This article has been accepted for publication in ANNRHEUMDIS following peer review.
The definitive copyedited, typeset version is available online at
10.1136/ard-2024-226129

Influence of study characteristics on harm estimates from randomised controlled trials in patients with inflammatory arthritis receiving biologic or synthetic anti-rheumatic drugs: A meta-epidemiological study

Johannes Iuel Berg,[1,2] Sabrina Mai Nielsen,[1,2] Esben Malm,[1,2], John P. A. Ioannidis,[3] Daniel E. Furst,[4] Josef S. Smolen,[5] Peter C. Taylor,[6] Lars Erik Kristensen,[1] Simon Tarp,[1] Torkell Ellingsen,[2] Robin Christensen[1,2]

Affiliations:

[1] Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark

[2] Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark

[3] Department of Medicine, Department of Epidemiology and Population Health, Department of Biomedical Data Science, Department of Statistics, and Meta- Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

[4] Division of Rheumatology, Department of Medicine, University of California at Los Angeles, LA, CA, USA

[5] Division of Rheumatology, Department of Medicine 3, Medical University of Vienna, Vienna, Austria

[6] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Corresponding author:

Johannes Iuel Berg, MD
The Parker Institute, Bispebjerg and Frederiksberg Hospital;
Nordre Fasanvej 57 {Vej 8, Indgang 19, 1. Sal};
DK-2000 Copenhagen, Denmark.
e-mail: Johanneskib@gmail.com
ORCID iD: 0009-0006-5411-1496

ABSTRACT

OBJECTIVE: To examine the association between study characteristics and the harms reported in randomised controlled trials (RCTs) on biological and targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) in patients with inflammatory arthritis (IA).

METHODS: We searched MEDLINE for all Cochrane reviews and for systematic reviews published since April 2015. RCTs were eligible if they included patients with IA receiving b/tsDMARD, compared to any comparator arm. Harms were evaluated based on number of withdrawals due to adverse events (WDdtAEs), total withdrawals (WDs), serious adverse events (SAEs), and deaths. Data were extracted for 48 trial/patient characteristics and meta-regression analyses were performed to relate the relative risk ratio (RRR) of harms to the trial characteristics.

RESULTS: A total of 284 trials (from 245 reviews) with 97,607 patients were included, contributing 490 comparisons for the primary analysis. Overall, the relative risk of WDdtAEs was lower when trials used active comparators (RRR, 0.74 [95%CI: 0.58-0.94]) and higher when requiring raised inflammatory markers at enrolment (RRR, 1.25 [1.01-1.55]). Our meta-regression analyses suggested that trials with eligibility criteria for minimum tender/swollen joint count and maximum disease duration decreased the risk of WDs, while previous b/tsDMARDs use at the time of enrolment increased the risk of SAEs.

CONCLUSIONS: Most study characteristics do not affect the reported harm measures. However, a trend was observed where trials selecting patients with higher baseline disease activity found a higher risk ratio of WDdtAEs and SAEs, but also a lower risk of WDs, compared to trials not selecting patients with a high disease activity.

SYSTEMATIC REVIEW REGISTRATION: CRD42020171124.

What is known on this subject?

- Empirical evidence shows that various established bias domains can help assess the risk of bias in individual studies, thereby helping to determine the confidence in the overall findings concerning health benefits.
- There might be differences amongst patient subgroups and in the clinical context, in terms of harm outcomes in randomised trials. Stratified analyses and effect modification are mostly thought of in terms of beneficial treatment effects (as opposed to their modification of harms) and have received most attention.

What this study adds:

- This study provides empirical evidence that frequently used bias domains in randomised trials are not associated with the reported harm estimates.
- Trials selecting patients with a higher disease activity are associated with higher risks of withdrawals due to adverse events and serious adverse events but also a lower risk of total withdrawals, compared to trials selecting patients with lower disease activity.

How this study might affect research, practice or policy

- This study highlights that specific trial design choices can impact reported harms. These findings suggest that both regulatory bodies, when evaluating clinical trial designs and approving new treatments, and those designing future studies may need to adopt more nuanced approaches. Future trials might consider these factors more carefully during the planning phase to ensure balanced and representative results, as a one-size-fits-all model may not effectively address variations in safety outcomes.

INTRODUCTION

Inflammatory arthritis (IA) is an umbrella term grouping diseases with autoimmune inflammation affecting the joints: Rheumatoid arthritis (RA), psoriatic arthritis (PsA), and axial spondyloarthritis.(1,2) Inflammatory cell infiltration of the joints leads to pain, swelling, and impaired mobility as well as cartilage and bone destruction.(3,4) IA can severely reduce patients' quality of life and lead to loss of function and premature death.(5,6) 80-100 adults in 100,000 develop IA every year and RA alone affects about 1% of the world's population, making it a global health concern.(7,8) Conventional synthetic disease-modifying anti-rheumatic drugs (csDMARDs) have been used to treat IA, with varying degrees of success.(9) The more recently introduced biological disease-modifying anti-rheumatic drugs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) can be highly efficacious,(9–11) but all currently available b/tsDMARDs also carry a risk of harms.*(12)* Indeed, some bDMARD and/or tsDMARD studies have indicated potential increased risk of malignancies(13,14), as well as higher rates of serious infections and especially H. zoster.(13,15–17) Recent reviews on tsDMARDs have found a possible increased risk of thromboembolism.(18,19), and some studies suggest that the rates of malignancies and serious infections are comparable between bDMARDs and tsDMARDs.(18,20)

Randomised controlled trials (RCTs) are performed to compare the benefits and harms amongst interventions and should employ rigorous methods that can achieve and preserve comparability of the intervention and comparator groups,(21,22). While it is evident that randomised trials lacking methodological rigour, and flaws in trial conduct can lead to various biases and misestimation of the reported intervention effect, it is unclear to what extent the same is the case for reported relative harm estimates.(26) Furthermore, there might be differences amongst patient subgroups and in clinical context, in terms of harm outcomes in randomised trials. Effect modification is mostly thought of in terms of beneficial treatment effects and this aspect has received much attention,(27,28) in contrast to modification of harmful effects.

In this meta-epidemiological study, we aim to examine the association between study characteristics and the relative harm signals reported in randomised trials on synthetic and biological anti-rheumatic drugs in patients with inflammatory arthritis.

METHODS

This meta-research study follows the PRISMA statement and reporting guidelines for meta-epidemiological research.(29,30) We registered the protocol with PROSPERO (CRD42020171124) on 2020-08-17, and it was published as a peer-reviewed article.(31) This study combines into a single database the data from a large representative sample of available trials already included in published systematic reviews.

Eligibility criteria

Systematic reviews were eligible if they reported on any phase RCTs that included adult patients with RA, PsA, or AxSpA, who received bDMARD or tsDMARD compared to any comparator intervention (ie. Placebo, active comparator, no intervention). The drugs examined in this study (bDMARD and tsDMARD) included those that, on the date the systematic search began, had already been approved by the European Medicines Agency (EMA) and/or the U.S. Food and Drug Administration (FDA) for treating Inflammatory arthritis (supplementary appendix S1). To minimize the inclusion of duplicate references, we limited our selection to reviews published from 2015 onwards. However, we applied no publication year restrictions to Cochrane reviews. Reviews that had been withdrawn or did not have an explicit reference list of included trials were excluded.

The RCTs sampled from the systematic reviews needed to have a readily available full text in English language, in addition to fulfilling the eligibility criteria mentioned above. There was no restriction on the publication year of the individual trials. During the study selection, it was decided that trials that were designed to investigate the effect of drug withdrawal, tapering, or dose reduction should be excluded.

Data sources and search strategy

On the 1st of April 2021, we searched MEDLINE (via PubMed) for systematic reviews (published since April 2015) or Cochrane reviews, using a predefined search string (**supplementary appendix S2**).(31) The RCTs were sampled from the reference lists of the reviews.

Study selection

Two reviewers (JIB and SMN/EM) independently evaluated the systematic reviews for eligibility. Subsequently, the same reviewers found trial references from the eligible reviews. If the reference list in a given review was deemed insufficient, the corresponding author was contacted to obtain a complete reference list. Disagreements were resolved by discussion or by consulting a third reviewer (RC).

We used Python 3.70 software to initially screen the reference lists of the systematic reviews, to minimise extracting duplicate trial references (**supplementary appendix S3**). We used Zotero version 5.0.96.3 software to manage the records retrieved from the search, and the systematic review study selection was conducted using the Covidence online software (www.covidence.org).

Data extraction and data items

A single reviewer (JIB), assisted by (SMN/EM) extracted data using a predefined, standardised data extraction form. In case of uncertainty, we consulted a third reviewer (RC). After assigning each included trial an ID number, we extracted data on first author, year of publication, trial registration, trial duration, disease, number of participants, arms and number of participants randomised to each arm, treatment given in the comparator and active arms, treatment dose, and concomitant medication in trial design. While an Individual Participant Data (IPD) meta-analysis requires direct access to raw data from each study, we utilized an aggregate data meta-analysis approach, relying on reported summary statistics.

For *trial eligibility criteria,* we extracted data on 10 potentially important contextual factor domains: 1) presence of at least one raised inflammatory marker (erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), or morning stiffness); 2-3) minimum TJC and SJC (pragmatically categorised as high [≥6] or low [≤5], respectively); 4-5) maximum TJC and SJC; 6-7) minimum and maximum disease duration; 8) rheumatoid factor (RF); 9) anti-cyclic citrullinated peptide (anti-CCP); and 10) anti-cyclic citrullinated peptide antibody 2 (anti-CCP2) status.

For *medication background,* we extracted data on five potentially important contextual factor domains: 1) DMARD history, defined as the DMARD status of participants at inclusion - categorised as having either never received any DMARDs (DMARD naïve), inadequate

response to csDMARD (csDMARD-IR), or inadequate response to bDMARDs and tsDMARDs (bDMARDs-IR and tsDMARDs-IR); 2-5) Concomitant medication use at inclusion in terms of methotrexate (MTX); csDMARDs other than MTX; bDMARDs or tsDMARDs; and corticosteroids or NSAID - all categorised as the population were either naïve, continued, discontinued in conjunction with inclusion or were not currently using the medication (ie, had previously used and discontinued before screening for inclusion).

For *patient baseline characteristics,* we extracted data on 15 potentially important contextual factor domains: 1) age; 2) proportion of females; 3) disease duration; 4) ESR; 5) CRP; 6) disease activity score (DAS); 7) proportion of rheumatoid factor positive (RF); 8) anti-CCP; 9) anti-CCP2-positive; 10) TJC; 11) SJC; 12) health assessment questionnaire-disability index (HAQ); 13) MD global assessment of disease activity; 14) patient global assessment of disease activity; and 15) patient-reported pain - the last three domains all being measured on a visual analogue scale (VAS).

Furthermore, for *bias-related trial characteristics,* we extracted data on seven potentially important risk-of-bias domains: 1) risk-of-selection bias; 2) performance bias; and 3) attrition bias (categorised as high, low or unclear). We also characterised trials by 4) type (multi-centre or single centre); 5) size (small [<100] or large [≥100]); 6) publication year (before 2005, 2006-2010, 2011-2015, and 2016 or later); and 7) funding (non-profit, 100% industry, mixed, or unclear/undisclosed).

Harm estimates in relation to the use of targeted therapies were analysed to determine whether they were affected by contextual factors amongst population, drug classes (eg, biologics vs small molecules), and trial characteristics. Generic outcome measures covering important harm domains were used, namely: the numbers of trial participants that (*i*) withdrew from the study overall (WDs); (*ii*) withdrew due to adverse events (WD d/t AEs); and (*iii*) experienced serious adverse events (SAEs) as well as mortality. These three outcome measures constituted the dependent variables in the meta-epidemiology database.(12,26) If the number of individual patients with SAEs was not reported, we extracted the number of reported SAEs instead.

Risk of bias assessment

Where possible, we included the risk-of-bias assessments on trials from the source systematic reviews. (22,32,33) The trials with no or insufficient risk-of-bias assessment reported in the systematic reviews were assessed using the Cochrane RoB V.1. (33) on three domains by a single reviewer (JIB) for the following critical RoB domains (32):

- Selection bias - biased allocation to comparison groups
- Performance bias - unequal provision of care apart from treatment under evaluation
- Attrition bias - biased occurrence and handling of deviations from protocol and loss to follow-up

After conducting the main meta-analyses, we performed additional analyses stratified by overall risk of bias. Stratifying the meta-analysis by risk of bias allowed us to assess the potential impact of bias (weather selection bias, performance bias, or attrition bias) on the overall findings.(22) These analyses were used to determine whether the observed summary measures for each type of adverse event were influenced predominantly by studies with a high risk of bias or if they remained consistent across studies with a low risk of bias.

Summary measures and statistical analysis

For each harm outcome, we calculated the risk ratios (RRs) and corresponding standard errors (SEs, for the logRR) based on the number of patients who experienced the outcome in the experimental intervention group and comparison group, and the total number of patients randomised to those groups (ie, corresponding to a trial 2×2 contingency table). The RR values were coded so that RR>1 indicated a potentially more harmful experimental intervention over the comparator. In the case of zero-events, we used the continuity correction suggested by Sweeting et al.(34) For trials with multiple comparisons, the comparator group was divided by the number of comparisons (ie, referred to as randomised comparisons). Furthermore, odds ratios (ORs) with 95% confidence intervals (CIs) were also calculated using the Peto method, for the purpose of sensitivity analysis (**supplementary table S5-8**).

It is important to explore and understand the reasons behind heterogeneity across studies in a meta-analysis. In meta-epidemiological studies, which examine study characteristics and biases there is a need to account for both within- and between-meta-analysis heterogeneity to provide accurate results. We applied restricted maximum likelihood (REML)-based meta-

analysis models with trial ID as a random-effects factor to synthesise the RRs for harms across comparisons (35,36). While the I^2 provides a percentage representation of the proportion of total variability due to heterogeneity in standard contrast-based meta-analysis, the estimate for τ^2 provides an actual measure of the absolute variance in effect sizes (measured in log[RR] units).(37) The association between each outcome's summary measure and each participant and trial characteristic was investigated by univariably adding the individual fixed-factor covariates for the characteristic to the model.(36) If introducing the covariate to the model reduces the estimated τ^2 (between trial variance), this corresponds to an indication of a potentially important effect modifier.(37,38)

All analyses were conducted in the statistical program R with the package *metafor.*(39)

RESULTS

As illustrated in **figure 1** ('Review part'), a total of 683 review records were identified through PubMed. Of these records, 245 systematic reviews were deemed eligible, comprising 2,283 potentially eligible RCTs. After removing duplicates, we assessed the remaining 2,148 records for eligibility (**fig 1,** 'Trial part'). Of these remaining records, 644 trial records representing 304 unique RCTs were eligible for qualitative synthesis. However, 20 trials could not be included in the quantitative synthesis due to incomplete data reporting. Thus, 284 RCTs (490 randomised comparisons) were included in the present data synthesis.

[Figure 1 approx. here]

Study characteristics

The eligible trials included a total of 97,607 patients (63,361 in the experimental intervention groups and 34,246 in comparator groups); a summary of the trial characteristics is shown in **Table 1**. The majority of trials (69.7%) included patients with RA (198 trials and 359 comparisons). Tumour necrosis factor inhibitor (TNFi) bDMARDs were the most frequently used experimental interventions (50.4%), and placebo was the most frequently used comparator (78.2%).

Table 1 Characteristics of the trials included in this study

Characteristics	RCTs (284 trials)
Sample size, no.	275 (20 to 1,759)
No. of comparisons	
1	156 (54.9%)
2	85 (29.9%)
3	23 (8.1%)
4	9 (3.2%)
5	7 (2.5%)
6	4 (1.4%)
Disease	
RA	198 (69.7%)
PsA	40 (14.1%)
AxSpA	46 (16.2%)
Intervention	
TNFi-bDMARDs	143 (50.4%)
non-TNFi-bDMARDs	90 (31.7%)
tsDMARD	42 (14.8%)
Mixed	9 (3.2%)
Trial groups treated with csDMARDs	
None	151 (53.2%)
MTX	117 (41.2%)

Other*	16 (5.6%)
Comparison	
Placebo	222 (78.2%)
Active comparator	54 (19.0%)
No intervention	8 (2.8%)
Trial duration, weeks	16 (2 to 104)

Data are no (%) trials or median (range).
*Other csDMARDs included azathioprine (AZA), Chloroquine (CQ), Hydroxychloroquine (HCQ), Leflunomide (LEF), parenteral gold therapy (PGT), Sulfasalazine (SSZ) and non-steroidal anti-inflammatory drugs (NSAIDs) in AxSpA (including AS) trials.
AxSpA, axial spondyloarthritis; bDMARD, biological disease-modifying anti-rheumatic drugs; MTX, Methotrexate; PsA, psoriatic arthritis; RA, rheumatoid arthritis; RTC, randomised controlled trials; TNFi, tumour necrosis factor inhibitor; tsDMARD, targeted synthetic disease-modifying anti-rheumatic drugs.

Harm estimates across all contexts

The four harm outcomes (WDdtAE, WD, SAE, and death) were reported in 260 (91.5%), 264 (93.0%), 257 (90.5%), and 219 (77.1%) trials, respectively. These figures corresponded to 455, 455, 444, and 403 randomised comparisons, respectively. Only 194 (68.3%) trials reported data on all four outcomes. Forest plots for each of the outcomes are available in **supplementary figure 1**.

The overall effect of adding a targeted therapy was RR, 1.14 (95%CI, 1.04 to 1.27; k=455 comparisons; I^2=24.6%) for WDdtAE; 0.73 (0.68 to 0.78; k=455; 51.4%) for WD; 1.06 (0.98 to 1.14; k=444; 16.5%) for SAE; and 1.06 (0.88 to 1.28; k=403; 0.0%) for death. For three outcomes (WDdtAE, WD, and SAE), sufficient data were available for 34 (71%) of the 48 possible population and trial characteristics. For the outcome death, there were insufficient data on the trial eligibility aCCP positive/negative criterion; thus, data were available for only 33 (69%) characteristics of the 48 possible population and trial characteristics. For the outcome WDdtAE, the between-trial variance (τ^2) was 0.187 (**table 2** and **supplementary table S1**). For the remaining three outcomes (WD, SAE and death), the between-trial variance (τ^2) was 0.147, 0.072, and 0.000, respectively (**supplementary table S2-4**).

Table 2 shows the analysis results of trial design, trial eligibility criteria, and trial characteristics for WDdtAE. Analysis results of all study characteristics for all four outcomes are available in **supplementary table S1-4**.

Table 2: Association between trial/population characteristics and withdrawals due to adverse events

Population and trial characteristics	Comparisons	Summary*	Association with harm outcome†	τ^2	p value
Overall	k=455	-	RR, 1.14 (1.04 to 1.27)	0.187	-
Trial design					
Disease	k=455			0.187	0.598
AxSpA		61 (13.4%)	RR, 1.10 (0.77 to 1.57)		
PsA		62 (13.6%)	RR, 1.01 (0.77 to 1.33)		
RA		332 (73.0%)	RR, 1.17 (1.05 to 1.31)		
Intervention	k=455			0.189	0.512
Mixed		27 (5.9%)	RR, 1.63 (1.02 to 2.63)		
Non-TNFibDMARD		157 (34.5%)	RR, 1.11 (0.94 to 1.31)		
TNFibDMARD		173 (38.0%)	RR, 1.14 (0.97 to 1.33)		
tsDMARD		98 (21.5%)	RR, 1.13 (0.90 to 1.43)		
Comparison	k=455			0.169	0.014
Active comparator		54 (11.9%)	RR, 0.90 (0.73 to 1.11)		
Placebo		393 (86.4%)	RR, 1.21 (1.08 to 1.35)		
No intervention		8 (1.8%)	RR, 2.29 (0.99 to 5.30)		
Trial duration, weeks	k=455	16.0 (2.0 to 104.0)	β, 1.00 (0.99 to 1.00)	0.188	0.701
Trial eligibility criteria					
ESR/CRP/morning stiffness criteria	k=455			0.173	0.044
No		143 (31.4%)	RR, 0.98 (0.81 to 1.17)		
Yes		312 (68.6%)	RR, 1.22 (1.09 to 1.37)		
Min swollen joint count‡	k=355			0.156	0.799
High		255 (71.8%)	RR, 1.22 (1.07 to 1.39)		
Low		100 (28.2%)	RR, 1.19 (0.98 to 1.43)		
Min tender joint count‡	k=352			0.156	0.565
High		263 (74.7%)	RR, 1.24 (1.09 to 1.41)		
Low		89 (25.3%)	RR, 1.16 (0.94 to 1.42)		
Min disease duration, months	k=232	6.0 (1.4 to 36.0)	β, 1.01 (0.98 to 1.04)	0.111	0.542
Max disease duration, months	k=84	60.0 (3.0 to 240.0)	β, 1.00 (1.00 to 1.00)	0.330	0.954
RF	k=79			0.374	0.478
Negative		8 (10.1%)	RR, 1.91 (0.74 to 4.94)		
Positive		71 (89.9%)	RR, 1.34 (1.02 to 1.76)		
aCCP	k=66			0.179	0.871
Negative		1 (1.5%)	RR, 1.00 (0.02 to 66.24)		
Positive		65 (98.5%)	RR, 1.41 (1.11 to 1.80)		
Trial characteristics					
RoB selection bias	k=455			0.185	0.392
High		5 (1.1%)	RR, 1.43 (0.59 to 3.44)		
Low		174 (38.2%)	RR, 1.22 (1.06 to 1.41)		
Unclear		276 (60.7%)	RR, 1.07 (0.93 to 1.23)		
RoB performance bias	k=455			0.190	0.790
High		31 (6.8%)	RR, 1.06 (0.75 to 1.50)		
Low		255 (56.0%)	RR, 1.13 (0.99 to 1.29)		
Unclear		169 (37.1%)	RR, 1.19 (1.01 to 1.41)		
RoB attrition bias	k=455			0.191	0.906
High		27 (5.9%)	RR, 1.22 (0.86 to 1.75)		
Low		396 (87.0%)	RR, 1.14 (1.02 to 1.26)		
Unclear		32 (7.0%)	RR, 1.20 (0.76 to 1.90)		
Trial type	k=451			0.189	0.649
Multicenter		439 (97.3%)	RR, 1.14 (1.03 to 1.26)		
Singlecenter		12 (2.7%)	RR, 1.44 (0.53 to 3.97)		
Trial size§	k=455			0.188	0.641
Large		217 (47.7%)	RR, 1.13 (1.01 to 1.27)		
Small		238 (52.3%)	RR, 1.19 (0.98 to 1.45)		
Publication year, calendar year	k=455	2013 (1998 to 2021)	β, 0.99 (0.97 to 1.01)	0.191	0.425

*Data are no. (%) comparisons, median percentage (range of percentages), or median of means (range of means) for aggregated data.
†To investigate the association between each of the population and trial characteristics and risk of harm (ie, logRRs for withdrawals due to AEs), separate REML-based meta-regression models with random effects, including a factor for the characteristic were performed. The slope β should be interpreted as the proportional increase (or decrease) in the RR per 1 unit increase in the characteristic. A slope of β=1.00 indicates no association with the treatment effect, whereas, for example, a slope of β=1.03 for disease duration means that for every 1 year, the RR increases by a factor 1.03 (ie, 3%). It was not possible to analyse three trial characteristics: max TJC, max SJC, and aCCP2 status, as well as the patient baseline characteristic aCCP2 positive percentage, due to missing and/or inconsistent reporting in the eligible trials.
‡High and low indicate six or more joints and five or fewer joints, respectively.
§Large and small indicate 100 or more participants and 99 or fewer participants, respectively.
aCCP, anti-cyclic citrullinated peptide antibody; bDMRD, biological disease-modifying anti-rheumatic drug; CRP, c-reactive protein; csDMARD, conventional synthetic disease-modifying anti-rheumatic drug; DMARD, disease-modifying anti-rheumatic drug; ESR, erythrocyte sedimentation rate; HAQ-DI, health assessment questionnaire-disability index; IR, inadequate responder; MD, medical doctor; MTX, methotrexate; NSAID, non-steroidal anti-inflammatory drug; Pt, patient; RF, rheumatoid factor; tsDMARD, targeted synthetic disease-modifying anti-rheumatic drug; VAS, visual analogue scale

Withdrawals due to adverse events, and estimates of intervention effects

The risk of WDdtAE was 26% lower in trials using an active comparator relative to trials using placebo or no intervention (RRR, 0.74 [0.58 to 0.94]), reducing the between-trial variance by 9.6% (τ^2=0.169) and with a statistically suggestive interaction (*P*=0.014; k=455 comparisons). In contrast, the risk of WDdtAE was increased by 25% when the eligibility criteria explicitly required a minimum level of the inflammatory markers ESR, CRP, and/or morning stiffness (RRR, 1.25 [1.01 to 1.55]), reducing the between-trial variance by 7.5% (τ^2=0.173), with a statistically suggestive interaction (*P*=0.044; k=455 comparisons).

Withdrawals and estimates of intervention effects

Relative to the comparator, the likelihood of WD in the intervention group was 18% lower in trials requiring participants to have a minimum TJC of 6 or more (RRR, 0.82 [0.69 to 0.98]) and 15% lower in trials requiring participants to have a minimum SJC of 6 or more (RRR, 0.85 [0.72 to 1.00]). For minimum TJC the between-trial variance was increased by 6.1% (τ^2=0.156) with a statistically suggestive interaction (*P*=0.029; k=348 comparisons) and for minimum SJC the between-trial variance was increased by 6.1% (τ^2=0.156) with a possible interaction (*P*=0.057; k=351 comparisons). Also, the likelihood of WD was reduced in the intervention group by 4% per year when the trials extended the criteria for maximum disease duration (β-years, 0.96 [0.94 to 0.98]), reducing the between-trial variance by 72.1% (τ^2=0.041), with a highly statistically significant interaction (*P*=<0.001; k=85 comparisons).

Serious adverse events and estimates of intervention effects

The likelihood of SAE in the intervention group was increased, corresponding to a 93% higher likelihood in trials where participants had previously been using a bDMARDs or tsDMARDs but discontinued before screening, compared to trials where participants were naïve (RRR, 1.93 [1.30 to 2.86]), reducing the between-trial variance by 1.4% (τ^2=0.071), with a statistically suggestive interaction (*P*=0.011; k=411 comparisons) across all four bDMARDs or tsDMARDs backgrounds. For death as an outcome and the remaining population and trial characteristic factors across all outcomes, no statistical interactions were observed.

Overall, analyses stratified by rheumatological conditions and other sensitivity analyses indicated that the conclusions from the main analysis were robust, as they did not lead to different interpretations. Although, due to the low number of events, some did not reach statistical significance, the direction of the effect remained consistent across all point estimates. Tables reporting these stratified analyses and sensitivity approaches, are available in **Supplementary Table S5-20.**

[Figure 2 approx. here]

DISCUSSION

In this large meta-epidemiological study of randomised RA, PsA, and AxSpA trials, we investigated whether contextual factors modify the risk of harms in trials testing targeted therapies. Trial designs with active comparators, and trials with the eligibility criteria ESR/CRP/morning stiffness criteria, minimum TJC, and maximum disease duration - as well as medication status of bDMARDs or tsDMARDs at enrolment - appear to influence the risk of harms, albeit not always in the same direction.(40) Overall, across the eligible randomised comparisons, there was a higher risk of withdrawals due to AEs, and a lower risk of total withdrawals in the intervention groups compared to the comparison groups when receiving a b/tsDMARD.

In general, although separate markers were present within each outcome, our results indicate that trials including patients who had raised inflammatory markers or were not currently (but had been) using bDMARDs or tsDMARDs are at greater relative risk of WDdtAE and SAE from the experimental intervention compared to the control comparator. However, our findings also reveal that trials with patients having more severe manifestations of their condition (e.g., higher tender joint count and longer disease duration), are less likely to withdraw during the trial period. This counterintuitive result may be attributed to these patients' greater familiarity with managing their condition and coping with various drugs and side effects. Consequently, their ability to remain in the trial could reflect their resilience and adaptability to treatment-related challenges. Thus, in this context, withdrawals (regardless of cause) might not necessarily indicate a lack of harm but could instead be interpreted as a measure of perceived benefit, reflecting the patients' commitment to remaining in the trial due to the advantages they experience, such as improved access to care or symptom management.

Patients who have discontinued a bDMARDs or tsDMARDs prior to the time of enrolment might have a history of AEs and/or SAEs from previous treatments and therefore might as a group be at higher risk compared to patients who are naïve or in active treatment at enrolment. These patients might also have a higher disease activity due to previous treatment failures and thus lack proper disease control at inclusion. Furthermore, patients with higher disease activity are thought to have a greater potential for improvement and can therefore be more motivated to stay longer in the trials. Similarly, subjects recruited from regions where they

have poor access to efficacious therapies can be highly motivated to receive medical care, particularly with a chance for active treatment, in a trial context.

In trials employing an active comparator, participants in the comparator group were more likely to withdraw due to AEs compared to participants in placebo trials. This outcome could indicate a potential bias in active comparator trials, wherein researchers may lean towards portraying the intervention as less harmful. Alternatively, the significantly greater efficacy noted in active comparator studies compared to placebo-controlled trials might influence a preference for remaining in the trial. This observation aligns with previous findings that demonstrate that RA patients in RCTs exhibit a higher likelihood of response compared to those in placebo-controlled trials testing the same drug.(41) This phenomenon has been attributed to psychological factors in both patients and trial assessors, such as expectations, hope, and perception, suggesting that these may contribute to an outcome measurement artifact. This implies that outcomes in RA trials may reflect more than just the reduction of inflammation achievable by the trial drugs.(42) The extent to which this phenomenon could explain the observed effect modification by trial type (placebo-controlled vs. active comparator-controlled) on harm outcomes remains an open question. However, since most recent trials incorporate bDMARDs as active comparators, the risk of AE might be different. Consequently, it appears that the AE profile of conventional DMARDs is similar to or potentially worse than that of b/tsDMARDs. Notably, potentially related to "healthy user effect", the opposite trend was observed for the outcome SAE, although this finding could not be confirmed as statistically significant.

The duration of follow-up varied across the trials, making it difficult to ensure consistency in drug exposure over time. Some adverse effects, such as malignancies, may emerge only after extended periods. Therefore, if the follow-up time was insufficient, caution should be exercised when generalizing our findings to all potential adverse events. Although our study was not specifically designed to evaluate individual or specific SAEs, the variability and in some cases, relative brevity in follow-up duration constitutes a limitation to ensure that our findings are accurately interpreted within the context of the follow-up periods relevant to randomised trials. Also, when reporting outcomes such as withdrawal and withdrawal due to adverse events in clinical trials, the attribution of attrition to adverse events, loss of efficacy, lack of efficacy, or a combination thereof presents a significant challenge. This complexity renders these harm

outcomes conceptually difficult to interpret and complicates the drawing of robust conclusions. In contrast, outcomes like SAEs and mortality, while more straightforward, are infrequent and thus do not present the same interpretive difficulties.

Relation to other studies

Previously, concerns have been voiced over the majority of patients in routine care not being eligible candidates for participating in randomised trials due to strict inclusion criteria for testing of new drugs in typical clinical trials.(43,44) The findings of this study show that trials including patients with higher disease activity also experience an increased risk of AEs in the treatment groups.

A previous meta-epidemiological study found that trials with highly selective inclusion criteria did not show increased efficacy of targeted therapies.(45) This finding raises the question of whether trial inclusion criteria for clinical trials on new treatment need to be revised. Our findings showed that trials extending the criteria for maximum disease duration lowered the risk of patients' withdrawing from the intervention. This finding is in line with previous meta-analysis,(45,46) which found an overall positive effect of longer disease duration on treatment efficacy outcomes in RA trials. The data presented here suggest that fewer withdrawals may be achieved by including patient populations with longer disease durations, as this may lead to patients being sufficiently satisfied with the results obtained, despite the potentially lower efficacy.(47)

Strengths and limitations

This study has several strengths. First, this study is the most comprehensive analysis to date on the risk of harm of bDMARDs and tsDMARDs across IA diagnoses, having collected trial data on nearly 500 randomised comparisons. Second, all FDA/EMA-approved drugs were included, which adds to the study's external validity. Third, most previous meta-epidemiological studies in this field have focused mainly on patient baseline characteristics, whereas we have also looked at trial eligibility criteria. Fourth, we included only RCTs to avoid inherent problems when investigating effect modifiers in non-RCTs.(48) Fifth, the study selection was performed by two independent reviewers.

This study also has limitations that should be considered. First, we did not do double

data extraction by two independent reviewers, so the amount of missing data might be slightly overestimated. Second, all extracted variables were limited to information from the trial reports, so it was impossible to investigate a number of the predefined characteristics due to sparse or inconsistent reporting. Also, as observed in a previous review, harms were generally heterogeneously and insufficiently reported in trials (eg, many RCTs chose only to report treatment-related SAE, whereas others reported on all SAEs.(49) Third, a notable issue that could be considered a caveat, is that the baseline characteristics prespecified for extraction were primarily relevant to rheumatoid arthritis (RA) and might not adequately address the other diseases under study. Indeed, in AxSpA and, to some degree, PsA studies, different clinical and biochemical measures are used, thus resulting in a great deal of missed data for those studies. Thus, pooling data between IA diseases has limitations, given they have different comorbidities, age at presentation, etc. Post hoc, we conducted additional analyses stratified by type of inflammatory condition. Robust findings should remain consistent and reliable across different analytical approaches and conditions. Indeed, all these exploratory secondary sensitivity analyses indicated that the findings were robust across different conditions. To judge the robustness of our findings after running a series of sensitivity analyses, we evaluated the consistency of results, statistical and clinical significance, variability, and provided detailed documentation (see **Supplementary Table S9-20**). Fourth, by relying on the search results of published reviews, we assumed that the retrieved sample of trials for our analysis was representative. However, eligible trials might have been omitted—due, for example, to specific eligibility criteria in the systematic reviews, or trials being too recent to have been included in a systematic review on the date of the search. Such omissions theoretically could have influenced our results, but there is no reason to believe that harms estimates would be affected in one particular direction. Fifth, many statistical tests were performed, increasing the risk of type I error (ie, finding statistically significant results that in actuality appeared by chance or could be due to meta-confounding). Therefore, we consider p-values <0.05 suggestive and p<0.005 as statistically significant.

Sixth, trial withdrawal is a multifaceted construct influenced by a range of factors, and different effect modifiers could potentially interact with each other, influencing the overall outcome. In this study, effect modification was assessed using a variable-by-variable approach due to limited statistical power. However, a multivariable approach, allowing for simultaneous modelling of

various trial characteristics and meta-confounders, could provide a more nuanced understanding of the factors influencing trial withdrawal.(50) This approach would potentially help clarify how different effect modifiers may interact and jointly impact outcomes. Future research with larger sample sizes and more comprehensive data collection should consider such multivariable modelling to fully capture the complex interplay of factors affecting trial withdrawal.

Seventh, the number of comparisons across contextual variables varied in our study, this may have resulted in insufficient power to detect meaningful effect modification for some factors. Consequently, while some variables showed reasonable changes in τ^2, these were not always statistically significant. A lack of statistical significance does not necessarily imply the absence of effect modification, and we interpret these findings with caution, as true effect modification may have been missed due to limited power.

Eighth, it is relevant to be mindful that this study only assessed the risk of harms. Future studies should evaluate the risk-benefit profile of the interventions, by also weighing against the efficacy of the intervention to hopefully provide a more comprehensive understanding of their clinical utility. Also, to review the largest possible sample of data, we included trials in several phases, and while most of the trials assessed standard dosing, we also included trials with differing and not always optimal dosing-regiments, which might have led to an over/underestimation of the harms.

Last, all data were obtained at trial level, including patient characteristics that ideally should have been collected at patient level. The use of aggregated data to draw conclusions on individuals should be done with care. When performing meta-regression analysis based on averages, the results might not reflect the true relationship within the study and hence might be misleading. This phenomenon is known as ecological fallacy.(37) Ideally, such population characteristics should have been investigated using individual patient data, although such data are unlikely to be available in the larger scale needed for the meta-epidemiological study of harms observed across RCTs.

Conclusions

In conclusion, our results suggest that four trial eligibility criteria - ESR/CRP/morning stiffness criteria, minimum TJC, maximum disease duration - as well as medication status of bDMARDs or

tsDMARDs at enrolment, might influence the reported risk of harms across the examined outcomes. Trials that select patients with a higher disease activity seem to report a higher risk ratio of WDdtAsE and SAEs but also a lower risk ratio of total WDs, compared to trials selecting patients with lower disease activity.

Acknowledgements

We would like to thank Prof Ching-Chi Chi, Department of Dermatology, Chang Gung Memorial Hospital, Linkou, Taiwan and Professor Young Ho Lee, Department of Rheumatology, Korea University, for providing the full text of their research, ensuring the article could undergo a proper assessment in the study selection and (potential) inclusion in our meta-analysis. We would like to thank Dorthe Bang Berthelsen, PhD student at Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark, and fellow of the OMERACT Safety Working Group, for assisting in interpreting the results.

Contributors: JIB, SMN, EM, and RC conceived and designed the study. JIB, SMN, and EM acquired the data. SMN carried out the statistical analysis. JIB, SMN, and RC interpreted the data. JIB, SMN and RC drafted the manuscript. All authors critically revised the article for important intellectual content and gave final approval for the article. The last author, RC, is the guarantor and accepts full responsibility for the work and the conduct of the study. RC also attests that all listed authors meet the authorship criteria and that no one meeting these criteria has been omitted.

Funding: This study was supported by no specific grant. JIB was supported by the PhD programme at the Faculty of Health Science, University of Southern Denmark. Section for Biostatistics and Evidence-Based Research, The Parker Institute, (SMN and RC) is supported by a core grant from The Oak Foundation (OCAY-18-774-OFIL), a group of philanthropic organizations giving grants to not-for-profit organizations around the world. The funding organisations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Competing interests: All authors have completed the ICMJE (International Committee of Medical Journal Editors) uniform disclosure form www.icmje.org/disclosure-of-interest/ and declare they received no financial support from any industry for the submitted work. However, DEF has received Grants and/or research support from Amgen, Corbus, CSL Behring, Galapagos, Gilead, GSK, Horizon, Kadmon, Novartis, Pfizer,Roche/Genetech, Talaris; Consulting fees from Amgen, Corbus, Galapagos, Horizon,Kadmon, Pfizer, Talaris and payment or honoraria from CME. JSS has received institution grants or contracts from Abbvie, Astra-Zeneca, Eli Lilly and Company, Galapagos; Consultant fees from AbbVie, Amgen, Ananda, Astro, BMS, Celltrion, Chugai, Eli Lilly and Company, Gilead, Immunovant, MSD, Novartis, Pfizer, Roche, R-Pharma, Samsung, Sanofi, UCB; Payment or honoraria from Eli Lilly, UCB, Chugai, Sandoz. PCT has received institution grants or contracts from Galapagos; Consulting fees from AbbVie, Biogen, Galapagos, Gilead Sciences, GlaxoSmithKline, Janssen, Nordic Pharma, Pfizer Inc, UCB, Eli Lilly, Fresenius, Acelyrin; Payment or honoraria from AbbVie; Support for attending meetings and/or travel from Lilly; Participation on a Data Safety Monitoring Board or Advisory Board for Sanofi, Kymab, Immunovant. LEK has

received consultant fees and payment or honoraria from AbbVie, Amgen, Biogen, BMS, Eli Lilly, Gilead, Janssen pharmaceuticals, MSD, Novartis, Pfizer, and UCB Pharma; Has stock options in Novartis, Merck, Eli Lilly, Novo Nordisk, UCB and is employed at Leo Pharmaceutical from the 1st of June 2024. RC has received consulting fees from Image Analysis Ltd, trading as IAG, Image Analysis Group (UK) and Compass Communications Ltd.; Payment or honoraria from Osteoarthritis and Cartilage, Acta Orthopaedica and Frontiers in Drug Safety and Regulation; A Leading member of OMERACT (Outcome Measures in Rheumatology), A member of the GRADE Working Group and editor for Cochrane Collaboration.

Ethical approval: Not required.

Data sharing: Details of the characteristics of the included studies were shared in the supplementary materials. The study-specific data included in the meta-analysis can be obtained from the corresponding author at Johanneskib@gmail.com.

Patient and Public Involvement: During the development of the objective and protocol, we aimed to integrate the patient perspective into the research procedures. This included the active involvement of members of the OMERACT Safety and Contextual Factors Working Groups. These patient collaborators provided invaluable insights derived from their direct encounters with safety issues and a nuanced comprehension of how various personal and contextual factors might impact their encounters with side effects stemming from rheumatology treatment. This considerable deliberations with OMERACT participants, particularly the patient research partners, ensured that the meta-research questions, although academically abstract, stayed pertinent and attuned to patient experiences and apprehensions.

REFERENCES

1. Ledingham J, Snowden N, Ide Z. Diagnosis and early management of inflammatory arthritis. BMJ. 2017 Jul 27;358:j3248.

2. Raychaudhuri SP, Deodhar A. The classification and diagnostic criteria of ankylosing spondylitis. J Autoimmun. 2014 Feb 1;48–49:128–33.

3. Miyabe Y, Miyabe C, Luster AD. LTB4 and BLT1 in inflammatory arthritis. Semin Immunol. 2017 Oct 1;33:52–7.

4. McInnes IB, Schett G. The Pathogenesis of Rheumatoid Arthritis. N Engl J Med. 2011 Dec 8;365(23):2205–19.

5. Janke K, Biester K, Krause D, Richter B, Schürmann C, Hirsch K, et al. Comparative effectiveness of biological medicines in rheumatoid arthritis: systematic review and network meta-analysis including aggregate results from reanalysed individual patient data. BMJ. 2020 Jul 7;370:m2288.

6. Schweikert B, Malmberg C, Núñez M, Dilla T, Sapin C, Hartz S. Cost-effectiveness analysis of ixekizumab versus secukinumab in patients with psoriatic arthritis and concomitant moderate-to-severe psoriasis in Spain. BMJ Open. 2020 Aug 1;10(8):e032552.

7. Savolainen E, Kaipiainen-Seppänen O, Kröger L, Luosujärvi R. Total incidence and distribution of inflammatory joint diseases in a defined population: results from the Kuopio 2000 arthritis survey. J Rheumatol. 2003 Nov 1;30(11):2460–8.

8. Hartman L, Rasch LA, Klausch T, Bijlsma HWJ, Christensen R, Smulders YM, et al. Harm, benefit and costs associated with low-dose glucocorticoids added to the treatment strategies for rheumatoid arthritis in elderly patients (GLORIA trial): study protocol for a randomised controlled trial. Trials. 2018 Jan 25;19(1):67.

9. Goulabchand R, Mouterde G, Barnetche T, Lukas C, Morel J, Combe B. Effect of tumour necrosis factor blockers on radiographic progression of psoriatic arthritis: a systematic review and meta-analysis of randomised controlled trials. Ann Rheum Dis. 2014 Feb 1;73(2):414–9.

10. Christensen AW, Tarp S, Furst DE, Døssing A, Amris K, Bliddal H, et al. Most Trial Eligibility Criteria and Patient Baseline Characteristics Do Not Modify Treatment Effect in Trials Using Targeted Therapies for Rheumatoid Arthritis: A Meta-Epidemiological Study. PLOS ONE. 2015 Sep 11;10(9):e0136982.

11. Callhoff J, Sieper J, Weiß A, Zink A, Listing J. Efficacy of TNFα blockers in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis: a meta-analysis. Ann Rheum Dis. 2015 Jun 1;74(6):1241–8.

12. Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better Reporting of Harms in Randomized Trials: An Extension of the CONSORT Statement. Ann Intern Med. 2004 Nov 16;141(10):781–8.

13. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V. Anti-TNF Antibody Therapy in Rheumatoid Arthritis and the Risk of Serious Infections and MalignanciesSystematic Review and Meta-analysis of Rare Harmful Effects in Randomized Controlled Trials. JAMA. 2006 May 17;295(19):2275–85.

14. 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.

15. Singh JA, Wells GA, Christensen R, Ghogomu ET, Maxwell LJ, MacDonald JK, et al. Adverse effects of biologics: a network meta-analysis and Cochrane overview. Cochrane Database Syst Rev [Internet]. 2011 [cited 2021 Dec 13];(2). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD008794.pub2/full

16. Singh JA, Cameron C, Noorbaloochi S, Cullis T, Tucker M, Christensen R, et al. Risk of serious infection in biological treatment of patients with rheumatoid arthritis: a systematic review and meta-analysis. The Lancet. 2015 Jul 18;386(9990):258–65.

17. Winthrop KL, Cohen SB. Oral surveillance and JAK inhibitor safety: the theory of relativity. Nat Rev Rheumatol. 2022 May;18(5):301–4.

18. Sepriano A, Kerschbaumer A, Smolen JS, Heijde D van der, Dougados M, Vollenhoven R van, et al. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2019 update of the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2020 Jun 1;79(6):760–70.

19. Mease P, Charles-Schoeman C, Cohen S, Fallon L, Woolcott J, Yun H, et al. Incidence of venous and arterial thromboembolic events reported in the tofacitinib rheumatoid arthritis, psoriasis and psoriatic arthritis development programmes and from real-world data. Ann Rheum Dis. 2020 Nov 1;79(11):1400–13.

20. Salgado E, Maneiro JR, Carmona L, Gomez-Reino JJ. Safety profile of protein kinase inhibitors in rheumatoid arthritis: systematic review and meta-analysis. Ann Rheum Dis. 2014 May 1;73(5):871–82.

21. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016 Oct 12;355:i4919.

22. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019 Aug 28;366:l4898.

23. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008 Mar 15;336(7644):601–5.

24. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995 Feb 1;273(5):408–12.

25. Savović J, Jones H, Altman D, Harris R, Jűni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. Health Technol Assess Winch Engl. 2012 Sep;16(35):1–82.

26. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. BMJ. 2016 Feb 1;352:i157.

27. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized

controlled trials and meta-analyses. CMAJ Can Med Assoc J J Assoc Medicale Can. 2020 Aug 10;192(32):E901–6.

28. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. J Clin Epidemiol. 2021 Jun;134:174–7.

29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71.

30. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. BMJ Evid-Based Med. 2017 Aug 1;22(4):139–42.

31. Malm E, Nielsen SM, Berg J, Ioannidis JPA, Furst D, Smolen JS, et al. Risk of harm in synthetic and biological intervention trials in patients with inflammatory arthritis: protocol for a metaepidemiological study focusing on contextual factors. BMJ Open. 2021 Sep 1;11(9):e049850.

32. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001 Jul 7;323(7303):42–6.

33. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

34. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004 May 15;23(9):1351–75.

35. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med. 1999 Feb 15;18(3):321–59.

36. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002 Feb 28;21(4):589–624.

37. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002 Jun 15;21(11):1559–73.

38. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med. 2002 Jun 15;21(11):1513–24.

39. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw. 2010 Aug 5;36:1–48.

40. Kerschbaumer A, Smolen JS, Herkner H, Stefanova T, Chwala E, Aletaha D. Efficacy outcomes in phase 2 and phase 3 randomized controlled trials in rheumatology. Nat Med. 2020 Jun;26(6):974–80.

41. Kerschbaumer A, Stimakovits NM, Smolen JS, Stefanova T, Chwala E, Aletaha D. Influence of active versus placebo control on treatment responses in randomised controlled trials in rheumatoid arthritis. Ann Rheum Dis. 2023 Apr;82(4):476–82.

42. Landewé RBM, Sepriano A, Bergstra SA. Why most (but perhaps not all) DMARDs work equally well. Semin Arthritis Rheum. 2024 Feb;64S:152316.

43. Sokka T, Pincus T. Eligibility of patients in routine care for major clinical trials of anti-tumor necrosis factor alpha agents in rheumatoid arthritis. Arthritis Rheum. 2003 Feb;48(2):313–8.

44. He J, Morales DR, Guthrie B. Exclusion rates in randomized controlled trials of treatments for physical conditions: a systematic review. Trials. 2020 Feb 26;21(1):228.

45. Christensen AW, Tarp S, Furst DE, Døssing A, Amris K, Bliddal H, et al. Most Trial Eligibility Criteria and Patient Baseline Characteristics Do Not Modify Treatment Effect in Trials Using Targeted Therapies for Rheumatoid Arthritis: A Meta-Epidemiological Study. PLOS ONE. 2015 Sep 11;10(9):e0136982.

46. Kanters S, Druyts E, Mills EJ, Thorlund K. What drives the comparative effectiveness of biologics vs methotrexate in rheumatoid arthritis? Meta-regression and graphical inspection of suspected clinical factors. Rheumatology. 2014 Jul 1;53(7):1264–73.

47. Hamann P, Holland R, Hyrich K, Pauling JD, Shaddick G, Nightingale A, et al. Factors Associated With Sustained Remission in Rheumatoid Arthritis in Patients Treated With Anti-Tumor Necrosis Factor. Arthritis Care Res. 2017 Jun;69(6):783–93.

48. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clin Epidemiol. 2017 Jun 8;9:331–8.

49. Hadi MA, McHugh GA, Conaghan PG. Quality of reporting of harms in randomised controlled trials of pharmacological interventions for rheumatoid arthritis: a systematic review. Evid Based Med. 2017 Oct;22(5):170–7.

50. Christensen R, Berthelsen DB. Controversy and Debate on Meta-epidemiology. Paper 3: Causal inference from meta-epidemiology: a reasonable goal, or wishful thinking? J Clin Epidemiol. 2020 Jul;123:131–2.