/CRC; 2009. p. 553–99. (Chapman & Hall/CRC handbooks of modern statistical methods).

222. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. Am J Epidemiol. 2008 Sep 15;168(6):656–64.

223. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ. 2019 Oct 23;367:l5657.

224. Li F, Thomas LE, Li F. Addressing Extreme Propensity Scores via the Overlap Weights. Am J Epidemiol. 2019 Jan 1;188(1):250–7.

225. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution—A Simulation Study. American Journal of Epidemiology. 2010 Oct 1;172(7):843–54.

226. Stürmer T, Webster-Clark M, Lund JL, Wyss R, Ellis AR, Lunt M, et al. Propensity Score Weighting and Trimming Strategies for Reducing Variance and Bias of Treatment Effect Estimates: A Simulation Study. American Journal of Epidemiology. 2021 Aug 1;190(8):1659–70.

227. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550–60.

228. Daniel RM, Cousens SN, Stavola BLD, Kenward MG, Sterne J a. C. Methods for dealing with time-dependent confounding. Statistics in Medicine. 2013;32(9):1584–618.

229. Hernán MÁ, 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):561–70.

230. Hernán MA, Robins JM. Selection Bias: How to adjust for selection bias. In: Causal Inference: What if [Internet]. forthcoming. Boca Raton: Chapman & Hall/CRC; 2020 [cited 2022 May 30]. p. 107–11. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

231. Sato T, Matsuyama Y. Marginal Structural Models as a Tool for Standardization. Epidemiology. 2003 Nov;14(6):680–6.

232. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. American Journal of Epidemiology. 2017 Oct 15;186(8):1010–4.

233. Dahabreh IJ, Robins JM, Hernán MA. Benchmarking Observational Methods by Comparing Randomized Trials and Their Emulations. Epidemiology. 2020 Sep;31(5):614–9.

234. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ [Internet]. 2009 Jun 29 [cited 2019 Jan 10];338. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2714692/

235. Rubin DB. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In: Proceedings of the survey research methods section of the American Statistical Association. American Statistical Association Alexandria, VA, USA; 1978. p. 20–34.

236. Buuren S van. Multivariate Missing Data: Fully conditional specification: The MICE algorithm. In: Flexible Imputation of Missing Data [Internet]. 2nd ed. CRC Press; 2018 [cited 2022 May 31]. Available from: https://stefvanbuuren.name/fimd/

237. SAS Documentation: FCS Methods for Data Sets with Arbitrary Missing Patterns [Internet]. [cited 2022 May 31]. Available from: https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_mi_details19.htm

238. White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med. 2009 Jul 10;28(15):1982–98.

239. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011 Feb 20;30(4):377–99.

240. Liu Y, De A. Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. International Journal of Statistics in Medical Research. 2015 Aug 19;4(3):287–95.

241. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used? Stat Methods Med Res. 2019 Jan 1;28(1):3–19.

242. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. Stat Methods Med Res. 2012 Jun 1;21(3):243–56.

243. Mohan K, Pearl J. Graphical Models for Processing Missing Data. Journal of the American Statistical Association. 2021 Apr 3;116(534):1023–37.

244. von Hippel PT. How to Impute Interactions, Squares, and other Transformed Variables. Sociological Methodology. 2009 Aug 1;39(1):265–91.

245. Seaman SR, Bartlett JW, White IR. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. BMC Med Res Methodol. 2012 Apr 10;12(1):46.

246. Buuren S van. Imputation in Practice: Derived Variables: Quadratic relations. In: Flexible Imputation of Missing Data [Internet]. 2nd ed. CRC Press; 2018 [cited 2022 Jun 2]. Available from: https://stefvanbuuren.name/fimd/

247. Leyrat C, Carpenter JR, Bailly S, Williamson EJ. Common Methods for Handling Missing Data in Marginal Structural Models: What Works and Why. American Journal of Epidemiology. 2021 Apr 6;190(4):663–72.

248. Blake HA, Leyrat C, Mansfield KE, Seaman S, Tomlinson LA, Carpenter J, et al. Propensity scores using missingness pattern information: a practical guide. Statistics in Medicine. 2020;39(11):1641–57.

249. Kleinbaum DG, Klein M. Introduction to Survival Analysis. In: Survival Analysis: A Self-Learning Text. 3rd ed. Springer New York; 2012. p. 10.

250. Hernán MA, Robins JM. Causal Survival Analysis: Hazards and Risks. In: Causal Inference: What if [Internet]. forthcoming. Boca Raton: Chapman & Hall/CRC; 2020 [cited 2022 May 16]. p. 209–11. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

251. Allison PD. Estimating Cox Regression Models with PROC PHREG: The Proportional Hazards Model. In: Survival Analysis Using SAS: A Practical Guide. 2nd ed. SAS Institute; 2010. p. 126–7.

252. Hernán MA. The Hazards of Hazard Ratios. Epidemiology. 2010 Jan;21(1):13–5.

253. Kaplan EL, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958 Jun 1;53(282):457–81.

254. Hernán MA, Robins JM. Causal Survival Analysis: From hazards to risks. In: Causal Inference: What if [Internet]. forthcoming. Boca Raton: Chapman & Hall/CRC; 2020 [cited 2022 May 16]. p. 211–4. Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

255. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine. 2020;39(8):1199–236.

256. Waldenlind K, Eriksson JK, Grewin B, Askling J. Validation of the rheumatoid arthritis diagnosis in the Swedish National patient register: a cohort study from Stockholm County. BMC Musculoskeletal Disorders. 2014 Dec 15;15(1):432.

257. Böhm SK. Risk Factors for Diverticulosis, Diverticulitis, Diverticular Perforation, and Bleeding: A Plea for More Subtle History Taking. Viszeralmedizin. 2015 Apr;31(2):84–94.

258. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Annals of the Rheumatic Diseases. 2009 Jun 1;68(6):954–60.

259. Listing J, Gerhold K, Zink A. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. Rheumatology (Oxford). 2013 Jan 1;52(1):53–61.

260. Strehl C, Bijlsma JWJ, Wit M de, Boers M, Caeyers N, Cutolo M, et al. Defining conditions where long-term glucocorticoid treatment has an acceptably low level of harm to facilitate implementation of existing recommendations: viewpoints from an EULAR task force. Annals of the Rheumatic Diseases. 2016 Jun 1;75(6):952–7.

261. Matsumoto Y, Sada K ei, Takano M, Toyota N, Yamanaka R, Sugiyama K, et al. Risk factors for infection in patients with remitted rheumatic diseases treated with glucocorticoids. Acta Med Okayama. 2011 Oct;65(5):329–34.

262. Doran MF, Crowson CS, Pond GR, O'Fallon WM, Gabriel SE. Predictors of infection in rheumatoid arthritis. Arthritis & Rheumatism. 2002;46(9):2294–300.

263. THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [Internet]. L 119/1 May 4, 2016. Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj

264. van Veen EB. Observational health research in Europe: understanding the General Data Protection Regulation and underlying debate. European Journal of Cancer. 2018 Nov 1;104:70–80.

265. Ludvigsson JF, Håberg SE, Knudsen GP, Lafolie P, Zoega H, Sarkkola C, et al. Ethical aspects of registry-based research in the Nordic countries. Clin Epidemiol. 2015 Nov 23;7:491–508.

266. Research material with personal data [Internet]. Swedish National Data Service. 2022. Available from: https://snd.gu.se/en/manage-data/plan/research-material-with-personal-data

267. Sugarman J, Califf RM. Ethics and Regulatory Complexities for Pragmatic Clinical Trials. JAMA. 2014 Jun 18;311(23):2381–2.

268. Riksdagsförvaltningen. Lag (2003:460) om etikprövning av forskning som avser människor [Internet]. 2003:460. Available from: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460

269. Etikprövningsmyndigheten [Internet]. Etikprövningsmyndigheten. [cited 2022 Jul 7]. Available from: https://etikprovningsmyndigheten.se/

270. Council for International Organizations of Medical Sciences. International ethical guidelines for health-related research involving humans [Internet]. World Health Organization; 2016 [cited 2020 Jul 17]. Available from: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

271. Barbulescu A, Delcoigne B, Askling J, Frisell T. Gastrointestinal perforations in patients with rheumatoid arthritis treated with biological disease-modifying antirheumatic drugs in Sweden: a nationwide cohort study. RMD Open. 2020 Jul 1;6(2):e001201.

272. Barbulescu A, Askling J, Chatzidionysiou K, Forsblad-d'Elia H, Kastbom A, Lindström U, et al. Effectiveness of baricitinib and tofacitinib compared with bDMARDs in RA: results from a cohort study using nationwide Swedish register data. Rheumatology. 2022 Feb 3;keac068.

273. Barbulescu A, Askling J, Saevarsdottir S, Kim SC, Frisell T. Combined Conventional Synthetic Disease Modifying Therapy vs. Infliximab for Rheumatoid Arthritis: Emulating a Randomized Trial in Observational Data. Clinical Pharmacology & Therapeutics

[Internet]. 2022 [cited 2022 Jul 27];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpt.2673

274. Hjern F, Wolk A, Håkansson N. Smoking and the risk of diverticular disease in women. British Journal of Surgery. 2011 Jul 1;98(7):997–1002.

275. Rempenault C, Lukas C, Combe B, Herrero A, Pane I, Schaeverbeke T, et al. Risk of diverticulitis and gastrointestinal perforation in rheumatoid arthritis treated with tocilizumab compared to rituximab or abatacept. Rheumatology. 2022 Mar 1;61(3):953–62.

276. Chan FKL, Cryer B, Goldstein JL, Lanas A, Peura DA, Scheiman JM, et al. A novel composite endpoint to evaluate the gastrointestinal (GI) effects of nonsteroidal antiinflammatory drugs through the entire GI tract. J Rheumatol. 2010 Jan;37(1):167–74.

277. Kuhn KA, Manieri NA, Liu TC, Stappenbeck TS. IL-6 Stimulates Intestinal Epithelial Proliferation and Repair after Injury. PLOS ONE. 2014 Dec 5;9(12):e114195.

278. Nakahara H, Song J, Sugimoto M, Hagihara K, Kishimoto T, Yoshizaki K, et al. Anti–interleukin-6 receptor antibody therapy reduces vascular endothelial growth factor production in rheumatoid arthritis. Arthritis & Rheumatism. 2003;48(6):1521–9.

279. Verheul HMW, Pinedo HM. Possible molecular mechanisms involved in the toxicity of angiogenesis inhibition. Nature Reviews Cancer. 2007 Jun;7(6):475–85.

280. Chang XW, Qin Y, Jin Z, Xi TF, Yang X, Lu ZH, et al. Interleukin-6 (IL-6) mediated the increased contraction of distal colon in streptozotocin-induced diabetes in rats via IL-6 receptor pathway. Int J Clin Exp Pathol. 2015 May 1;8(5):4514–24.

281. Zhang L, Hu L, Chen M, Yu B. Exogenous Interleukin-6 Facilitated the Contraction of the Colon in a Depression Rat Model. Dig Dis Sci. 2013 Aug 1;58(8):2187–96.

282. Heise CP. Epidemiology and Pathogenesis of Diverticular Disease. J Gastrointest Surg. 2008 Aug 1;12(8):1309–11.

283. Ito H, Takazoe M, Fukuda Y, Hibi T, Kusugami K, Andoh A, et al. A pilot randomized trial of a human anti-interleukin-6 receptor monoclonal antibody in active Crohn's disease. Gastroenterology. 2004 Apr 1;126(4):989–96.

284. Smolen JS, Aletaha D. Interleukin-6 receptor inhibition with tocilizumab and attainment of disease remission in rheumatoid arthritis: the role of acute-phase reactants. Arthritis Rheum. 2011 Jan;63(1):43–52.

285. Traves PG, Murray B, Campigotto F, Galien R, Meng A, Paolo JAD. JAK selectivity and the implications for clinical inhibition of pharmacodynamic cytokine signalling by filgotinib, upadacitinib, tofacitinib and baricitinib. Annals of the Rheumatic Diseases [Internet]. 2021 Mar 19 [cited 2021 Apr 20]; Available from: https://ard-bmj-com.proxy.kib.ki.se/content/early/2021/03/19/annrheumdis-2020-219012

286. Ytterberg SR, Bhatt DL, Mikuls TR, Koch GG, Fleischmann R, Rivas JL, et al. Cardiovascular and Cancer Risk with Tofacitinib in Rheumatoid Arthritis. N Engl J Med. 2022 Jan 27;386(4):316–26.

287. Boers M, Hartman L, Opris-Belinski D, Bos R, Kok MR, Silva JAD, et al. Low dose, add-on prednisolone in patients with rheumatoid arthritis aged 65+: the pragmatic randomised, double-blind placebo-controlled GLORIA trial. Annals of the Rheumatic Diseases. 2022 Jul 1;81(7):925–36.

288. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, van Denderen JC, et al. Randomised comparison of combined step-down prednisolone, methotrexate and

sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. The Lancet. 1997 Aug 2;350(9074):309–18.

289. Fleischmann R, Genovese MC, Lin Y, St John G, van der Heijde D, Wang S, et al. Long-term safety of sarilumab in rheumatoid arthritis: an integrated analysis with up to 7 years' follow-up. Rheumatology. 2020 Feb 1;59(2):292–302.

290. Heijde DVD, Keystone EC, Curtis JR, Landewé RB, Schiff MH, Khanna D, et al. Timing and Magnitude of Initial Change in Disease Activity Score 28 Predicts the Likelihood of Achieving Low Disease Activity at 1 Year in Rheumatoid Arthritis Patients Treated with Certolizumab Pegol: A Post-hoc Analysis of the RAPID 1 Trial. The Journal of Rheumatology. 2012 Jul 1;39(7):1326–33.

291. Curtis JR, Yang S, Chen L, Park GS, Bitman B, Wang B, et al. Predicting low disease activity and remission using early treatment response to antitumour necrosis factor therapy in patients with rheumatoid arthritis: exploratory analyses from the TEMPO trial. Annals of the Rheumatic Diseases. 2012 Feb 1;71(2):206–12.

292. Aletaha D, Alasti F, Smolen JS. Optimisation of a treat-to-target approach in rheumatoid arthritis: strategies for the 3-month time point. Annals of the Rheumatic Diseases. 2016 Aug 1;75(8):1479–85.

293. Khosrow-Khavar F, Kim SC, Lee H, Lee SB, Desai RJ. Tofacitinib and risk of cardiovascular outcomes: results from the Safety of TofAcitinib in Routine care patients with Rheumatoid Arthritis (STAR-RA) study. Annals of the Rheumatic Diseases. 2022 Jun 1;81(6):798–804.