

Combined Conventional Synthetic Disease Modifying Therapy vs. Infliximab for Rheumatoid Arthritis: Emulating a Randomized Trial in Observational Data

Andrei Barbulescu^{1,*} , Johan Askling^{1,2}, Saedis Saevarsdottir^{1,3}, Seoyoung C. Kim^{4,5}  and Thomas Frisell¹

Observational studies are often considered unreliable for evaluating relative treatment effectiveness, but it has been suggested that following target trial protocols could reduce bias. Using observational data from patients with rheumatoid arthritis (RA) in the Swedish Rheumatology Quality Register (SRQ), between 2006 and 2020, we emulated the protocol of the Swedish Farmaco-therapy trial (SWEFOT) and compared the results. SWEFOT was a pragmatic trial nested in SRQ, between 2002 and 2005, where methotrexate (MTX) insufficient responders were randomized to receive additional infliximab or sulfasalazine (SSZ) + hydroxychloroquine (HCQ). Patients with RA initiating infliximab ($N = 313$) or SSZ + HCQ ($N = 196$) after MTX were identified in SRQ and the Prescribed Drugs Register, mimicking the SWEFOT eligibility criteria. The primary outcome was the proportion of European Alliance of Associations for Rheumatology (EULAR) good responders at 9 months, classifying patients who discontinued treatment as “nonresponders.” Through sensitivity analyses, we assessed the impact of relaxing eligibility criteria. The observed proportions reaching EULAR good response were close to those reported 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 (CI) 1.04–1.86), increasing to 1.48 (95% CI 0.98–2.24) after confounding adjustment, compared to 1.59 (95% CI 1.10–2.30) in SWEFOT. Results remained close to SWEFOT when relaxing eligibility criteria until allowing prior disease-modifying anti-rheumatic drug (DMARD) use which reduced the observed difference between treatments. By applying a prespecified trial emulation protocol to observational clinical registry data, we could replicate the results of SWEFOT, favoring infliximab over SSZ + HCQ combination therapy at 9 months.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Randomized controlled trials comparing the efficacy of adding TNF inhibitors vs. adding conventional synthetic disease-modifying drug combinations to methotrexate (MTX) for the treatment of rheumatoid arthritis (RA) showed mixed results. Observational studies assessing the same question indicated that TNF inhibitors in combination with MTX might be more effective, but they are often considered biased.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ This study emulated the protocol of one of the trials using observational data and compared the results with those of the trial.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ The results of this observational emulation were close to those of its target trial, supporting an increased effectiveness of TNF inhibitors plus MTX compared with conventional synthetic disease modifying drug combinations, within the first year of RA treatment.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Emulating target trials (conducted or theoretical) could reduce bias in observational studies allowing them to credibly answer more questions about treatments in clinical use, at lower costs than possible with trials.

¹Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden; ²Rheumatology, Theme Inflammation and Ageing, Karolinska University Hospital, Stockholm, Sweden; ³Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland; ⁴Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital/Harvard Medical School, Boston, Massachusetts, USA; ⁵Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital/Harvard Medical School, Boston, Massachusetts, USA.
*Correspondence: Andrei Barbulescu (andrei.barbulescu@ki.se)

Received March 23, 2022; accepted May 19, 2022. doi:10.1002/cpt.2673

Randomized controlled trials (RCTs) are the gold standard for evaluating the relative efficacy and safety of drugs, but practical constraints preclude well-powered head-to-head RCTs of all therapeutic options in all clinically relevant patient groups. Whereas “real-world studies,” providing evidence on treatment outcomes in clinical practice are increasingly recognized as necessary complementary information sources for patients not included in clinical trials,^{1,2} they are also criticized for their susceptibility to bias.³ Comparative effectiveness studies using real-world data are the main target of such criticism because in clinical practice treatment decisions are mainly based on perceived predictors of treatment benefit, introducing confounding by indication.⁴ In the treatment of rheumatoid arthritis (RA), for example, switching to a second conventional synthetic disease-modifying drug (csDMARD) or to a combination of csDMARDs is recommended for patients who have failed initial methotrexate (MTX) and who do not have indicators of a poor prognosis, whereas switching to a biological DMARD (bDMARD) is recommended for patients with poor prognosis.^{5,6} To adjust for confounding, measurements of prognosis indicators which influenced treatment decision and are also associated with the outcome (e.g., disease activity at treatment initiation) must be available to the investigator. Nonetheless, even if sufficient data are available to adequately adjust for confounding, more insidious biases, such as selection bias or immortal time bias, may still occur in observational studies, especially when these are based on secondary longitudinal data which allow conditioning on events that happened after treatment initiation.^{7,8} It has been argued that selection and immortal time biases could be avoided if observational studies were designed to explicitly follow target RCT protocols,^{9,10} thus aligning eligibility assessment, treatment assignment, and the start of follow-up. Additionally, harmonizing definitions of eligibility criteria, exposure, outcome, and causal contrast of interest between observational emulations and existing target RCTs facilitates their comparison. Sensitivity analyses aimed at explaining persisting differences in results could then be informed by any remaining differences in key protocol elements.¹¹ Finally, formally adopting and recording pre-analysis study plans increases the transparency and credibility of observational studies.

A sensible approach for testing if observational studies following the trial emulation methodology can produce valid results is to emulate published RCTs and compare the results. One such recent effort is the DUPLICATE initiative where US claims data were used to emulate 10 cardiovascular outcome RCTs, with mixed results.¹² Also in the field of cardiology, a pragmatic trial nested in a Swedish register has been emulated in an observational study using the same register as data-source, and the results were broadly comparable.¹³

A recent review aimed at assessing how many comparative effectiveness observational studies in RA emulated target trials found only one study which explicitly described the target trial in some detail and no study was designed to replicate an existing trial.¹⁰

Thus, in the current study, we aim to evaluate how closely we are able to mimic the protocol and replicate the primary end point results of the Swedish Pharmacotherapy trial (SWEFOT)¹⁴ in an observational study using Swedish register data. SWEFOT was an open-label RCT which recruited patients with early RA from Sweden between October 2002 and December 2005 and treated

them with MTX monotherapy for 3 months, subsequently randomizing nonresponders to either the tumor necrosis factor inhibitor (TNFi) infliximab or a combination of sulfasalazine (SSZ) and hydroxychloroquine (HCQ) added to MTX. At 9 months after randomization a 59% higher European Alliance of Associations for Rheumatology (EULAR) good responder proportion was observed among patients who received additional infliximab. We identified SWEFOT as a good candidate for emulation since it was nested in the Swedish Rheumatology Quality Register (SRQ) which allowed for identification of other patients initiating the same treatments in clinical practice. Thus, we could emulate the trial in a non-overlapping population defined using the same data source, with access to the necessary primary outcome data. The secondary aim of our study was to explore if and how treatment effect estimates changed when relaxing eligibility criteria.

METHODS

As proposed previously,¹⁵ to make the RCT emulation process as transparent and reliable as possible, we first drafted a structured analysis protocol describing in detail how criteria and definitions from SWEFOT would be implemented. Draft definitions of exposure and eligibility criteria were applied to the available data for initial feasibility assessment of sample size and covariate balance, without analyzing the association between treatment and outcome, and were revised aiming for a sample size of 200+ patients per arm (according to the power calculation in SWEFOT).¹⁴ The final version of the protocol was published on [ClinicalTrials.gov](https://www.clinicaltrials.gov/ct2/show/study/NCT05051137) (NCT05051137). The details of the emulation are summarized in [Table S1](#) and the study design in [Figure 1](#).

Data sources and population

We used patient-level data from the SRQ¹⁶ linked via the personal identity number of each patient to several Swedish national registers¹⁷—the Prescribed Drugs Register (PDR), the National Patients Register (NPR), the Cancer Register—and demographic data from Statistics Sweden. The earliest date of data availability in all sources was July 2005 (start of PDR). The SRQ has good coverage for bDMARDs, containing the dates of decision to start and stop treatment.

We identified patients with a primary diagnosis of RA which debuted after July 2005, as recorded in the SRQ, in order to have complete prescription data for all patients from RA debut onward and to reduce overlap with the SWEFOT population. We then identified study treatment initiations (i.e., baseline) and excluded patients not fulfilling the emulation eligibility criteria at baseline, patients who, due to migrations, lacked full 5-year data before baseline, and patients with < 9 months of follow-up to January 2021 (end of study period). The main eligibility criteria were: a Disease Activity Score (DAS)28-ESR (disease activity score using 28 joints and erythrocyte sedimentation rate) > 3.2, use of MTX for > 30 days before baseline, and no use of other DMARDs before baseline (details in [Table S1](#)).

Treatment strategies

Infliximab arm. The baseline in this arm was the first decision to initiate infliximab, as recorded in the SRQ. Protocol treatment was ended at the recorded infliximab treatment stop or at the initiation of a DMARD (other than infliximab or MTX), whichever came first. Consecutive infliximab treatment episodes on the same or different products, with less than 90 days in between, were merged. Mimicking SWEFOT, patients who stopped infliximab for safety reasons, and initiated etanercept within 90 days were considered to stay on treatment until the end of etanercept treatment.

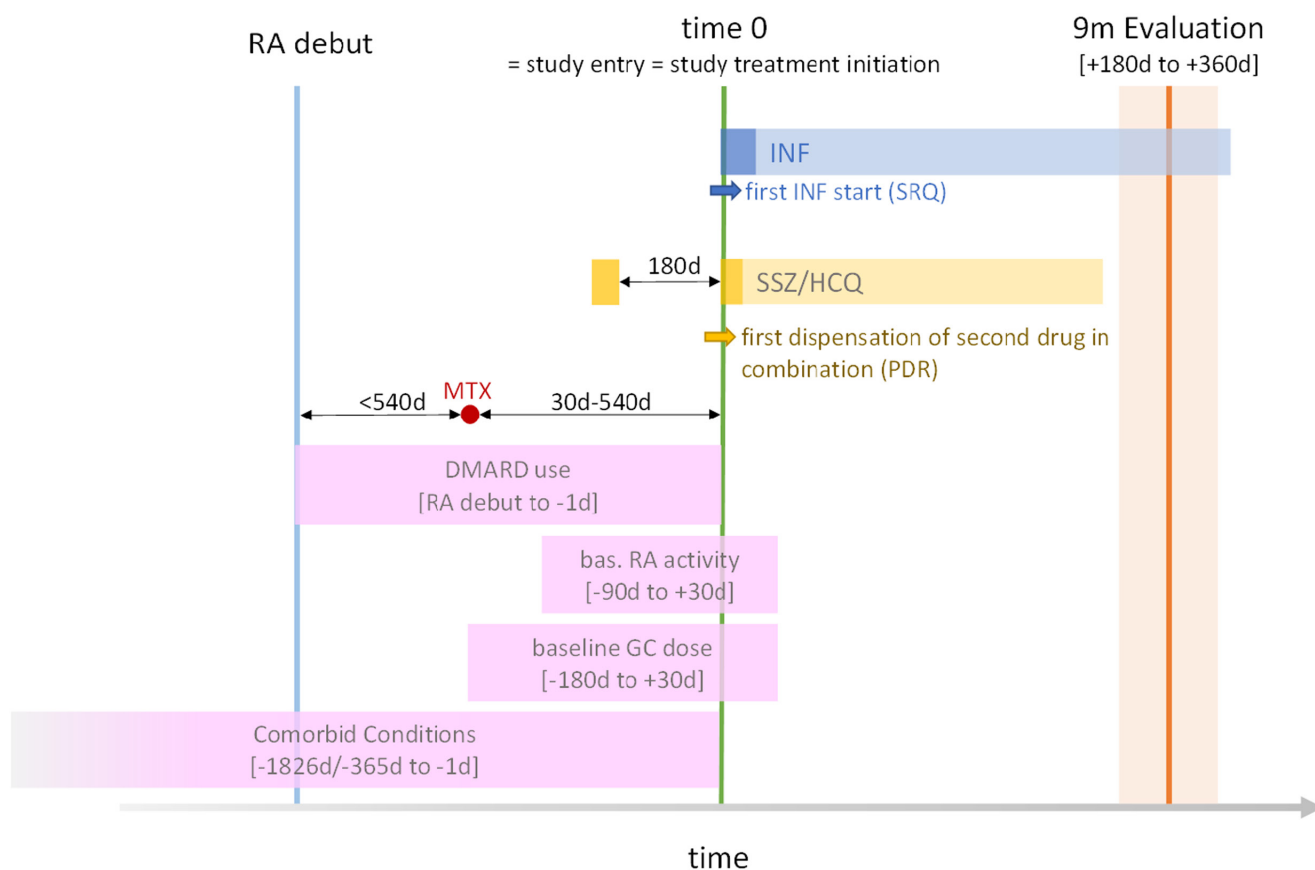


Figure 1 Study design. The time unit is days and time-windows are measured relative to baseline (i.e., time 0); for example, the time window for the 9-month outcome evaluation (+180 days to +360 days) represents a time window between 180 and 360 days after baseline. DMARD, disease-modifying anti-rheumatic drug; GC, glucocorticoids; HCQ, hydroxychloroquine; INF, infliximab; MTX, methotrexate; PDR, Prescribed Drugs Register; RA, rheumatoid arthritis; SRQ, Swedish Rheumatology Quality Register; SSZ, sulfasalazine. [Colour figure can be viewed at wileyonlinelibrary.com]

Sulfasalazine and hydroxychloroquine arm. Prescriptions for SSZ (ATC = A07EC01) and HCQ (ATC = P01BA02) were identified in the PDR. We included patients who initiated both drugs no more than 180 days apart, and who collected at least one prescription for the first drug after the initiation of the second drug. The decision to allow a sequential initiation of the two drugs within this time window encountered in clinical practice was based on feasibility analyses with the goal of achieving sample sizes of ~ 200 patients per arm. The dispensation date for the second drug in the combination was considered the baseline in this arm (i.e., the date when the patient started being treated with both drugs). The end of each treatment in the combination was estimated at the first gap in the sequence of prescriptions. To identify such gaps, the duration of each prescription was first calculated using the dispensed quantity and assuming a daily dose of 2,000 mg/day for SSZ and 200 mg/day for HCQ. Then, a gap in the sequence of prescriptions was defined as not collecting a subsequent prescription within 90 days after the current dispensed duration. If the patient stopped both SSZ and HCQ and initiated cyclosporin A within 90 days, the end of protocol treatment was extended to the end of the cyclosporin A treatment. Protocol treatment was ended at the initiation of another DMARD.

Outcomes

Similar to SWEFOT, outcomes were evaluated at 9 months after baseline using data from the SRQ (rheumatology visits). The primary outcome measure was the achievement of a EULAR good (vs. moderate or no) response defined as a DAS28-ESR ≤ 3.2 at evaluation and a decrease

in DAS28-ESR larger than 1.2 units at evaluation compared with baseline. Patients who stopped the protocol treatment before the 9-month evaluation were classified as “nonresponders.” Patients who discontinued follow-up due to death or emigration were excluded from the analysis.

A secondary outcome was a EULAR good or moderate (vs. no) response defined as: DAS28-ESR ≤ 5.1 at evaluation and a decrease > 0.6 units compared with baseline, or a DAS28-ESR > 5.1 at evaluation and a decrease > 1.2 units compared with baseline, again with nonresponder imputation for those who stopped treatment.

The baseline DAS28-ESR was measured within a window spanning from 90 days before to 30 days after baseline and end point DAS28-ESR was measured between 180 to 360 days after baseline. The closest measurement to baseline and day 270 (end point), respectively, was picked. If several measurements were recorded at the same distance from baseline or evaluation, respectively, the average value was calculated.

Covariates

Several baseline patient characteristics that were considered outcome predictors were specified in the protocol for inclusion in the propensity score (PS): sex, age, country of origin, year of treatment start, RA duration at treatment start, rheumatoid factor positivity, disease activity: DAS28-ESR, Clinical Disease Activity Index (CDAI), counts of swollen and tender joints, pain assessment (measured on a visual analogue scale) and Health Assessment Questionnaire Disability Index (HAQ-DI), use of MTX, nonsteroidal anti-inflammatory drugs (NSAIDs), and of glucocorticoids, history of: cancer, diabetes, acute coronary syndrome, stroke, venous thromboembolism, peripheral

vascular disease, obstructive respiratory disease, anemia, psoriasis, neuropathies, pain syndromes, osteoporosis, depression or anxiety, joint surgery, retinopathy, and hospitalized infections, the number of days spent in the hospital within 5 years before treatment start, and smoking status. Definitions of all measured covariates are presented in **Table S2**.

Statistical analysis

We estimated crude and adjusted responder proportions and their ratios using generalized linear models with log link function, binomial outcome distribution, and robust standard error estimation.

Confounding was adjusted for by stabilized inverse-probability of treatment weighting (IPTW), where the conditional probability of receiving infliximab (vs. SSZ + HCQ) as a function of baseline patient characteristics (PS) was modeled using logistic regression. Stabilized weights were calculated for each observation as the marginal probability of receiving the treatment actually received divided by the corresponding conditional probability.¹⁸ Balance of baseline patient characteristics between treatment arms was assessed before and after weighting using standardized mean differences.¹⁹

Missing outcome and covariate values were addressed with multiple imputation using fully conditional specification, 30 imputations and 25 burn-in iterations. Each missing variable was imputed as a function of all other variables used in the analysis (plus their transformations), including exposure and outcome, and an “on-treatment at evaluation” indicator. Predictive mean matching was used for quantitative variables and logistic regression for categorical variables. IPTW estimation and outcome modeling were conducted in each imputation, pooling final estimates using Rubin’s rules.²⁰

Sensitivity analyses

A series of analyses tested how altering the pre-recorded analysis plan would change the results compared with the protocol analysis.

The first analysis was identical to the main analysis but included several additional covariates which were excluded from the protocol due to their low prevalence (congestive heart failure, renal or liver diseases, interstitial respiratory disease, and inflammatory bowel disease), and did not allow treatment continuation on etanercept or cyclosporin A.

The next analysis included only observations with complete baseline and outcome data where treatment responses of all patients included at baseline were measured around the 9-month end point regardless of treatment discontinuation/switch (without nonresponder imputation).

Subsequently, eligibility criteria were relaxed one by one: first, a baseline DAS28-ESR ≤ 3.2 was allowed; second, no restriction on the timing of MTX initiation relative to RA debut and baseline was imposed (but MTX had to be initiated between RA debut and baseline); third, the use of other non-MTX DMARDs before baseline was permitted (but not after baseline). Baseline DAS28-ESR, RA duration, and use of DMARDs before baseline were adjusted for. Trimming of observations based on the PS, as suggested by Stürmer *et al.*,²¹ was used to improve comparability of treatment arms and avoid extreme IPTW after allowing use of DMARDs before baseline. The PS was estimated separately in each imputation and observations with PS lower than the 5th percentile in the infliximab arm or higher than the 95th percentile in the SSZ + HCQ arm were excluded.

Finally, the SSZ + HCQ arm was restricted to patients who initiated the two drugs simultaneously to avoid potential selection and immortal time biases built in the protocol definition.⁷ Selection bias occurs when initiation of the first drug in the combination is allowed before baseline only in the SSZ + HCQ arm. Patients in this arm who respect inclusion criteria (i.e., high enough baseline DAS28-ESR) after an initial treatment with the first drug may have a higher prevalence of other common causes of baseline DAS28-ESR and later treatment response compared with those in the infliximab arm. Immortal time bias occurs when a prescription for the first drug is required after baseline (to ensure addition of the second drug over the first and not just a treatment switch).

This decreases the probability that patients in the SSZ + HCQ arm stop protocol treatment (thus be classified as nonresponders) because protocol treatment continues until both drugs are stopped/replaced. Time when the outcome (i.e., nonresponse) cannot occur is deemed “immortal.” No similar “immortal time” is introduced in the infliximab arm, biasing results in favor of SSZ + HCQ. There was no need to ensure continued treatment with one drug after baseline when both drugs were initiated at baseline, thus avoiding immortal time bias.

Ethics

The study was approved by the Ethical Review Board in Stockholm (DNR: 2015/1844-31/2, amended 2016/1986-32, 2017/2473-32, 2020-01756). In accordance with Swedish law, participant consent is not necessary for register-based studies with pseudo-anonymized data.

RESULTS

After screening 57,288 patients with RA identified in the SRQ, and applying the eligibility criteria we included 509 patients who initiated either infliximab ($N = 313$) or SSZ + HCQ ($N = 196$) between January 2006 and April 2020. A diagram of how the study population was identified and progressed through follow-up is presented in **Figure 2**.

Baseline characteristics

Baseline characteristics of patients in the emulation and in SWEFOT are presented in **Table 1**. The RA duration was similarly distributed in both emulation arms, with 75% of patients at less than 1.4 years since the RA debut. Patients who initiated SSZ + HCQ were marginally older, more likely women, and of Swedish origin than the ones initiating infliximab and the proportion of women were overall lower in the emulation than in SWEFOT. In the emulation population, patients who initiated infliximab had a higher disease activity at treatment initiation than patients who initiated SSZ + HCQ. The baseline use of glucocorticoids and NSAIDs was balanced between emulation arms. The use of NSAIDs was more common in SWEFOT. Patients initiating infliximab were less likely to have suffered from cancer or serious infections and more likely to have suffered from diabetes.

The proportions of missing data are presented in **Table S3**. Missingness in baseline RA disease activity, disability, and rheumatoid factor positivity was under 5% in both groups. Smoking status data was missing in high proportion in both groups (60%–70%).

There was substantial overlap in PS distribution between treatment groups (**Figure S1**). Stabilized IPTW ranged from 0.6 to 8.1 in the infliximab arm and between 0.4 and 28.1 in the SSZ + HCQ arm. Weights higher than 10 were observed for only 3 patients in the SSZ + HCQ arm and were truncated to 10. The distribution of IPTW before truncation over all imputations is presented in **Figure S2**. The balance in baseline characteristics before and after weighting is displayed in **Figures S3** and **S4** and in **Table S4**. Baseline characteristics were well-balanced by the truncated IPTW, with most standardized differences lower than 0.1.

Primary and secondary outcome

No patient was lost during follow-up due to emigration. One patient in the SSZ + HCQ arm was lost due to death (and excluded from the analysis). In the infliximab arm, 81 of 313 (26%;

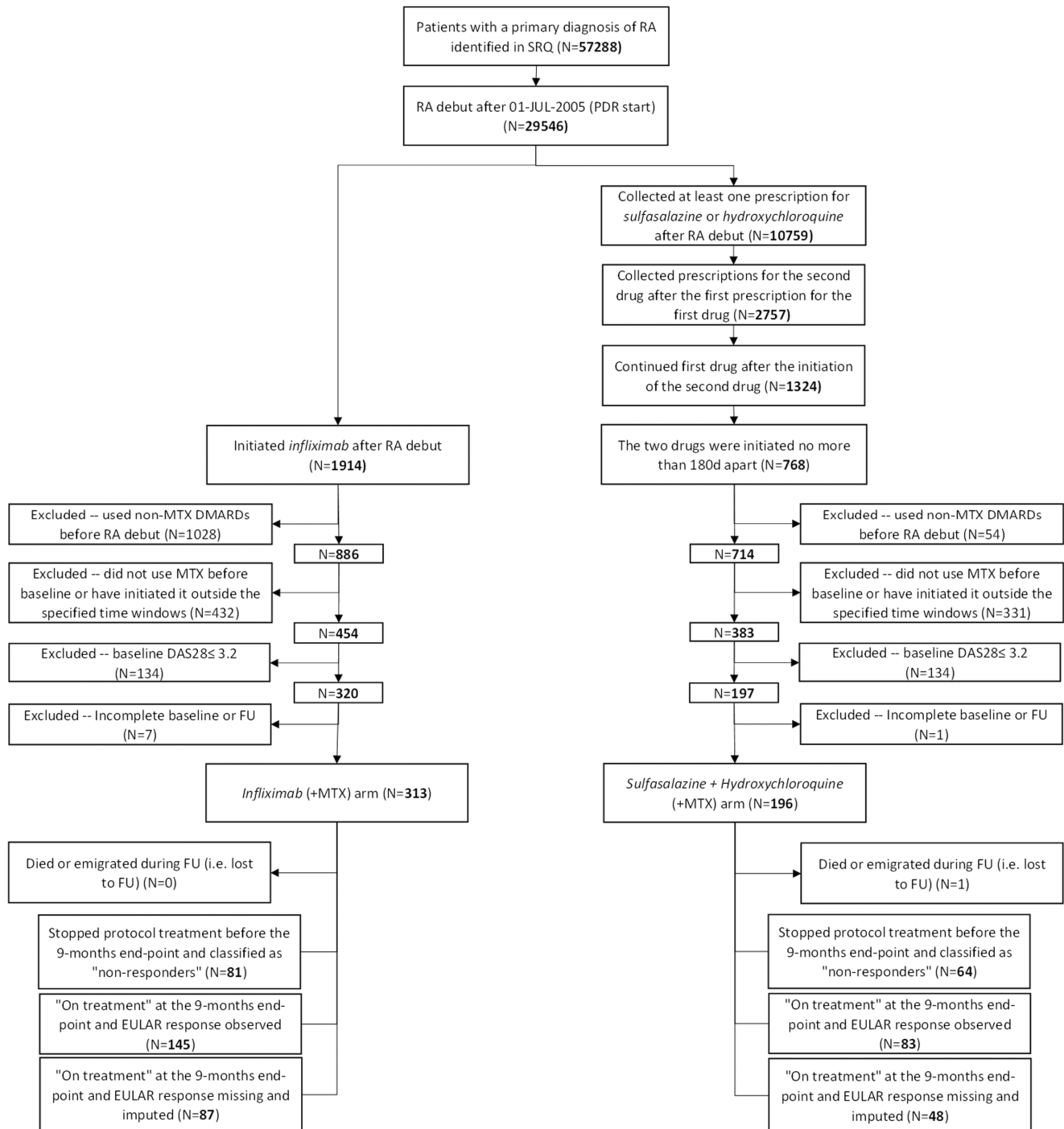


Figure 2 Identification of the emulation population and progression through follow-up. DAS28, Disease Activity Score (28 joints); DMARD, disease-modifying anti-rheumatic drugs; EULAR, European Alliance of Associations for Rheumatology; FU, follow-up; MTX, methotrexate; PDR, Prescribed Drugs Register; RA, rheumatoid arthritis; SRQ, Swedish Rheumatology Quality Register.

Figure 2) patients stopped protocol treatment before the 9-month evaluation and were classified as “nonresponders”; the corresponding number for SSZ + HCQ was 64 of 196 (33%; **Figure 2**). **Figure 2** and **Table S3** show that, among patients on treatment at 9-month evaluation, treatment response was missing for 87 of 313 (28%) subjects in the infliximab arm and 48 of 196 (25%) subjects in the SSZ + HCQ arm.

Table 2 summarizes the proportions of EULAR “good” and “good or moderate” responders after multiple imputation of missing responses. In the infliximab arm, we observed 39% “good” EULAR responders, the same proportion as in SWEFOT. In the SSZ + HCQ arm, the crude proportion of responders was higher than in SWEFOT, at 28% vs. 25%. The crude response ratio was numerically lower but with overlapping confidence intervals (CIs)

Table 1 Patients characteristics at baseline in the observational emulation and SWEFOT

Characteristic	Observational emulation		SWEFOT ^a	
	Infliximab	SSZ+HCQ	Infliximab	SSZ+HCQ
N	313	196	124	127
Start year	2012 (2010–2015)	2010 (2007–2015)	2004 (2003–2005)	2004 (2004–2005)
Age, years	56.0 (46.0–65.0)	57.5 (47.0–66.5)	54.0 (42.0–60.5)	56.0 (46–0–63.0)
Female	211 (67.4)	139 (70.9)	94 (75.8)	98 (77.2)
Origin (Swedish)	260 (83.1)	169 (86.2)	110 (88.7)	116 (91.3)
RA duration, years	1.0 (0.7–1.4)	1.0 (0.7–1.4)	0.8 (0.6–1.0)	0.8 (0.6–1.0)
Rheumatoid factor positivity	225 (72.8)	124 (66.0)	89 (72.4)	86 (68.8)
Disease activity score (DAS28ESR)	5.1 (4.4–5.9)	4.6 (4.0–5.2)	4.9 (4.2–5.5)	4.7 (3.9–5.5)
HAQ disability score	1.0 (0.6–1.5)	0.9 (0.5–1.1)	0.9 (0.5–1.3)	0.9 (0.6–1.4)
Swollen joints count	7.0 (4.0–11.0)	5.0 (2.5–7.0)	6.0 (3.0–10.0)	6.0 (3.0–9.0)
Tender joints count	7.0 (4.0–11.0)	5.0 (3.0–8.0)	6.0 (4.0–10.0)	5.0 (2.0–10.0)
Global health (patient assessment)	62.0 (44.0–75.0)	46.0 (26.5–64.5)	49.5 (34.5–67.5)	40.0 (22.0–64.0)
Pain (VAS scale)	63.0 (43.0–74.5)	43.5 (29.0–62.0)	47.0 (30.0–63.0)	38.0 (21.0–60.0)
Clinical Disease Activity Index	25.5 (19.5–34.5)	19.5 (15.0–24.0)	23.0 (17.0–28.8)	19.0 (13.0–28.5)
Glucocorticoid dose (mg/day)	8.8 (5.4–10.1)	8.8 (4.2–10.0)	–	–
Methotrexate comedication	297 (94.9)	179 (91.3)	122 (98.4)	126 (99.2)
NSAID comedication	104 (33.1)	72 (36.7)	79 (63.7)	67 (52.8)
Joint surgery	23 (7.3)	14 (7.1)	–	–
Cancer	8 (2.6)	14 (7.1)	–	–
Diabetes	20 (6.4)	4 (2.0)	–	–
Hospitalized infections	4 (1.3)	4 (2.0)	–	–
Acute coronary syndrome	1 (0.3)	4 (2.0)	–	–
Congestive heart failure	1 (0.3)	0 (0.0)	–	–
Stroke	6 (1.9)	2 (1.0)	–	–
Venous thromboembolism	5 (1.6)	3 (1.5)	–	–
Peripheral vascular diseases	4 (1.3)	2 (1.0)	–	–
Obstruct lung diseases	4 (1.3)	5 (2.6)	–	–
Interstitial lung diseases	1 (0.3)	0 (0.0)	–	–
Liver diseases	0 (0.0)	1 (0.5)	–	–
Renal diseases	3 (1.0)	0 (0.0)	–	–
Neuropathies	3 (1.0)	2 (1.0)	–	–
Pain syndromes	13 (4.2)	6 (3.1)	–	–
Mood disorders	15 (4.8)	10 (5.1)	–	–
Anemias	7 (2.2)	3 (1.5)	–	–
Retinopathies	9 (2.9)	5 (2.6)	–	–
Smoking status (never)	48 (39.0)	28 (44.4)	–	–
Smoking status (past)	53 (43.1)	29 (46.0)	–	–
Smoking status (current)	22 (17.9)	6 (9.5)	–	–
Days in hospital (last 5years)	0.0 (0.0–4.0)	0.0 (0.0–4.5)	0.0 (0.0–3.5)	0.0 (0.0–3.0)

HAQ, Health Assessment Questionnaire; HCQ, hydroxychloroquine; NSAID, nonsteroidal anti-inflammatory drug; RA, rheumatoid arthritis; SSZ, sulfasalazine; SWEFOT, Swedish Pharmacotherapy trial; VAS, visual analogue scale.

Median (interquartile range) presented for quantitative variables. Number (percentage) presented for binary indicators.

^aBaseline characteristics in SWEFOT were measured at randomization using the available registry data (data from the Prescribed Drugs Register and National Patients Register was not available or incomplete). No data was available for four patients in the SWEFOT infliximab arm and three patients in the SSZ+HCQ arm.

Table 2 Proportions of “good” and “good or moderate” EULAR responders in the protocol analysis

	Infliximab	SSZ + HCQ	
Emulation <i>N</i>			
Total	313	195	
With non-missing outcome data	226	147	
SWEFOT ^a <i>N</i>	128	130	
EULAR good response proportions at 9 months (95% CI)			Ratio (95% CI)
Crude	39% (33–46%)	28% (22–37%)	1.39 (1.04–1.86)
IPTW adjusted ^b	40% (34–47%)	27% (19–39%)	1.48 (0.98–2.24)
SWEFOT ^a	39%	25%	1.59 (1.10–2.30)
EULAR good or moderate response proportions at 9 months (95% CI)			Ratio (95% CI)
Crude	60% (54–66%)	47% (40–56%)	1.26 (1.04–1.52)
IPTW adjusted ^b	61% (55–68%)	48% (38–60%)	1.27 (0.99–1.63)
SWEFOT ^a	60%	49%	1.22 (0.98–1.53)

CI, confidence interval; EULAR, European Alliance of Associations for Rheumatology; HCQ, hydroxychloroquine; IPTW, inverse probability of treatment weighting; SSZ, sulfasalazine; SWEFOT, Swedish Pharmacotherapy trial.

^aFrom van Vollenhoven *et al.*¹⁴

^bAdjusted for: age, sex, country of birth, smoking status (63.3% missing), RA duration, rheumatoid factor (2.4% missing), year of treatment start, disease activity measures (5.9% missing), comedication, and comorbidity. Missing data was accounted for by multiple imputation.

compared with SWEFOT, 1.39 (95% CI: 1.04–1.86) vs. 1.59 (95% CI: 1.10–2.30). After inverse probability weighting adjustment, the ratio rose to 1.48 (95% CI: 0.98–2.24). The proportions of “good or moderate” responders were similar in our emulation and SWEFOT with minimal changes after confounding adjustment.

Sensitivity analyses

Table 3 presents the results of sensitivity analyses.

Adding disease history covariates and not allowing treatment continuation on etanercept and cyclosporin A, respectively (base case), left the results unchanged.

Complete case analysis without nonresponder imputation. The crude response rates were higher than in the main analysis, with 47% responders in the infliximab arm (vs. 39%) and 40% in the SSZ + HCQ arm (vs. 28%), which corresponds to a crude response ratio of 1.19 (95% CI: 0.91–1.55). After confounding adjustment, the response ratio rose to 1.23 (95% CI: 0.85–1.83), which was lower than 1.48 in the main analysis.

Eligibility criteria relaxation. Relaxing the eligibility criteria unsurprisingly increased sample size but also lowered the contrast between infliximab and SSZ + HCQ especially after allowing a longer RA duration and prior use of non-MTX DMARDs. The lowest contrast was observed after allowing the use of non-MTX DMARDs before baseline—adjusted response ratio 1.19 (95% CI: 0.86–1.66) and 1.23 (95% CI: 0.88–1.73) after initial PS-based trimming.

Alternative SSZ + HCQ definition. Allowing only simultaneous initiation of SSZ and HCQ produced an adjusted response ratio of 1.47 (95% CI: 0.91–2.37), very similar to the one obtained in the main analysis.

DISCUSSION

Using the same source population (SRQ) as in SWEFOT and striving to harmonize eligibility criteria, treatment definitions, outcome measurements, and causal contrasts between the non-randomized emulation and the trial, we obtained similar results in both primary and secondary end points despite remaining differences between protocols. These results indicate an increased effectiveness of infliximab compared to SSZ + HCQ when initiated after unsuccessful treatment with MTX monotherapy.

Several other observational studies also found combinations of TNFi^{22,23} or other bDMARDs²⁴ with MTX superior to the SSZ + HCQ + MTX at 6 months and 1 year. One of the studies used data from the SRQ within a time period overlapping with SWEFOT, but substituted infliximab for a wider bDMARD group, defined different outcomes and identified the SSZ + HCQ + MTX group using a different algorithm and data source than in our study.²⁴ Another study of Norwegian patients compared EULAR good response proportions at 6 months and found 40% in the TNFi + MTX arm and 20% in the SSZ + HCQ + MTX arm. An American study was able to stratify the analysis into bDMARDs-exposed and bDMARDs-naïve patients and observed reduced proportions of responders in both arms and a lower contrast among the biologics-exposed.²² Due to very few patients who had used any non-MTX DMARD before SSZ + HCQ, we were unable to replicate this analysis. However, the lower contrast in the analysis allowing DMARD use before baseline could indicate such effect modification. Alternatively, the low contrast in this analysis could be explained by residual confounding given the extreme IPTW (up to 400) which had to be truncated, but ameliorating comparability and extreme weights by initial PS trimming yielded similar results. The possibility of obtaining extreme weights may be seen as a downside of IPTW because they lower the precision of estimation. However, extreme weights can alert the analyst to the presence of

Table 3 Proportions of “good” EULAR responders in sensitivity analyses

		N	Crude response proportions (95% CI)	Crude response ratio (95% CI)	IPTW adjusted ^a response ratio (95% CI)
Base case ^b	Infliximab	313	40% (34%–46%)	1.37 (1.02–1.83)	1.45 (0.98–2.15)
	SSZ+HCQ	195	29% (23%–37%)	Ref	Ref
Complete cases without nonresponder imputation ^c	Infliximab	187	47% (40%–55%)	1.19 (0.91–1.55)	1.23 (0.85–1.83)
	SSZ+HCQ	121	40% (32%–49%)	Ref	Ref
Relaxed eligibility criteria					
+ No DAS28 restriction	Infliximab	438	37% (31%–43%)	1.48 (1.10–1.98)	1.39 (0.96–2.00)
	SSZ+HCQ	366	25% (20%–32%)	Ref	Ref
+ No RA duration restriction	Infliximab	800	37% (33%–41%)	1.35 (1.08–1.70)	1.24 (0.96–1.60)
	SSZ+HCQ	546	27% (22%–33%)	Ref	Ref
+ Prior DMARDs allowed ^d	Infliximab	1,465	30% (28%–34%)	1.16 (0.92–1.46)	1.19 (0.86–1.66)
	SSZ+HCQ	579	26% (21%–32%)	Ref	Ref
+ Prior DMARDs allowed + PS trimming ^e	Infliximab	1,114	33% (29%–36%)	1.18 (0.92–1.53)	1.23 (0.88–1.73)
	SSZ+HCQ	497	28% (22%–34%)	Ref	Ref
Alternative definition of the SSZ+HCQ arm					
Simultaneous SSZ+HCQ initiation	Infliximab	313	39% (33%–46%)	1.30 (0.96–1.77)	1.47 (0.91–2.37)
	SSZ+HCQ	166	30% (23%–39%)	Ref	Ref

CI, confidence interval; DAS28, disease activity score (28 joints); DMARD, disease-modifying anti-rheumatic drug; EULAR, European Alliance of Associations for Rheumatology; GC, glucocorticoid; HCQ, hydroxychloroquine; IPTW, inverse probability of treatment weighting; MTX, methotrexate; PS, propensity score; RA, rheumatoid arthritis; SSZ, sulfasalazine.

^aAdjusted for age, sex, country of birth, smoking status, RA duration, rheumatoid factor, year of treatment start, disease activity measures, comorbidity (congestive heart failure, inflammatory bowel disease, renal and liver disease, and interstitial lung disease added to protocol analysis adjustment set). ^bBase case: all inclusion criteria are in place (as in the protocol analysis) but treatment switch to etanercept (infliximab) or cyclosporin A (SSZ+HCQ) is not allowed. ^cOnly observations without missing outcome or covariate data; smoking excluded from the adjustment set due to >60% missingness; actual EULAR response analyzed regardless of treatment discontinuation (without nonresponder imputation). ^dUse of other DMARDs than MTX allowed before baseline and adjusted for in the analysis. ^eUse of other DMARDs than MTX allowed before baseline and adjusted for in the analysis + exclusion of observations with PS lower than the 5th percentile of PS among INF or higher than the 95th percentile of PS among SSZ + HCQ; separate PS trimming in each imputation, therefore numbers of observations in each arm were averaged over imputations.

population strata where the use of one treatment is unlikely. Such strata hold little information about the outcome under the unlikely treatment (therefore the imprecision in effect estimation) and may harbor unmeasured confounding, thus they could be trimmed away.²¹

Not all studies agree that TNFi+MTX are superior to SSZ+HCQ+MTX. Two RCTs did not find etanercept+MTX superior to SSZ+HCQ+MTX.^{25,26} It could be argued that these results are not directly comparable to SWEFOT because etanercept was used instead of infliximab. Besides, the two trials were blinded whereas SWEFOT was open-label (as were of course the observational studies). Whereas limiting the interpretability of results in terms of “treatment efficacy,” the open-label design was a reason for choosing SWEFOT as the emulation target in our study because blinding is impossible in observational data, and cannot be emulated by statistical means.⁹ Being able to use similarly collected data from the same context (Swedish rheumatology) but from a non-overlapping calendar period allowed us to emulate SWEFOT in most respects. Emulating a trial using real-world data from different settings would impose additional challenges for the emulation, requiring creative solutions to overcome them (for example, treatment response measured via proxies, such as an increase in the dose of glucocorticoids in the absence of RA disease activity measurements).

Nonetheless, we could not emulate the SWEFOT protocol perfectly. Most difficult was identifying a sufficiently large SSZ+HCQ

arm, as this has gradually become a less common treatment choice in clinical practice. In order to increase the size of our study population, we allowed the initiation of the drugs in sequence, as long as they were not initiated more than 180 days apart and as long as the first drug was not stopped before initiating the second. This treatment definition may introduce selection bias by initiating a treatment component before baseline and immortal time bias by demanding treatment continuation on one component during follow-up. Restricting to patients who initiated both drugs simultaneously produced similar effect estimates as in the main analysis since it turned out that 67% of the 196 patients initiating SSZ+HCQ in the main analysis initiated the drugs at the same time and 50% stopped SSZ+HCQ due to initiation of another DMARD regardless of continuation of the first drug in combination.

Other compromises were made in the interest of sample size. First, the protocol analysis was not restricted to treatments initiated within the first year of RA debut. However, the median RA duration was 1 year in both arms. Second, continuation of MTX treatment to baseline was not imposed. Nonetheless, more than 90% of patients in both arms were on treatment with MTX at baseline. Finally, no patients were excluded based on study treatments contraindications (such as congestive heart failure, chronic infections, porphyria, or retinopathies) or baseline glucocorticoid use. Instead, these covariates were measured and considered for adjustment.

Despite relaxing eligibility criteria, an important limitation of our study (as of SWEFOT) remains the small sample size. Due to small sample size, the lower confidence limit reached 0.98 in our IPTW adjusted main results and would be interpreted as “no statistically significant difference between treatments” based on a binary hypothesis test, disagreeing with SWEFOT. Additionally, outcome data was incomplete due to the unstructured follow-up in real clinical practice. We used a window spanning from 90 days before to 90 days after the 9-month end point to capture as much DAS28-ESR data as possible but we still had to impute outcome data for ~25% of observations, adding uncertainty to our estimates.

Analyzing only observations with complete data and using the observed 9-month treatment response regardless of protocol treatment discontinuation produced higher response proportions and a lower response ratio. Such results are expected in the absence of bias because patients who discontinued due to ineffectiveness would switch to alternative treatments which may improve response.

The foremost strength of our emulation is the access to the same source population as the target trial and to disease activity data which is arguably the most important confounding factor. In addition, adjusting for confounding via IPTW produces effect estimates averaged (marginalized) over the entire study population, thus comparable to RCT estimates. Our results were not sensitive to minor deviations from the protocol and the impact of confounding was modest after applying broad exclusion criteria.

To conclude, we observed a 48% increased EULAR good response proportion for infliximab vs. SSZ + HCQ compared with 59% increase observed in SWEFOT, with largely overlapping CIs. The fact that this emulation managed to obtain similar results to its target trial is not a guarantee for observational studies in general. Nevertheless, previous observational studies comparing similar treatment strategies but not emulating SWEFOT obtained results with similar qualitative interpretation (despite some heterogeneity in treatments, outcome measures and effect estimates).^{22–24} The current study shows that close emulation of a trial protocol using similar data can also produce quantitatively similar results. This holds promise for conducting in parallel pragmatic trials and observational trial emulations in large healthcare registries in order to broaden the scope of evidence to the entire population requiring treatment in real-world clinical practice. Observational extensions could also include additional comparator arms, outcome measurements, and longer follow-up, all with the purpose of better informing clinical and regulatory decisionmaking.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

FUNDING

This work was supported by the Swedish Research Council (grant 2016-01355). S.C.K. is supported by the National Institutes of Health (NIH; grant K24AR078959).

CONFLICTS OF INTEREST

J.A. has received research grants to Karolinska Institutet from AbbVie, Bristol Myers Squibb, MSD, Eli Lilly, Pfizer, Roche, Samsung Bioepis, and UCB. S.C.K. has received research grants to the Brigham and Women's Hospital from Roche/Genentech, Pfizer, Bristol Myers Squibb,

Roche, and AbbVie for unrelated studies. S.S. is a part-time employee of deCODE genetics, unrelated to this study. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

A.B., T.F., J.A., S.C.K., and S.S. wrote the manuscript. A.B., T.F., J.A., and S.C.K. designed the research. A.B. performed the research. A.B. analyzed the data.

© 2022 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

- Zink, A. 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 Rheum.* **54**(11), 3399–3407 (2006).
- Vashisht, P., Sayles, H., Cannella, A.C., Mikuls, T.R. & Michaud, K. Generalizability of patients with rheumatoid arthritis in biologic agent clinical trials. *Arthritis Care Res.* **68**(10), 1478–1488 (2016).
- Nash, P. Real-world evidence needs careful interpretation. *J. Rheumatol.* **48**(1), 1–2 (2021).
- Vandenbroucke, J.P. & Psaty, B.M. Benefits and risks of drug treatments: how to combine the best evidence on benefits with the best data about adverse effects. *JAMA* **300**(20), 2417–2419 (2008).
- Smolen, J.S. et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann. Rheum. Dis.* **69**(6), 964–975 (2010).
- Smolen, J.S. et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann. Rheum. Dis.* **79**(6), 685–699 (2020).
- Hernán, M.A., Sauer, B.C., Hernández-Díaz, S., Platt, R. & Shrier, I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* **79**, 70–75 (2016).
- Choi, H.K., Nguyen, U.S., Niu, J., Danaei, G. & Zhang, Y. Selection bias in rheumatic disease research. *Nat. Rev. Rheumatol.* **10**(7), 403–412 (2014).
- Hernán, M.A. & Robins, J.M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**(8), 758–764 (2016).
- Zhao, S.S., Lyu, H., Solomon, D.H. & Yoshida, K. Improving rheumatoid arthritis comparative effectiveness research through causal inference principles: systematic review using a target trial emulation framework. *Ann. Rheum. Dis.* **79**(7), 883–890 (2020).
- Lodi, S. et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. *Am. J. Epidemiol.* **188**(8), 1569–1577 (2019).
- Franklin, J.M. et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies. *Circulation* **143**(10), 1002–1013 (2021).
- Matthews, A.A. et al. Comparing effect estimates in randomized trials and observational studies from the same population: an application to percutaneous coronary intervention. *J. Am. Heart Assoc.* **10**(11), e020357 (2021).
- van Vollenhoven, R. 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. *Lancet* **374**(9688), 459–466 (2009).
- Franklin, J.M. et al. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized

- trial replication project. *Clin Pharmacol Therap* **107**(4), 817–826 (2020).
16. Eriksson, J.K., Askling, J. & Arkema, E.V. The Swedish rheumatology quality register: optimisation of rheumatic disease assessments using register-enriched data. *Clin. Exp. Rheumatol.* **32**(5 Suppl 85), S-147–S-149 (2014).
 17. Laugesen, K. *et al.* Nordic health registry-based research: a review of health care systems and key registries. *CLEP* **13**, 533–554 (2021).
 18. Austin, P.C. & Stuart, E.A. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat. Med.* **34**(28), 3661–3679 (2015).
 19. Austin, P.C. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.* **28**(25), 3083–3107 (2009).
 20. White, I.R., Royston, P. & Wood, A.M. Multiple imputation using chained equations: issues and guidance for practice. *Stat. Med.* **30**(4), 377–399 (2011).
 21. Stürmer, T. *et al.* Propensity score weighting and trimming strategies for reducing variance and bias of treatment effect estimates: a simulation study. *Am. J. Epidemiol.* **190**(8), 1659–1670 (2021).
 22. Curtis, J.R. *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 Res.* **73**(8), 1114–1124 (2021).
 23. Lie, E., van der Heijde, D., Uhlig, T. *et al.* Treatment strategies in patients with rheumatoid arthritis for whom methotrexate monotherapy has failed: data from the NOR-DMARD register. *Ann. Rheum. Dis.* **70**(12), 2103–2110 (2011).
 24. Källmark, H. *et al.* Sustained remission in patients with rheumatoid arthritis receiving triple therapy compared to biologic therapy: a Swedish Nationwide register study. *Arthritis Rheumatol.* **73**(7), 1135–1144 (2021).
 25. Moreland, L.W. *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 Rheum.* **64**(9), 2824–2835 (2012).
 26. O'Dell, J.R. *et al.* Therapies for active rheumatoid arthritis after methotrexate failure. *N. Engl. J. Med.* **369**(4), 307–318 (2013).