BMJ Open

Risk of harm in synthetic and biological intervention trials in patients with inflammatory arthritis: protocol for a metaepidemiological study focusing on contextual factors

Esben Malm ☉,1,2 Sabrina Mai Nielsen ☉,1,2 Johannes Berg,1,2 John P. A. Ioannidis ☉,3 Daniel Furst,4 Josef S. Smolen,5 Peter C. Taylor ☉,6 Lars Erik Kristensen,1 Simon Tarp,1 Torkell Ellingsen ☉,2 Robin Christensen1,2


Methods and analysis We will search MEDLINE (via PubMed) for systematic reviews published since April 2015 and all Cochrane reviews. From these reviews, randomised trials will be eligible if they include patients with an IA condition with at least one group randomly allocated to bDMARD and/or tsDMARD treatments. A predefined form will be used for extracting data on population characteristics (eg, baseline characteristics or eligibility criteria, such as medication background) and specific harm outcome measures, such as number of withdrawals, numbers of patients discontinuing due to adverse events and number of patients having SAEs. RoB in individual trials will be assessed using a modified Cochrane RoB tool. We will estimate the potentially causal harm effects related to the experimental intervention compared with control comparator as risk ratios, and heterogeneity across randomised comparisons will be assessed statistically and evaluated as inconsistency using the I² Index. Our meta-regression analyses will designate population and trial characteristics and each RoB domain as independent variables, whereas the three harm domains will serve as dependent variables.

ABSTRACT

Introduction Inflammatory arthritis (IA) conditions, including rheumatoid arthritis, psoriatic arthritis and axial spondyloarthritis, are characterised by inflammatory infiltration of the joints. Biological disease-modifying antirheumatic drugs (bDMARDs) and targeted synthetic disease-modifying antirheumatic drugs (tsDMARDs), respectively, reduce the effects of proinflammatory cytokines and immune cells to ameliorate disease. However, immunosuppression can be associated with high rates of serious adverse events (SAEs), including serious infections, and maybe an increased risk of malignancies and cardiovascular events. Currently, there is no empirical evidence on the extent to which contextual factors and risk of bias (RoB) domains may modify these harm signals in randomised trials.

INTRODUCTION

Background Inflammatory arthritis (IA) is a heterogeneous group of autoimmune diseases that includes rheumatoid arthritis (RA), psoriatic arthritis (PsA), axial spondyloarthritis (AxSpA, including ankylosing spondylitis (AS)).1 2 These diseases are characterised by inflammatory infiltration of the joints,3 resulting in pain, swelling, stiffness and restricted movement.4 Ultimately, they can have a detrimental impact on quality of life and can cause progressive disability and premature death.5 6 Not only are the implications severe but also the diseases are a global concern; RA alone affects about 1% of the world’s population.7

Aside from conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) including methotrexate (MTX), targeted therapies, consisting of biological...
Rationale
Designing, conducting and reporting RCTs should incorporate methods—such as concealing randomised allocation, blinding participants and personnel, and appropriately engaging the intention-to-treat population—that avoid biases resulting from incompatibilities between the intervention and control groups. Empirical evidence suggests that the absence of rigorous methodology can lead to biased intervention effect estimates (ie, net benefits).28 Contextual factors, such as population and trial characteristics, also have been shown to be possible effect modifiers when assessing benefits.8 29

The effect on harms of RoB, trial characteristics and contextual factors (eg, population characteristics) has yet to be investigated, and using metaepidemiology is the method of choice for doing so.30 Previous metaepidemiological research within rheumatology has primarily investigated the effect of methodological quality (internal validity) on treatment effects (benefit).3 29 31 32 To our knowledge, no one has yet investigated the importance of IA therapies of contextual factors (ie, population and trial characteristics) and RoB when the outcome is harmful effects.

Aims and objectives
This exploratory metaepidemiological study aimed to improve harm reporting by investigating the influence of RoB domains, trial characteristics and contextual factors on the three harm measures: all withdrawals (WD), withdrawals due to adverse events (WD d/t AEs) and serious adverse events (SAEs), as well as mortality among patients who were treated with bDMARDs and/or tsDMARDs. Our objective is to explore whether specific participant, drug classes, and trial characteristics (ie, contextual factors including RoB domains) have a quantitative influence in terms of effect modification and/or distortion due to biases on the observed likelihood of harm from an experimental intervention compared with control comparators in randomised trials. If contextual factors have such an influence, we hope to shed light on their importance to future trial reporting and interpretation of harm when reporting randomised trials.17

Protocol and registration
This protocol was developed in accordance with the V.1.07 2018 Methodological Expectations for Cochrane Intervention Reviews recommendations of the Cochrane Collaboration33 and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Protocols guidelines.34 The study protocol was prespecified and registered in PROSPERO on 17 August 2020 (CRD number: CRD42020171124).

Table 1 Research objective described using PICO

<table>
<thead>
<tr>
<th>Participation</th>
<th>Adult patients with rheumatoid arthritis, psoriatic arthritis, axial spondyloarthritis or anklyosing spondylitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>Targeted therapies (ie, biological disease-modifying antirheumatic drugs and targeted synthetic disease-modifying antirheumatic drugs)</td>
</tr>
<tr>
<td>Comparison</td>
<td>Placebo, standard care or waiting list/no intervention, active comparator and unclear (ie, comparison described in insufficient detail to fit within the other categories)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Number of withdrawals, number of withdrawals due to adverse events, number of serious adverse events and number of patients who died</td>
</tr>
</tbody>
</table>

PICO, participants, intervention, comparison and outcomes.
METHODS AND ANALYSIS

This metaepidemiological study combines data from a large representative sample of available trials into a single database by assembling trials already included in published systematic reviews. Generic outcome measures that cover important harm domains, namely, (1) WD, (2) WD d/t AEs and (3) SAEs,\textsuperscript{17} and mortality will constitute the dependent variables in the database. Harm effects in relation to the use of targeted therapies will be analysed to determine whether they are affected by contextual factors among population, drug classes (e.g., biologics vs small molecules) and trial characteristics.

Eligibility criteria

We will search MEDLINE (via PubMed) for systematic reviews (published since April 2015) or Cochrane reviews. From these systematic reviews, included randomised trials will (independent of publication year) be eligible if they fulfill our inclusion criteria (see Table 1 for participants, intervention, comparison and outcomes (PICO) framework), with at least one group randomly allocated to bDMARD and/or tsDMARD.

The interventions of interest are targeted therapies (i.e., bDMARDs and tsDMARDs) approved by either the European Medicines Agency or the U.S. Food and Drug Administration, to date, treating IA conditions in adult populations; this pharmacological distinction will also be used for separate stratified analyses. These therapies will include bDMARDs: abatacept, adalimumab, anakinra, certolizumab, etanercept, golimumab, guselkumab, infliximab, ixekizumab, rituximab, sarilumab, secukinumab, tocilizumab and ustekinumab, and tsDMARDs: apremilast, baricitinib, filgotinib, tofacitinib and upadacitinib.

Only RCTs included in an existing systematic review of patients with IA (RA, PsA and AxSpA (including AS)) will be considered for eligibility. We will exclude reviews (i.e., not look for trials) that have been withdrawn.
Information sources and search strategy
We will search MEDLINE (via PubMed) for eligible meta-analyses or systematic reviews of trials published since April 2015 and eligible Cochrane reviews using the search algorithm shown in box 1.

Study selection
Two independent reviewers (JB and EM, with support from SMN/RC) will screen the systematic reviews based on title and abstract, in accordance with eligibility criteria. The same two reviewers will assess the full systematic review texts for eligibility of the reviews and subsequently select the RCTs from the reviews that are eligible, according to our objectives. We will obtain the full text if at least one of the reviewers considers an RCT to be potentially eligible during the screening process. Disagreements will be resolved by consensus or by consulting a third reviewer (RC). EndNote VX9.2 software will be used to manage the reviews and RCT records retrieved from the search.

Data collection process
Two reviewers (JB and EM) will extract data using a predefined, standardised data extraction form, and in case of uncertainty, a third reviewer (SMN or RC) will be consulted. If a trial is included in more than one review, the trial will be registered and counted/included once.

Review level
From the reviews, we will extract data on review registration number, year of publication, first author’s name, number of RCTs eligible for our study, and the condition and intervention studied, according to the review title.

Randomised trial level
From the trials, we will extract data on the first author’s name, publication identification (ID), trial duration, duration until switch (eg, relevant for adaptive trial designs), rescue, early escape or crossover, number of participants in each arm, and treatment given in the active and comparator arms (ie, following the PICO framework). Treatment in the comparator arm will be grouped into the following predefined categories: (1) placebo; (2) standard care or waiting list/no intervention; (3) active comparator; or as an ultimate last option, (4) unclear. Outcome measures and data extraction of trial and participant characteristics are specified as follows and listed in box 2.

Participant characteristics in individual studies
Data on the following trial eligibility criteria will be extracted

- Binary indicator (0 or 1) of inflammation scored as 1 with ≥1 of the following criteria: erythrocyte sedimentation rate (ESR) ≥28 mm/hour, CRP level ≥0.3 mg/dL, and/or morning stiffness lasting ≥45 min.
- Minimum and maximum required number of swollen joint counts (SJC)s.
- Minimum and maximum required number of tender joint counts (TJC)s.
- Minimum and maximum allowed disease duration.

- Rheumatoid factor (RF) and/or anticyclic citrullinated peptide antibody (CCP) and/or anticyclic citrullinated peptide antibody 2 (CCP2) status.
- Many published reviews include some form of RoB assessment (ie, at least for the most frequently used RoB domains) for each individual trial; these assessments will be mapped to the best of our ability to correspond to the Cochrane RoB V.1.35 The RoB will also be assessed for at least one previous bDMARD or tsDMARD.

Data on the following aggregate (average/median) patient baseline characteristics will be extracted
- Age, female (proportion), disease duration (years), ESR (mm/hour), CRP (mg/dL), Disease Activity Score, RF positive (proportion), anti-CCP positive (proportion), anti-CCP2 positive (proportion), SJC, TJC, Health Assessment Questionnaire–Disability Index, physician global assessment (eg, Visual Analogue Scale (VAS)), patient global assessment (eg, VAS), and patient-reported pain on VAS of 0–100 mm (eg, VAS).

Rob and trial characteristics in individual studies
Many published reviews include some form of RoB assessment (ie, at least for the most frequently used RoB domains) for each individual trial; these assessments will be mapped to the best of our ability to correspond to the Cochrane RoB V.1.35 The RoB will also be assessed by one of the two reviewers (JB or EM) and subsequently...
compared with the original bias and internal validity assessments in the review, from which it was sampled.

Within each full-text trial report, we will apply the following domains of the original Cochrane RoB tool (ie, RoB V.1.0), which comprise methods for:

- Sequence generation/maintaining allocation concealment (selection bias).
- Blinding both patients and personnel (performance bias).
- Management/analysis/reporting of incomplete outcome data (attrition bias).
- Each of these three domains will be rated as high risk, low risk or unclear RoB. In case of uncertainty, another reviewer (SMN/RC) will be consulted. Any discrepancy between RoB assessment sources will be resolved by discussion among the authors.

The following additional bias sources will also be assessed, and data will be extracted from them:

- Single versus multisite trials.
- Small versus large trials (where small will pragmatically be defined as <100 patients per arm).
- Source of funding grouped into one of the following five categories: (1) 100% industry (pharmaceutical/device company) funded; (2) mixed funding (eg, non-industry and industry); (3) provision of drug only; (4) 100% non-profit funded; and (5) unclear/undisclosed funding.

Outcome measures

Data extraction for each RCT will include the number of patients who died and the numbers of patients with any of the following ‘generic events’: (1) all WDs, (2) WDs due to AEs and (3) SAEs. If the number of individual patients with SAEs is not reported, we will extract the number of reported SAEs instead.

Where reasons are provided to explain the SAEs, they will be categorised according to MedDRA V.23.0: (1) infections and infestations; (2) neoplasms benign, malignant and unspecified (including cysts and polyps); (3) cardiac and vascular disorders; or (4) others. If patients in the comparison group later are offered the intervention (ie, switching to the intervention group), the number of events before the switch will be used as the endpoint of choice (ie, before introducing adaptive trial designs).

Summary measures

For each binary outcome, we will extract data corresponding to a 2×2 table, summarising the number of patients who experienced the outcome in each comparison group as reported in the randomised trial and the total number of patients randomly assigned in each group. For outcomes collected and reported corresponding to different time points, we will select the time point with the longest follow-up while still respecting the primary research design (eg, before introducing an adaptive trial design, and/or open label extension).

Because all the outcome measures are dichotomous, the relative risk ratios (RRs) with 95% CIs will be the preferred measure of relative effect, and these can subsequently be applied to the baseline or control group risks to generate absolute risks. When appropriate for the very rare adverse events, we will use a continuity correction (adding to all cells a factor proportional to the reciprocal of the size of the contrasting study group), as suggested by Sweeting and colleagues, to take into account zero cell counts in one group only. This continuity correction will be applied when no events are observed in one study arm of a trial. The correction is inversely proportional to the relative size of the opposite of the study. Continuity correction for the experimental intervention arm is 1/(R+1); R is the ratio of control group to intervention group sizes (ie, R=n_c/n_i). Similarly, the continuity correction for the control comparator arm is R/(R+1).

We expect that most trials will report only a few SAEs and deaths, so the ORs and 95% CIs will also (for the purpose of sensitivity analysis) be calculated with the use of the Peto method. Because the expected events are sparse and all trials will have similar durations of follow-up for their treatment groups, the use of both RRIs and ORs should represent a valid approach to assessing the ‘risk’ associated with the use of synthetic or biological intervention in patients with IA. The number of studies to be included will be determined by the number of eligible trials available in the systematic reviews; hence, a formal sample size calculation would not be meaningful. However, we expect the search to bring around 175 eligible trials (RA: 117, PsA: 15, AxSpA: 43) based on a few large systematic reviews within the field.

Synthesis of results

Outcome events will be coded so that an RR direction of more than one indicates a potentially harmful effect of the experimental intervention (ie, RR=[r_c/n_c]/[r_i/n_i]). Mixed effects restricted maximum likelihood (REML) meta-analyses will be used to combine the harm effects across RCTs (based on their log(RR)); we will apply the trial ID as a random effect while applying the review ID (from which the trial was sampled) as a fixed effect in order to model the hierarchical structure of the data sampling. Heterogeneity across randomised comparisons will be assessed by using the Cochrane Q test, interpreted based on the I² inconsistency index, and quantified by the estimate for between-trial variance (τ²), estimated as T². The effect of population and trial characteristics (listed further) on the between-trial variance will be calculated by univariately adding a fixed factor for the specific characteristic in the model. If introducing a specific covariate into the metamodel reduces the observed between-trial variance (T²), this result will be considered an indication of a potentially important effect modifier.

Further analyses might (if possible) include multivariable models. Observational by nature, metaepidemiological studies like the present should be expected to have some degree of metaconfounding because prognostic factors might be unequally distributed between...
studies exposed (trial characteristic positive) or not. We will attempt to adjust for possible confounding covariates (ie, ‘deconfound’) the inference; that is, we will explore the ability to adjust for any important pre-exposure covariate C (ie, something that happened before the ‘study design variable of interest was generated’) that is potentially unequally distributed between the metaepidemiological exposure groups (E: exposed vs unexposed), which is also an ancestor (ie, likely cause) of the trial outcome (ie, a more exaggerated risk of harm): $M_k \leftarrow C \rightarrow Y_k$.

Power and sample size considerations

Although the number of trials being eligible in the synthesis is fixed based on the premises, there are several factors that can influence the statistical power in meta-analyses, such as the total number of eligible RCTs ($k$), their individual sample sizes ($N_1, N_2, \ldots, N_k$) and the number of harm events observed in each of the trials ($\epsilon_1, \epsilon_2, \ldots, \epsilon_k$). Since we want to study harmful effects of various experimental interventions, even very small effect sizes would be informative. Substantial heterogeneity—between exposed versus unexposed subgroups—will affect the precision of our meta-analytic estimates, and thus our potential to find significant differences between strata. Thus, on an ad hoc basis, we performed some analyses exploring how large the difference between two strata had to be in order for us to be able to detect it, given the expected number of studies at our disposal (say, at least 200 trials across conditions).^{8,29,57,58} An ad hoc subgroup power analysis was conducted in R using the power.analysis.subgroup function, based on the approach described by Hedges and Pigott:^{59,60} assuming that we want to compare two independent ORs, $OR_1=2$ vs $OR_2=1$, with a conservative guestimate of the SE of the log(OR) values of 0.15. We would have a good statistical power to detect a difference between strata (90.4%). The output also tells us that, all else being equal, the effect size difference ($\log(OR_1)$ vs $\log(OR_2)$) needs to be at least 0.595 in order to reach sufficient power.

### Box 3 Risk of bias

<table>
<thead>
<tr>
<th>Cochrane risk of bias domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of selection bias.</td>
</tr>
<tr>
<td>Risk of performance bias.</td>
</tr>
<tr>
<td>Risk of attrition bias.</td>
</tr>
<tr>
<td>Overall risk of bias.*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional bias sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single versus multisite trials.</td>
</tr>
<tr>
<td>Small versus large trials.</td>
</tr>
<tr>
<td>Source of funding.</td>
</tr>
</tbody>
</table>

*For each trial, the overall RoB will be classified as low (ie, low RoB for all three domains), high (ie, high RoB for one or more domains) or unclear (ie, unclear RoB for one or more domains in the absence of high RoB).

RoB across studies

Stratified meta-analyses will be used for tests of interaction between harms and the trial RoB (listed in box 3) and trial characteristics collected as described previously.

**DISCUSSION**

The Outcome Measures in Rheumatology initiative established the Contextual Factors Working Group to guide the understanding, identification and handling of contextual factors for clinical trials, with most of the current emphasis being on net benefit inferred from a rheumatology trial.^{62}

This metaepidemiological study will work from the original definition of what constitutes a contextual factor, as defined in the introduction. We will explore and hope to reveal the possible impact of contextual factors (ie, population and trial characteristics) and RoB domains of three different (but related) harm measures, as well as deaths.^{63}

Knowing which factors are associated with a causal model for harms—either as effect modifiers or distortions of the outcome due to bias—is important for improving investigation and reporting of harms in future trials.^{64} If future trials report harms according to important contextual factors, meta-analyses would be able to investigate contextual factors for harms without relying on getting access to individual patient data. Such metaresearch might identify subgroups among patients with rheumatic disease that are at higher risk of experiencing harms. Such information would provide important evidence for future treatment guideline development. Ultimately, the doctor will be able to differentiate the risk of intervention based on the patient’s characteristics. This has the potential to enhance informed decision-making and effect therapeutic interventions applied in practice, leading to safer treatment of individual patients and increase efficiency in the healthcare system.

This study has several strengths. This will be a comprehensive risk of harm analysis of bDMARDs and tsDMARDs across multiple IA diagnoses unlike previous assessments. The study selection, data extraction and RoB assessment will be done by two reviewers; the data extraction will be extensive as it will include data on many different contextual factors and RoB items. Finally, we will only include RCTs to avoid inherent problems when investigating potential effect modifiers from non-RCT evidence.

Nevertheless, several limitations should be considered. First, the analyses may be limited by poor reporting, and data are likely to be lacking for certain contextual factors (ie, exposures) and uncommon harms (ie, outcomes). Second, by relying on the search results of published reviews, we assume that we retrieve a representative sample for our analysis; however, it is possible that potentially useful trials may be omitted, for example, due to very specific eligibility criteria in the reviews or very recent trials that have not yet been included in a systematic review, which could theoretically influence our results. Third, other characteristics that we do not consider may be relevant. Fourth, metaregression assessments,
despite including only RCTs, are observational by nature and their results may by themselves be confounded by other factors. Finally, the metaregression assessments involving aggregated data may be affected by ecological fallacy and could potentially be misleading. Patient and public involvement

Patients and/or the public were not involved in designing study concept or drafting of the protocol.

ETHICS AND DISSEMINATION

Because our study does not collect primary data, no formal ethical assessment and informed consent are required. The study will be disseminated in a peer-reviewed publication.

Author affiliations

1Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
2Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark
3Department 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
4Division of Rheumatology, Department of Medicine, University of California at Los Angeles, LA, CA, USA
5Division of Rheumatology, Department of Medicine, Medical University of Vienna, Vienna, Austria
6Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Acknowledgements

The Parker Institute is grateful for the financial support received from public and private foundations, companies and private individuals over the years.

Contributors

Contributions: study concept and design and drafting of the protocol: EM, JB, SMN, TE, LEK, RC, JPN, ST, DF, JSS and PCT; critical revision of the protocol for important intellectual content and final approval before submission: all authors; guarantor of the protocol: RC; obtained funding: TE and RC.

Funding

The Parker Institute, Bispebjerg and Frederiksberg Hospital, is supported by a core grant from the Oak Foundation (OCWF-18-774-QF0L). The Oak Foundation is a group of philanthropic organisations that, since its establishment in 1983, has given grants to not-for-profit organisations around the world.

Disclaimer

The views expressed in the submitted protocol are the authors’ own and not an official position of the institution or funder.

Competing interests

None declared.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication

Not required.

Provenance and peer review

Not commissioned; externally peer reviewed.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Ebbe Rasmussen http://orcid.org/0000-0002-8016-6646
Sabrina Mai Nielsen http://orcid.org/0000-0003-2857-2484
John P.A. Ioannidis http://orcid.org/0000-0003-3118-6859
Peter C. Taylor http://orcid.org/0000-0001-7766-6167
Torkell Ellingsen http://orcid.org/0000-0003-0426-4962

REFERENCES

1 Ledingham J, Snowden N, Ide Z. Diagnosis and early management of inflammatory arthritis. BMJ 2017;358:j3248.
25 Rahman MU, Buchanan J, Doyle MK, et al. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for...