Seminars in Arthritis and Rheumatism 51 (2021) 607–617

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

# Harms reported by patients in rheumatology drug trials: a systematic review of randomized trials in the cochrane library from an OMERACT working group

Dorthe B. Berthelsen[a,b], Thasia G. Woodworth[c], Niti Goel[d], John P.A. Ioannidis[e], Peter Tugwell[f], Dan Devoe[g], Paula Williamson[h], Caroline B. Terwee[i], Maria E. Suarez-Almazor[j], Vibeke Strand[k], Amye L. Leong[l], Philip G. Conaghan[m,n], Maarten Boers[o], Beverley J. Shea[p], Peter M. Brooks[q], Lee S. Simon[r], Daniel E. Furst[c,#], Robin Christensen[a,s,#,*], on behalf of the OMERACT Safety Working Group

[a] *Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark*
[b] *Department of Rehabilitation, Municipality of Guldborgsund, Nykoebing F, Denmark*
[c] *David Geffen School of Med. Division Rheumatology. UCLA, USA*
[d] *Duke University School of Medicine, Durham, NC, USA*
[e] *Departments of Medicine, Epidemiology and Population Health, Biomedical Data Science and Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, California, USA*
[f] *Department of Medicine, School of Epidemiology, Public Health and Community Medicine, University of Ottawa, Canada*
[g] *Department of Medicine, University of Calgary, Cumming School of Medicine, Calgary, Canada*
[h] *MRC North West Hub for Trials Methodology Research, Institute of Translational Medicine, University of Liverpool, Liverpool, UK*
[i] *Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics and the Amsterdam Public Health Research Institute, Amsterdam, NL*
[j] *Department of Health Services Research and Section of Rheumatology and Clinical Immunology, University of Texas MD Anderson Cancer Centre, Houston, TX, USA*
[k] *Division of Immunology/Rheumatology, Stanford University, Palo Alto CA, USA*
[l] *Healthy Motivation; Global Alliance for Musculoskeletal Health, Bone and Joint Decade, Santa Barbara, California USA*
[m] *Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK*
[n] *NIHR Leeds Biomedical Research Centre, Leeds, UK*
[o] *Department of Epidemiology and Biostatistics and the Amsterdam Rheumatology and immunology Centre, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, NL*
[p] *Ottawa Hospital Research Institute, Clinical Epidemiology Program and School of Epidemiology and Public Health, University of Ottawa, Canada*
[q] *Centre for Health Policy Melbourne School of Population and Global Health University of Melbourne and Northern Health, Australia*
[r] *SDG LLC Cambridge, MA 02138, USA*
[s] *Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark*

ARTICLE INFO

*Keywords:*
Harms
Adverse events
Core Outcome Set
Rheumatology
OMERACT

ABSTRACT

*Background:* Underreporting of harms in randomized controlled trials (RCTs) may lead to incomplete or erroneous assessments of the perceived benefit-to-harm profile of an intervention. To compare benefit with harm in clinical practice and future clinical studies, adverse event (AE) profiles including severity need to be understood. Even though patients report harm symptoms earlier and more frequently than clinicians, rheumatology RCTs currently do not provide a reporting framework from the patient's perspective regarding harms. Our objective for this meta-research project was to identify AEs in order to determine harm clusters and whether these could be self-reported by patients. Our other objective was to examine reported severity grading of the reported harms.
*Methods:* We considered primary publications of RCTs eligible if they were published between 2008 and 2018 evaluating pharmacological interventions in patients with a rheumatic or musculoskeletal condition and if

*Abbreviations:* ACR, American College of Rheumatology; AE, adverse event; EULAR, European League Against Rheumatism; CDSR, Cochrane database of systematic reviews; CMSG, Cochrane Musculoskeletal Group; COMET, Core Outcome Measures in Effectiveness Trials; CONSORT, consolidated Standards of Reporting Trials; COS, core outcome set; CRs, Cochrane reviews; DMARDs, disease-modifying antirheumatic drugs; IQR, interquartile range; MedDRA, Medical Dictionary for Regulatory Activities; NSAIDs, nonsteroidal anti-inflammatory drugs; OARSI, OsteoArthritis Research Society International; OMERACT, Outcome Measures in Rheumatology; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRO-CTCAE, Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; PROSPERO, International prospective register of systematic reviews; RCT, randomized clinical trials; RCTC, Rheumatology Common Toxicity Criteria; SD, standard deviation

* Corresponding author.
*E-mail address:* Robin.Christensen@regionh.dk (R. Christensen).
# Shared last author.

https://doi.org/10.1016/j.semarthrit.2020.09.023

they were included in Cochrane reviews. We extracted data on harms such as reported AE terms together with severity (if described), and categorized AE- and severity-terms into overall groups. We deemed all AEs with felt components appropriate for patient self-reporting.
*Results:* The literature search identified 187 possible Cochrane reviews, of which 94 were eligible for evaluation, comprising 1,297 articles on individual RCTs. Of these RCTs, 93 pharmacological trials met our inclusion criteria (including 31,023 patients; representing 20,844 accumulated patient years), which reported a total of 21,498 AEs, corresponding to 693 unique reported terms for AEs. We further sub-categorized these terms into 280 harm clusters (i.e., themes). AEs appropriate for patient self-reporting accounted for 58% of the AEs reported. Among the reported AEs, we identified medical terms for all of the 117 harm clusters appropriate for patient reporting and lay language terms for 86%. We intended to include severity grades of the reported AEs, but there was no evidence for systematic reporting of clinician- or patient-reported severity in the primary articles of the 93 trials. However, we identified 33 terms suggesting severity, but severity grading was discernible in only 9%, precluding a breakdown by severity in this systematic review.
*Conclusions:* Our results support the need for a standardized framework for patients' reporting of harms in rheumatology trials. Reporting of AEs with severity should be included in future reporting of harms, both from the patients' and investigators' perspectives.
*Registration:* PROSPERO: CRD42018108393



## Introduction

Balanced adequate reporting of harms, as well as benefits, of an intervention in randomized controlled trials (RCTs) and future research is essential to allow patients and clinicians to make the most appropriate treatment decisions concerning a specific intervention [1]. However, the reporting of harms (adverse events, AEs) in studies of health care interventions is typically less comprehensive than that of benefit (efficacy) [2–4]. Further, regional differences in reporting of harms may reflect underreporting of AEs as well [5]. Such underreporting may lead to incomplete or erroneous judgments on the benefit-to-harm profile of an intervention [2,6]. Even though the harm extension of *Consolidated Standards of Reporting Trials* (CONSORT) statement provides guidance on items to include when reporting harms in RCTs [7], the quality of reporting RCTs in the literature is poor based on examination of articles published in high impact-factor journals in general medicine and rheumatology [8].

*Outcome Measures in Rheumatology* (OMERACT) is an independent international organization of health care professionals and patient research partners, which strives to improve outcome measurement and instrument methodology in studies assessing rheumatology treatments. Beginning in 1992, OMERACT has developed Core Outcome Sets (COS) for many rheumatologic conditions [9,10] and has actively involved patients since 2002 [11]. A COS is a minimum consensus-based set of outcome domains that should be measured and reported in all RCTs and longitudinal observational studies of a specific health condition and/or intervention [12]. OMERACT uses the term 'Core Domain Set' to distinguish it from the 'Core Outcome Measurement Set' that specifies instruments for each of the core domains. Many initiatives other than OMERACT are also establishing COS (see e.g. the *Core Outcome Measures in Effectiveness Trials* [COMET] database) [13], and although it is recommended that COS or systematic reviews covering multiple intervention types should address the potential for AEs, only one-third of COS explicitly call for AEs to be recorded [14]. To correct this apparent oversight, OMERACT recently recommended that benefits *and* harms should be equally and explicitly considered when developing COS [10].

Specifically, we in the OMERACT Safety Working Group aim to improve the guidance on what and how to measure and report harms, explicitly including the patient perspective [15]. Thus, the group developed the Rheumatology Common Toxicity Criteria 2.0 (RCTC 2.0) [16], which encourage standardization of assessment and reporting of AEs in RCTs and longitudinal observational studies in rheumatology. However, the RCTC 2.0 does not provide guidance on how to collect harm information taking into account whether clinicians or patients are in the best position to assess specific AEs. Nevertheless, focusing on the patient perspective to complement the clinician perspective on harms is highly relevant because patients report harm symptoms earlier and more frequently than clinicians [17], and because clinicians tend to systematically downgrade the severity, i.e., the intensity, of patients' symptoms [18–20].

A measurement instrument suitable for assessing and reporting patient perspectives on harms experienced during treatment for rheumatologic conditions is lacking [21], but such instruments have been developed in other conditions e.g., the Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE) within oncology [22]. To address this need in rheumatology and to identify candidate Core Outcome Domains as part of developing a reporting framework for patient-reported harms in rheumatology [23], we provide a systematic review of harms reported in primary publications of RCTs published between 2008 and 2018 included in Cochrane reviews. The results of our systematic review will inform a Delphi process. Our primary objective for this meta-research project was to identify all harm domains reported in those RCTs of pharmacological interventions in rheumatic and musculoskeletal conditions evaluated in systematic reviews by the Cochrane Musculoskeletal Group (CMSG), in order to determine if we could identify harm clusters appropriate to be self-reported by patients. Our other objective was to examine reported severity grading of the identified harms.

## Methods

We registered the study protocol on the international prospective register of systematic reviews (PROSPERO: CRD42018108393) and report our findings according to the guidance in Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) statement [24], with additional guidance of knowledge synthesis from PRISMA Extension for Scoping Reviews (PRISMA-ScR) when feasible [25].

### *Data sources and searches*

We searched the Cochrane Database of Systematic Reviews (CDSR) for all harms reported in RCTs of pharmacological interventions. Cochrane reviews (CRs) examine large numbers of trials and are recognized to be thorough in searching for eligible studies [26]. Thus, by searching CRs, we obtained a broad sampling across rheumatology indications, as well as industry and non-industry sponsored trials. Using the website https://www.cochranelibrary.com, we browsed by Cochrane Review Group, selecting Musculoskeletal (across all years available), limiting Type by intervention and Topics by Rheumatology. We conducted our search on 16 October 2018.

*Study selection*

Two reviewers (DBB supported by RC) screened all identified CRs by reviewing titles and abstracts. We excluded protocols without data and withdrawn reviews. We then used reference lists of included articles in the selected CRs to identify eligible rheumatology trials. Trials were eligible if they investigated any type of pharmacological intervention against any comparator(s) in patients with rheumatic and musculoskeletal conditions. We identified primary publications from the reference lists of the included reviews (i.e., referred to as major publications in CRs), and excluded manuscripts/reports of unpublished data and publications that were not journal articles. We removed articles not written in English and article duplicates; for practical reasons we included only articles published between 2008 and 2018.

*Data extraction*

We used a standardized data extraction form to collect information from eligible trials. At review level, we extracted CR-registration number, author, year of publication, and rheumatic or musculoskeletal condition. At trial level, we assigned all trials an ID and extracted data on author, year of publication, condition, intervention, trial duration (i.e., duration for reported harms), funding source, surveillance method for AEs, sample size (i.e., total number of patients randomized), number of completers of the trial, number of withdrawals, and number of withdrawals due to AEs. When not explicitly reported, we estimated total patient-years per trial of exposure by assuming a linear dropout rate between baseline and end of the trial period (i.e., the area under the curve) [27]. Further, we extracted patient characteristics i.e., participants' age, weight, BMI, sex (number of included women), and disease duration.

We categorized type of condition by topic categories of conditions in the CMSG library. Interventions were categorized according to American College of Rheumatology (ACR), European League Against Rheumatism (EULAR), and Osteoarthritis Research Society International (OARSI) recommendations and guidelines [28–39]. Categories included comparator interventions: placebo/sham, usual care/no intervention, and active treatment (such as non-pharmacological interventions). Trial duration was categorized as <27 weeks (short), 27–52 weeks (intermediate), or >52 weeks (long-term). Funding source was categorized as industry-sponsored (for any industry involvement in funding or any role in design, conception, analysis, and reporting of the trial); non-industry sponsored; neutral (such as industry's providing the study drug with no other role); or unclear. Further, we categorized surveillance of AEs as active (e.g., when the method of collecting harms was based on systematic recording at each follow up), passive, or unclear.

For each trial, we (DBB and RC) extracted all AEs by the reported term presented in the article and tabulated the number of reports for each AE. From each article, we extracted harm information from tables and supplemented by description in the main text in the most specific way for each AE. I.e., we only extracted domains of AEs, such as "musculoskeletal and connective tissue signs and symptoms" if no specific AEs (e.g., "myalgia") were mentioned. For each reported AE, we extracted the verbatim severity of the specific AE if provided in the article. If severity was not clearly described for the specific harm, we extracted overall categories possibly related to severity (e.g., serious AEs, AEs of interest or AEs leading to withdrawal), if reported. When such wording was not available, we implemented a reasonable, consistent, well-defined approach. First, we considered the regulatory definition of a serious AE: results in death; is life threatening; requires inpatient hospitalization or results in prolongation of existing hospitalization; results in persistent or significant disability/incapacity; is a congenital anomaly/birth defect; or is a medically important event or reaction [40]. We then considered previous work in rheumatology [16] and oncology [41], and categorized severity as grades 1-5, rating as follows: mild (1), moderate (2), severe (3), life threatening (4), and death (5). Although it's mandatory to report serious AEs, we modified the regulatory definition and categorized serious AEs as grade 4, because we assumed AEs resulting in death to be reported as so, and because "life-threatening" in the definition of "serious" refers to an event/reaction in which the patient was at risk of death at the time of the event/reaction [40]. We did this to emphasize the patient perspective, which may be different from the regulatory approach and less clear but, in our view, is just as important. To ensure the patient's perspective in this process, we included patients among the reviewers. We avoided double counting (e.g., severity reported as "AEs of interest, serious infection" counted only as serious infection).

*Data analysis*

DBB organized the extracted data in a customized spreadsheet enabling analysis in collaboration with TGW and DEF. Two reviewers (DBB and TGW) identified overall terms covering the same severity (e.g., "mild" would include "mild adverse events" and "mild in nature") and overall AE terms covering the same type of harm (e.g., "abdominal abscess" would include "abdominal wall abscess" and "peridiverticular abscess"). We also categorized the severity of each of these harm clusters as mild, moderate, severe, life threatening or fatal. If a group of extracted AEs fell into the same harm cluster, but none of the AE terms was appropriate as the overall term for the harm cluster, we added an appropriate term (e.g., the overall term "antibodies to biologics" was used to cover related terms such as "antibodies to certolizumab pegol," "antibodies to golimumab," and "antibodies to pegloticase").

Referring to the OMERACT filter 2.1 (Supplementary Fig A.1 and A.2), the two reviewers (DBB and TGW) independently also categorized each cluster of harms under one of the three areas (that is *life impact* [e.g., patient perception of health or quality of life]; *pathophysiologic manifestations* [e.g., body function and structure or biomarkers and surrogate measures that accompany a condition]; and *death*) [9,10]. Area of life impact included harm clusters most likely to be felt and reported by the patients (such as nausea and diarrhea), whereas the area of pathophysiologic manifestations included harm clusters most likely to be observed/measured and reported by clinicians (such as neutropenia or peripheral vascular disease). Further, each harm cluster's appropriateness for patient self-reporting was categorized according to being best assessed from an internal (patient) view when the AE is mostly felt (previous referred to as "subjective" [such as headache or nausea]); best assessed from a mixed (patient/clinician) view when the AE is mostly felt with observed components (such as vomiting or constipation) and mostly observed with felt components (such as rash or fever); and best assessed form an external (clinician) view when the AE is mostly observed (Fig. 1). For the last category, we distinguished clinically/measurable observable (such as pneumonia or abdominal abscess) and laboratory/biomarker-based (such as hyperlipidemia or increases in liver transaminase levels) [22]; we deemed the external category as harms inappropriate for patient self-reporting. Our categorization allowed that a patient would report AEs with a degree of observable components, as the patient might still be in the best position to report these as a patient reported outcome.

Harm clusters were then mapped into categories of system organ classes according to the OMERACT Rheumatology Common Toxicity Criteria v. 2.0 (RCTC) [16]. When the RCTC 2.0 did not list clusters in any category, we mapped the clusters into an RCTC-category considered relevant for the specific cluster. Finally, we added a lay language term and a medical term to each harm cluster. We used the overall term for the harm cluster as either the lay language or the medical term; if none of the extracted AE terms were appropriate for the lay language term, we added a synonym if it was evident (e.g., joint pain was added as a lay language term for arthralgia). We resolved discrepancies between the two reviewers through discussion. In case of

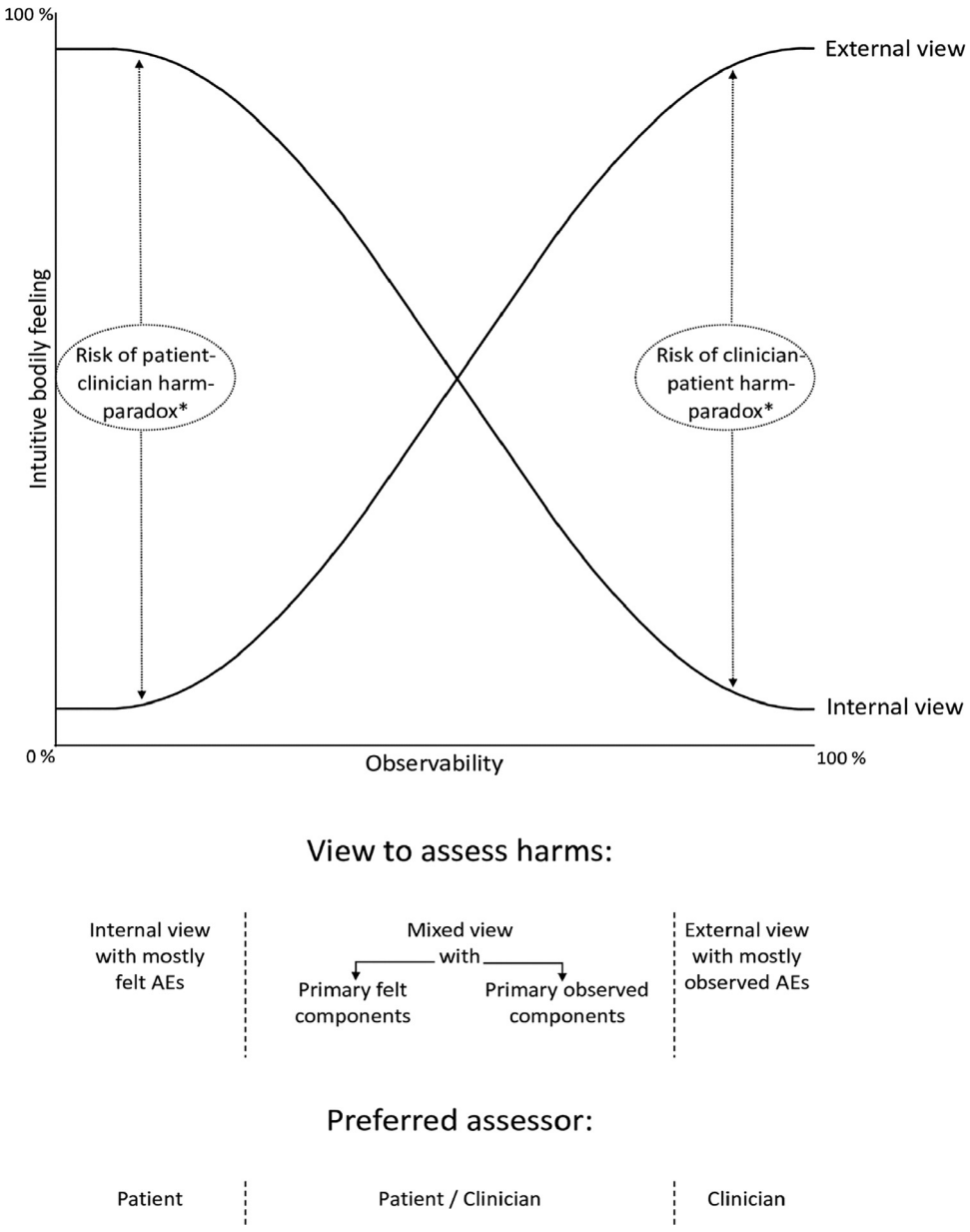


**Fig. 1.** Perspective on outcome assessment to cover harms. *Harm-paradoxes occur when harms appear unequally important/severe when observed from two different points of view.

uncertainty, we consulted a third reviewer (NG, DEF or RC). To ensure that the study objectives were assessed from patient's point of view, we included patients among the reviewers.

### *Statistical analysis*

We present descriptive statistics for categorical variables of trial characteristics using counts and proportions. For continuous variables, we reported mean (±SD) or medians (with interquartile ranges [IQRs]) as appropriate.

Agreement between the two reviewers assessing harms appropriateness for patient reporting was estimated (by unweighted Cohen's *k*-statistic) in terms of dichotomous assessment (i.e., harms appropriate for patient self-reporting or harms not appropriate for patient self-reporting) and interpreted according to Landis and Koch [42]: *k* values of <0 were considered poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 almost perfect agreement.

## Results

### *Eligible reviews and trials*

As presented in Fig. 2, our search retrieved 187 Cochrane reviews. We excluded protocols, and after screening titles and abstracts, we excluded reviews not including RCTs. This process narrowed the field to 94 eligible Cochrane reviews, encompassing 1,297 potentially eligible articles, from which we identified 98 eligible articles with 96 distinct RCTs. We excluded three trials that did not examine rheumatologic conditions, yielding a total of 93 trials included in the final analysis (Supplementary Table A).

### *Characteristics of included trials*

The reviewed trials included 31,023 participants, representing 20,844 patient years. Patients' mean (SD) age was 54 [7], disease duration 7 [4] years, weight 84 [17] kilos with a BMI of 31 [4] and

59% of patients were female. Table 1 shows that most participants suffered from rheumatoid arthritis (45%), osteoarthritis (26%) and gout (22%). The most commonly studied active interventions were biologic DMARDs (bDMARDs), as monotherapy or in combination with conventional synthetic DMARDs (csDMARDs) (23%); and urate-lowering therapy (16%). Placebo or sham interventions (13%) were the most commonly used comparators. Overall, 7,280 (24%) participants withdrew from the trials with 1,777 (6%) withdrawing due to AEs.

On an individual trial level, the median sample size in the included trials was 164 (IQR 26-499) participants; the median trial duration was 24 weeks (IQR 12–52) with 60 (65%) trials of less than 27 weeks', 19 (20%) of 27–52 weeks', and 14 (15%) of more than 52 weeks' duration. In 52 trials (56%), investigators used active surveillance of harms, whereas surveillance was passive in one trial (1%). In 40 trials (43%), the method of surveillance was unclear. Most trials (61 [66%]) were industry-sponsored; 14 (15%) were non-industry-funded; and funding sources were unclear or neutral in 14 (15%) and 4 trials (4%), respectively.

*Harms reported in rheumatology drug trials*

In the 93 included trials, we identified 21,498 reported AEs, covering 693 unique reported terms for AEs (Supplementary Table B). By categorizing these 693 terms into overall groups covering the same harms, we narrowed the field to 280 harm clusters. Most of the harm clusters were within the core area of pathophysiological manifestations: 194 (69%); fewer were in the areas of life impact: 85 (30%) or death: 1 (<1%).

Among the 280 harm clusters, we judged 117 (42%) to be appropriate for patient self-reporting: 29% mostly felt, 16% mostly felt with observed components, and 13% mostly observed with felt components.

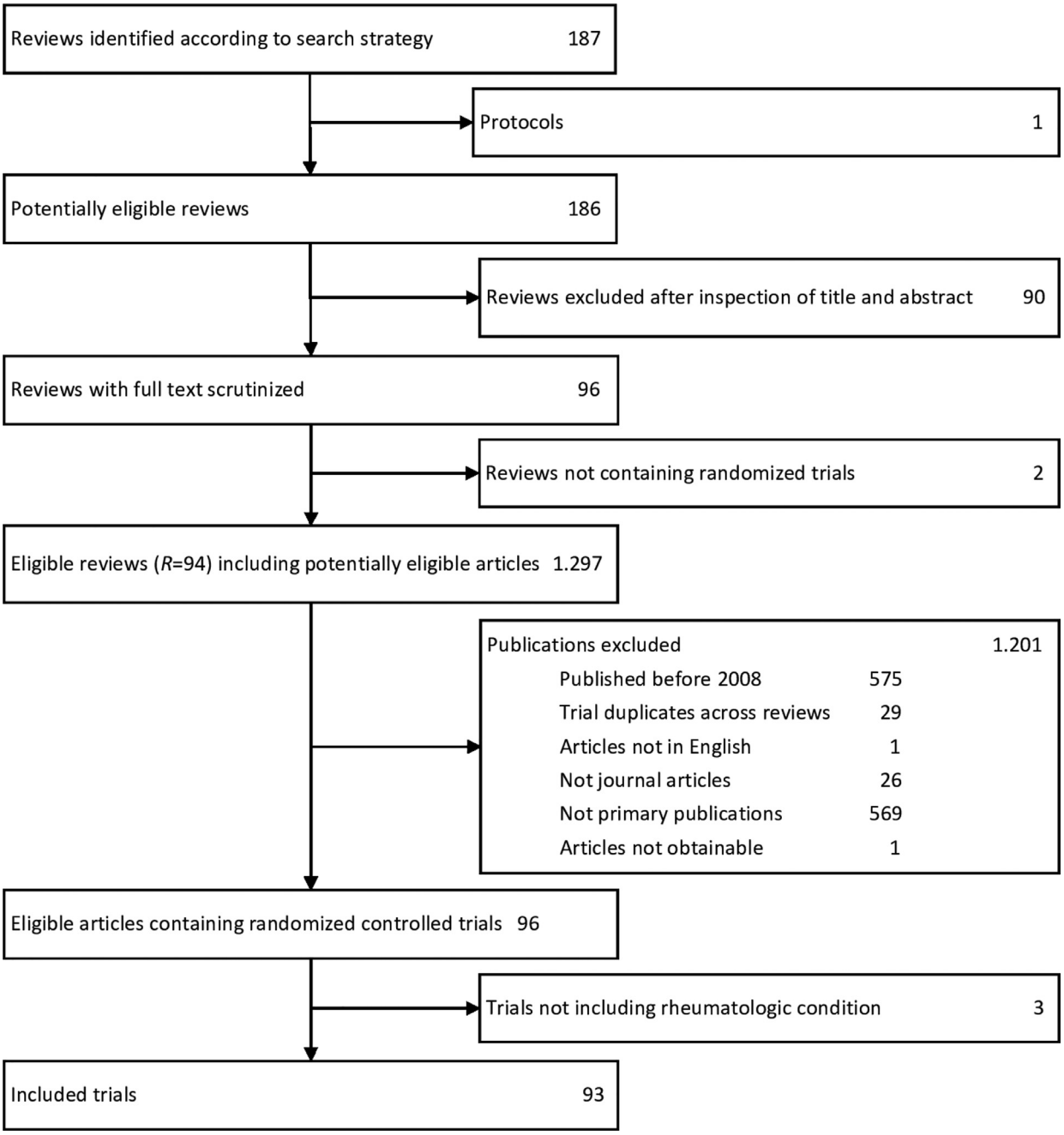


**Fig. 2.** Flow diagram for the study selection. *R* = review.

**Table 1**
Characteristics of included trials (*k*=93).

| | Trials, *k* | % | Patients, *n* | % |
|---|---|---|---|---|
| Total | 93 | 100 | 31,023 | 100.0 |
| **Condition**[*] | | | | |
| Rheumatoid arthritis | 29 | 31 | 13,897 | 44.8 |
| Osteoarthritis | 32 | 34 | 8,147 | 26.3 |
| Gout | 16 | 17 | 6,823 | 22.0 |
| Spondyloarthropathy (incl. PsA and AS) | 6 | 6 | 1,252 | 4.0 |
| Soft tissue disorders | 4 | 4 | 252 | 0.8 |
| Mixed | 2 | 2 | 240 | 0.8 |
| Osteoporosis | 1 | 1 | 173 | 0.6 |
| Lupus erythematosus | 2 | 2 | 138 | 0.4 |
| Fibromyalgia | 1 | 1 | 101 | 0.3 |
| **Intervention**[†] | | | | |
| bDMARDs + csDMARDs | | | 7,228 | 23.3 |
| Urate-lowering therapy | | | 5,097 | 16.4 |
| Placebo/sham | | | 3,941 | 12.7 |
| csDMARDs + placebo | | | 3,198 | 10.3 |
| bDMARDs | | | 2,868 | 9.2 |
| Nutraceuticals | | | 2,530 | 8.2 |
| Opioids | | | 1,840 | 5.9 |
| NSAIDs | | | 1,233 | 4.0 |
| Glucocorticoid and intraarticular hyaluronate | | | 795 | 2.6 |
| csDMARDs | | | 521 | 1.7 |
| Other pharmacological interventions[‡] | | | 107 | 0.3 |
| Colchicine | | | 385 | 1.2 |
| bDMARDs + placebo | | | 367 | 1.2 |
| Antiresorptive and osteoanabolic drugs | | | 234 | 0.8 |
| Active treatment[§] | | | 115 | 0.4 |
| Other combination of interventions[‖] | | | 107 | 0.3 |
| NSAIDs + placebo | | | 76 | 0.2 |
| bDMARDs + NSAIDs | | | 74 | 0.2 |
| Usual care/no intervention | | | 22 | 0.1 |
| **Sample size, median (IQR)** | | | 164 (26-499) | |
| **Funding source** | | | | |
| Industry sponsored | 61 | 66 | | |
| Non-industry funded | 14 | 15 | | |
| Unclear | 14 | 15 | | |
| Neutral | 4 | 4 | | |
| **Surveillance of harms** | | | | |
| Active | 52 | 56 | | |
| Passive | 1 | 1 | | |
| Unclear | 40 | 43 | | |
| **Trial duration** | | | | |
| <27 weeks | 60 | 65 | | |
| 27-52 weeks | 19 | 20 | | |
| >52 weeks | 14 | 15 | | |
| **Trial duration (weeks), median (IQR)** | 24 (12-52) | | | |

AS = ankylosing spondylitis; bDMARDs = biologic disease-modifying antirheumatic drugs; csDMARDs = conventional synthetic disease-modifying antirheumatic drugs; IQR = interquartile range; NSAIDs = nonsteroidal anti-inflammatory drugs; PsA = psoriatic arthritis.

[*] Index according to Rheumatology topics in the Cochrane Library.

[†] Categorized according to American College of Rheumatology (ACR), European League Against Rheumatism (EULAR), and Osteoarthritis Research Society International (OARSI) recommendations and guidelines.

[‡] E.g., doxycycline or botulinum toxin.

[§] E.g., acupuncture or exercise therapy.

[‖] E.g., aspiration plus corticosteroid injection plus horizontal therapy or hyaluronate plus exercise.

A total of 58% of the harm clusters were considered mostly observed; i.e., not appropriate for patient self-reporting: 51% clinically/measurable and 7% laboratory-/biomarker-based. Our judgement of whether they were appropriate for patient self-report is presented in Supplementary Fig B Reviewers agreed on 80% of the assessments (kappa = 0.61).

Table 2 shows the 117 harm clusters appropriate for patient self-reporting. However, as it was difficult to achieve consensus, Supplementary Table C explains the reasons for the specific categorization of the harms appropriate for patient self-reporting where ambiguity might exist. From the unique reported terms, we identified or added medical terms describing all of the harm clusters appropriate for patient reporting, although we were only able to identify or add lay language terms for 86% of these clusters. We judged 73% of the harm clusters appropriate for patient self-reporting to be within the core area of life impact, while 27% were within the area of pathophysiological manifestations.

The 117 harm clusters appropriate for patient self-reporting accounted for 58% of the AEs reported in the included trials. As harms not appropriate for patient self-reporting accounted for 42% of the total number of AEs reported, the (rate) ratio of reporting a harm appropriate for patient self-reporting compared with a harm not appropriate for patient self-reporting was 1.41 (95% CI, 1.37-1.44).

### *Severity of harms*

We intended to include severity grades of the reported AEs, but there was no evidence for systematic reporting of clinician or patient-reported severity in the primary articles of the 93 trials. As shown in Table 3, we identified 33 overall terms suggesting severity in the primary articles.

Only 2% of the events described severity in terms of “mild” 326 (2%), “moderate” 1 (<1%) or “severe” 11 (<1%). We further considered 5 (<1%) AEs described as ”slight” to be in the same grade as “mild”. Furthermore, 1280 (7%) of the reported AEs were “life threatening”, while 8 (<1%) fatal events were reported in terms of “adverse events leading to death”. Thus, of 21,498 reported AEs in the included trials, only 9% were broken down by severity in the articles.

### *Harm domains*

When we categorized the 280 harm clusters into system organ classes according to the RCTC 2.0, general 56 (20%), gastrointestinal 41 (15%) and musculoskeletal 36 (13%) were the most used categories (Supplementary Table D). The least used categories were laboratory data: hematology 9 (3%), chemistry 7 (3%), and urinalysis 1 (<1%). However, we lacked categories for mapping 15 (5%) harm clusters (e.g., somnolence, lymphoma, and abdominal wall abscess) into the RCTC 2.0. We found, for example, the following gaps: non-specific terms (e.g., fracture), hyperlipidemia (secondary to AEs associated with interleukin [IL]-6 and Janus kinase [JAK] inhibitors), specific infections (e.g., viral, opportunistic, mycobacterial associated with biologics and JAK inhibitors), and cancer-related terms (e.g., basal cell carcinoma). Further, there were no clear groupings for harms related to the renal system and to reproductive organs.

## Discussion

In our critical review of 93 RCTs in rheumatology, we (DBB, TGW with support from NG, DEF and RC) identified 117 out of a total of 280 harm clusters that could be appropriate for patient self-reporting. These 117 accounted for more than half of AEs reported in the primary publications. Medical terms could describe all harm clusters appropriate for patient reporting whereas lay language terms described 86% of the clusters. The observer- or patient-reported severity was poorly reported for more than 90% of the identified harms. Further, we identified important and frequently reported harms that we could not map as the RCTC 2.0 presently lacks domains such as infections, malignancies, fractures, and neurological terms such as somnolence.

Building on the premise that patients’ and clinicians’ different perspectives on a disease might influence the assessment of effects in RCTs [43], we feel patients should assess harms and their severity when the harm involves “felt” components. Likewise, clinicians should assess harms when “observed” components are involved.

**Table 2**
Harms appropriate for patient self-reporting*

| Mostly felt AEs | | Mostly felt AEs with observed components | | Mostly observed AEs with felt components | |
|---|---|---|---|---|---|
| No of reported harms | Harm cluster† | No of reported harms | Harm cluster† | No of reported harms | Harm cluster† |
| 1138 | Headache | 713 | Diarrhea | 1685 | Upper respiratory tract infection |
| 1038 | Nausea | 649 | Musculoskeletal and connective tissue signs and symptoms (none) | 507 | Injection-site reactions |
| 401 | Dizziness | 597 | Constipation | 327 | Joint-related signs and symptoms (none) |
| 268 | Fatigue | 565 | Nasopharyngitis (*common cold*) | 303 | RA flare |
| 204 | Arthralgia (*joint pain*) | 520 | Vomiting | 266 | Gout flare |
| 184 | Pruritus (*itching*) | 288 | **Injury, poisoning, and procedural complications** | 243 | **Rash** |
| 180 | Abdominal pain | 255 | Somnolence (*sleepiness*) | 241 | Lower respiratory tract infection (*bronchitis*) |
| 172 | Gastrointestinal symptoms (none) | 162 | Back pain | 62 | Erythema (*redness*) |
| 130 | Dyspepsia (*indigestion*) | 134 | Influenza (*flu syndrome*) | 62 | Infusion reaction |
| 56 | Pain | 94 | Sinusitis | 62 | Mouth ulcers |
| 27 | Injection site pain | 89 | Dry mouth | 41 | Psychiatric disorders (none) |
| 24 | Asthenia (*feeling weak*) | 87 | Pharyngitis (*sore throat*) | 35 | Pyrexia (*fever*) |
| 19 | Depression | 70 | Cough | 32 | Muscle-related signs and symptoms: muscle cramps, muscle twitching, night cramps (none) |
| 10 | Pain in the study joint | 65 | Skin injuries | 29 | Allergic reactions |
| 8 | Itch or dizziness | 41 | Vertigo (*spinning sensation*) | 24 | Osteoarthritis (none) |
| 8 | Myalgia (*muscle pain*) | 39 | Dyspnea (*shortness of breath*) | 16 | Joint effusion (*joint swelling*) |
| 7 | Pain in extremity | 38 | Sun sensitivity | 13 | Eczema |
| 5 | Joint stiffness | 34 | Peripheral oedema (*swelling*) | 8 | Allergic conjunctivitis (none) |
| 4 | Dysphagia (*difficulty in swallowing*) | 33 | Paresthesia (*'pins and needles'*) | 8 | Contusion (*bruise*) |
| 3 | Burning | 24 | Rhinitis (*runny nose*) | 6 | **Colitis (none)** |
| 3 | Malaise (*feeling badly*) | 21 | Chest pain | 6 | **Effusion (none)** |
| 2 | Change of bowel habit | 21 | Flare | 4 | Recurrent falls |
| 2 | Flatulence (*passing gas*) | 16 | Nephrolithiasis (*renal colic*) | 3 | Hospitalized |
| 2 | Increased appetite | 16 | Urticarial (*hives*) | 3 | Induration (none) |
| 2 | Stinging | 15 | Insomnia (*difficulty sleeping*) | 3 | **Optic neuritis (none)** |
| 2 | Tendon pain | 13 | **Pleurisy (none)** | 1 | Abdominal hernia, obstructive (none) |
| 1 | Ear pain | 11 | Gastroenteritis (*stomach flu*) | 1 | Abdominal wall abscess |
| 1 | Feeling of warmth | 10 | Flushing | 1 | Alopecia (*hair loss*) |
| 1 | Hallucination (*sensing things that are not real*) | 6 | Angina pectoris (*angina*) | 1 | **Anal fistula (none)** |
| 1 | Hyperesthesia (*increased sensitivity of any sense*) | 6 | Palpitations | 1 | Blepharitis (*eyelid inflammation*) |
| 1 | Hypoesthesia (*reduced sensitivity of any sense*) | 4 | Dental pain | 1 | Increased body weight |
| 1 | Lack of appetite | 3 | Gastritis | 1 | Infected tophus (none) |
| 1 | Pain in rectum | 2 | Abdominal distension (*bloating*) | 1 | Inguinal hernia (none) |
| 1 | Restless legs syndrome (*restless legs*) | 2 | Anxiety attack | 1 | Mastitis (*inflamed breast*) |
| 1 | Straining | 2 | Muscular weakness (*muscular weakness in the area around the study joint*) | 1 | Menometrorrhagia (*abnormally heavy, prolonged, and irregular uterine bleeding*) |
| | | 1 | **Asthma** | 1 | **Ptosis (***droopy eyelid***)** |
| | | 1 | **Ataxia (***impaired coordination***)** | 1 | **Yellow discoloration of urine** |
| | | 1 | Constipation-related bloating | | |
| | | 1 | **Cystitis (***bladder inflammation***)** | | |
| | | 1 | Irritable bowel syndrome | | |
| | | 1 | Neuralgia (*nerve pain*) | | |
| | | 1 | Skin peeling | | |
| | | 1 | Syncope (*fainting, losing consciousness*) | | |
| | | 1 | Tooth abscess | | |
| | | 1 | Tremor | | |

* Sample is based on harms reported in primary articles of both industry and non-industry trials.
† When difference between medical and lay language terms exits, terms are described in medical term (*lay language term*). Underscore indicates terms added by authors. "None" indicates that no lay language term was identified. Harms in **bold** print indicate disagreements that were resolved by discussion until consensus was reached among authors as whether appropriate for patient self-reports.
AE = adverse event; RA = rheumatoid arthritis.

However, if the patient can also observe the AE, then the patient may still be the best person to report it as a patient reported outcome. Each perspective provides clinically meaningful information although a patient-clinician or clinician-patient harm-paradox might occur if harms appear unequally important or severe when observed from two different points of view (Fig. 1).

Our study showed that most harm terms reported in the selected articles were in medical (e.g., pyrexia) rather than lay language (e.g.,

**Table 3**
Terms for reported severity of harms in the primary articles.

| Unique terms (frequency) | Overall terms (%) | Severity (%) |
|---|---|---|
| Mild (317); Mild adverse effects (8); Mild in nature (1) | Mild (2) | Mild (2) |
| Slight (5) | Slight (<1) | |
| Moderate (1) | Moderate (<1) | Moderate (<1) |
| Severe (8); Severe intensity adverse events (2); Severe AE (1) | Severe (<1) | Severe (<1) |
| Serious AEs (620); Serious adverse events (482); SAE (66); SAEs (47); Serious AE (47); Serious adverse event (9); SAEs not assigned pegloticase causality (7); Serious event (2) | Serious adverse events (6) | Life threatening (7) |
| Serious infections (108); Serious infectious events (43); Serious infections and infestations (24); AEs of interest, serious infection (13); Serious infectious AEs (4) | Serious infections (<1) | |
| Serious TEAEs (9); Treatment-emergent serious adverse events (6) | Treatment-emergent serious adverse events (<1) | |
| Serious noninfectious adverse events (12) | Serious noninfectious adverse events (<1) | |
| Adjudicated CV events (9) | Adjudicated cv events (0<1) | |
| SAEs were assigned causality (5) | SAEs were assigned causality (<1) | |
| Adverse events leading to death (8) | Adverse events leading to death (<1) | Death (<1) |
| Adverse events (5020); AEs (2455); AE (1490); Adverse event (418); Side effects (70); Adverse effects (55); Adverse effect (4) | Adverse events (44) | |
| Treatment-emergent adverse events (1,914); TEAEs (740); TEAE (664); Treatment-emergent gastrointestinal adverse events (293) | Treatment-emergent adverse events (17) | |
| Common adverse events (488); Common AEs (397); Most commonly reported (361) | Common adverse events (6) | |
| AEs of interest (493); Adverse events of interest (32) | Adverse events of interest (2) | |
| Infectious adverse events (173); Infectious AEs (25) | Infectious adverse events (<1) | |
| Noninfectious adverse events (149) | Noninfectious adverse events (<1) | |
| Other events (119); Other adverse events (6) | Other adverse events (<1) | |
| Non-serious adverse events (112) | Non-serious adverse events (<1) | |
| Events that occurred in 10% (99) | Events that occurred in 10% (<1) | |
| Adverse drug reactions (81); Adverse reactions (7); Adverse reaction (5) | Adverse reactions (<1) | |
| Gastrointestinal adverse events (89) | Gastrointestinal adverse events (<1) | |
| Acute infusional events (60) | Acute infusional events (<1) | |
| Adverse events of special interest (45) | Adverse events of special interest (<1) | |
| Reasons for withdrawals (17); Reasons for withdrawal (10); AEs leading to withdrawal (3) | AEs leading to withdrawal (<1) | |
| Injection-site reactions (24) | Injection-site reactions (<1) | |
| Non-APTC events (19) | Non-APTC events (<1) | |
| Mild or moderate (15); Mild to moderate (2) | Mild to moderate (<1) | |
| Laboratory abnormalities (7) | Laboratory abnormalities (<1) | |
| Bowel movement (6) | Bowel movement (<1) | |
| APTC events (4) | APTC events (<1) | |
| Laboratory evaluations (3) | Laboratory evaluations (<1) | |
| Transient non-specific symptoms (2) | Transient non-specific symptoms (<1) | |
| NA (3,658) | NA (17) | |

AE=adverse event; APTC = Antiplatelet Trialists' Collaboration; CV = cardiovascular; NA = not available; SAE=serious adverse event; TEAE = treatment emergent adverse events.

fever). Though most trials used active surveillance to collect AE information, it is unclear whether the collection method was based on e.g., interview or patients' own input. Regardless, "felt" AEs were likely to have been collected from patients in lay language terms and to be spontaneously reported or reported in answer to a question, either general or specific. Then, they were subsequently analyzed and described ("coded") in medical terms e.g., industry typically uses the Medical Dictionary for Regulatory Activities (MedDRA) to harmonize data reporting. As the OMERACT safety working group intends to develop a framework for patient self-reported harms, it is necessary to identify lay language descriptor terms to represent analogous medical terms – initially, to inform a Delphi process including all stakeholders (e.g., patients, clinicians, researchers, ancillary personnel) with the purpose to reach consensus on harm-domains to measure.

Our study revealed a major deficiency in the reporting of harm severity in the published literature, though less so for SAEs. We had planned to categorize the severity level of the reported AEs but, even though severity might be systematically reported to trial databases, in clinical study reports, or to regulators, we found no evidence for systematic reporting of the level of severity in the primary articles. It was also difficult to determine how severity was categorized and whether severity of the AEs was assessed by clinicians or patients, though in industry trials, it is typically assessed by the investigator. From the given (lack of) reporting, it was not possible to formally address harm severity in our study, as a meaningful severity assessment would require more consistent reporting than was found in the included trial literature.

Although it is mandatory to report SAEs in trials relevant to regulatory oversight, seriousness of an AE may not always correlate with severity of the AE though we categorically assessed SAEs as life-threatening for our analysis. Severity is a measure of intensity, whereas seriousness is defined by the criteria presented previously. An AE of severe intensity need not necessarily to be considered serious, e.g., nausea that persists for several hours may be considered severe nausea, but not a serious AE. Alternatively, a stroke that results in hospitalization but minimal to no permanent disability may be considered mild by an investigator but would be a serious AE. From the patient's perspective, one could consider that a patient would also deem the latter scenario severe – thus there is a risk of a patient-clinician harm paradox. The lack of information on harm severity in primary articles makes it difficult to assess the true benefit-harm profile of an intervention, thereby complicating decision making for patients and clinicians alike when considering medical treatment. Because clinicians tend to systematically downgrade the severity of patients' symptoms [18–20] (our study revealed that most AEs reported in trials within rheumatology involved harms with felt components), a fair assessment of severity should include the patients' perspective [44].

To stimulate a balanced and transparent reporting of harms, with emphasis on the rheumatic diseases, we suggest reporting the severity level of harms based on uniform criteria, such as that in the RCTC 2.0 [16]. To achieve complete understanding, harms and their severity should be assessed by both the investigator and the patient, and the reporting of harms should reflect both perspectives. The predominating clinician perspective on harms in the selected articles might

explain why less than one-third (85/280, 30%) of the harm clusters concerned domains in the core area of life impact (a patient domain). Possibly such patient-reported harms may also have been reflected to a certain degree in the score of a health-related quality of life instrument, but these instruments may not cover all harms relevant to patients. A comprehensive collection of patient-reported harms and their impact is essential because patient self-reports reflect impact on daily health [17]. Since other patient-reported harm-instruments, e.g., the PRO-CTCAE, allow severity for some AEs to be based on interference with activities of daily living [22], some might argue that we need a measure that reports life impact of AEs instead. Ultimately, a standardized reporting structure for patient-reported harms within rheumatology RCTs and longitudinal observational studies needs to be developed with patient input.

Our results show a diversity of reporting for harms. Some trials reported harms based on system organ class (e.g., gastrointestinal disorders), whereas other trials reported harms using more specific terms (e.g., preferred terms such as vomiting, dizziness or headache). More non-specific terms (e.g., hospitalized or infections) were also reported. Differences in grouping and reporting of harms between trials might lead to more biased, less reliable and less reproducible results [45]. We did not systematically analyze reporting levels of all reported AEs according to MedDRA, as we were aiming to optimize reporting according to RCTC 2.0. Industry-sponsored trials will report preferred terms due to use of MedDRA [46] which is less likely to occur with non-industry sponsored trials or investigator-initiated studies - this may also explain the observed difference in reporting levels. MedDRA is a licensed tool and thus not often available to the academic investigator. Also, MedDRA is not always easy to use: observers must be trained to code of AE terms accurately. The RCTC 2.0 might be more accessible, and easier for rheumatologists to use to classify harms for standardized reporting.

Our categorization of harm clusters identified some missing categories in RCTC 2.0 [16]. E.g., there were no clear groupings for harms related to specific infections, cancer-related terms, the renal system, and reproductive organs. We incorporated some of these in the General category, thus making it the most used category (20% of the harm clusters), which may or may not be ideal. While RCTC 2.0 is quite usable, these gaps clearly indicate a need for a revision, and periodic updating, of the RCTC 2.0, as has been suggested previously [47]. A revision of the RCTC should also address appropriate use of preferred terms, and match classification to MedDRA for easy cross-referencing.

Our study has some strengths. It included a large amount of data from trials during a 10-year period. It comprised an exhaustive compilation of harms and a collaborative, consensus-driven consolidation of terms into groupings that can be used to further develop standardized harm instruments. It utilized an international team of experts in the field. It brought to the forefront the need for a separate patient-oriented instrument to report and assess harms from the patient's point of view. It also highlighted the way forward for an update of a specific rheumatology-oriented, relatively easy-to-use harms instrument.

A limitation of our study, there is likely underreporting of harms. In published trial literature of health care interventions, harms are underreported in general [2,3], and some of the selected publications from the included trials only described events that were "reported by ≥5% of patients" or "most frequent AEs". Limiting reporting of AEs based on frequency may be important to identify true signals for harm from a single clinical trial based on the rule of three [48]. However, reporting all events that occur can assist in subsequent meta-analyses of data to detect true signals for rare AEs. The use of nonspecific terms to describe AEs might also explain why half of the AEs we found to be appropriate for patient self-reporting were reported fewer than 10 times in the publications of the included trials. Our extraction of data from the included trials most likely worsened underreporting in our study [2]. We excluded secondary publications; we did not examine appendices; we did not seek unpublished data such as clinical study reports, Summary Basis of Approvals, or European Public Assessment Reports; and we extracted the most specific AE terms, not including data such as "total number of AEs" as we could not classify them. Despite these limitations, we established 693 specific unique reported terms for AEs.

We cannot be confident that we identified all harms important to rheumatology patients in this study. We chose specifically to explore rheumatology in this review - such as previously done within cancer [22]. To expand the list of reported harms with felt AEs, additional relevant harms might be identified via review of publications from trials in fields other than rheumatology [7], review of unpublished data [2], and input from patients.

Our study also has other limitations. We selected trials included in systematic reviews conducted by the CMSG over a 10-year period. We cannot exclude the possibility that other rheumatologic trials would describe relevant harm-information not identified from the included trials (e.g., a significant increase in the number of published articles within psoriatic arthritis occurred from 2016 to 2018 and these recent trials were not yet included in selected Cochrane reviews; many large trials of systemic lupus erythematosus were also not available in Cochrane reviews). We also did not request Freedom of Information data, as it is a very lengthy process which might have delayed this project indefinitely. Further, some Cochrane reviews might deal only with efficacy but not safety. As we selected only primary publications from RCTs included in Cochrane reviews, we might have missed secondary papers on safety. Finally, two authors (supported by a third author when in doubt) did the clustering and classification of AEs, and some classification of harms might have been done differently if more authors had been involved.

We believe our results suggest that the development of a framework for patient self-reported harms can potentially provide a more balanced account of treatment experiences as well as a more balanced assessment of treatment strategies when deciding on new treatments. To inform a Delphi process, we need patients and experts globally both to identify lay language terms to cover medical terms for the harm clusters and to identify relevant additional harms. When deciding on which outcomes to measure in the framework, we need a standardized reporting structure for patient-reported harms including severity – a structure that we should develop in collaboration with patients. Further, we also need a revision and expansion of domains included in the RCTC 2.0, and the relative weights to give to the patient perspective and the harms related to pathophysiology etc. will need to be addressed in future research.

In conclusion, we found that 42% of the AEs described in the rheumatology trial literature are appropriate for patient self-reporting, and these represent the majority (58%) of the total number of AEs reported in primary articles of rheumatology clinical trials. For more than 90% of the identified harms, the AE severity was poorly reported. Our results support the development of a standardized reporting framework for patient-reported harms in rheumatology RCTs and longitudinal observational studies to ensure reliable reporting of AEs with severity grading according to both patients and investigators.

## Author contributions

Berthelsen and Christensen had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Berthelsen and Christensen.

Acquisition: Berthelsen with support from Christensen.

Extraction: Berthelsen and Woodworth.

Analysis and interpretation: Berthelsen and Woodworth with support from Furst, Goel and Christensen.

Drafting of the manuscript: Berthelsen and Christensen.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Berthelsen and Christensen.

## Declaration of Competing Interest

Berthelsen: none
Woodworth: none
Ioannidis: none
Tugwell: none relevant
Devoe: none relevant
Williamson: none
Terwee: none
Suarez-Almazor: none
Strand: none relevant
Leong: none
Goel: none relevant
Conaghan: none relevant
Boers: none relevant
Shea: none relevant
Brooks: none relevant
Simon: none relevant
Furst: none relevant
Christensen: none relevant

## Acknowledgments

The Parker Institute is supported by grants from the Oak Foundation. The Oak Foundation is a group of philanthropic organizations and, since its establishment in 1983; it has given grants to not-for-profit organizations around the world.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL); the Oak Foundation had no role in the study design, data collection and analyses, interpretation or reporting of this work, or the decision to submit the work for publication. P.G. Conaghan is supported in part by the UK National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the Department of Health.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2020.09.023.

## References

[1] Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. BMC Med Res Methodol 2007;7:32.
[2] Golder S, Loke YK, Wright K, Norman G. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. PLoS Med 2016;13(9):e1002127.
[3] Tang E, Ravaud P, Riveros C, Perrodeau E, Dechartres A. Comparison of serious adverse events posted at ClinicalTrials.gov and published in corresponding journal articles. BMC Med 2015;13:189.
[4] Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. BMJ (Clinical research ed) 2014;349:g6501.
[5] Keebler D, Teng E, Chia J, Galanter J, Peake J, Tuckwell K. Regional variations in adverse event reporting rates and ACR responses in placebo/standard-of-care arms of rheumatoid arthritis trials. Rheumatology (Oxford) [Internet]. Available from: http://academic.oup.com/rheumatology/advance-article/doi/10.1093/rheumatology/keaa043/5809194
[6] Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. BMJ Open 2013;3(9):e003436.
[7] Ioannidis JP, Evans SJ, Gotzsche 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;141(10):781–8.
[8] 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;22(5):170–7.
[9] Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67(7):745–53.
[10] Boers M, Beaton DE, Shea BJ, Maxwell LJ, Bartlett SJ, Bingham CO, et al. OMERACT filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. J Rheumatol 2019;46(8):1021–7.
[11] Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials 2007;8:38.
[12] Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
[13] COMET Initiative [Internet]. [cited 2020 May 30]. Available from: http://www.comet-initiative.org/
[14] Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol 2018;96:84–92.
[15] Klokker L, Tugwell P, Furst DE, Devoe D, Williamson P, Terwee CB, et al. Developing an OMERACT core outcome set for assessing safety components in rheumatology trials: the OMERACT safety working group. J Rheumatol 2017;44(12):1916.
[16] Woodworth T, Furst DE, Alten R, Bingham 3rd CO, Yocum D, Sloan V, et al. Standardizing assessment and reporting of adverse effects in rheumatology clinical trials II: the rheumatology common toxicity criteria v.2.0. J Rheumatol 2007;34(6):1401–14.
[17] Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009;101(23):1624–32.
[18] Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 2008;14(8):530–9.
[19] Efficace F, Rosti G, Aaronson N, Cottone F, Angelucci E, Molica S, et al. Patient- versus physician-reporting of symptoms and health status in chronic myeloid leukemia. Haematologica 2014;99(4):788–93.
[20] Yen JC, Abrahamowicz M, Dobkin PL, Clarke AE, Battista RN, Fortin PR. Determinants of discordance between patients and physicians in their assessment of lupus disease activity. J Rheumatol 2003;30(9):1967–76.
[21] Klokker L, Berthelsen DB, Woodworth T, Andersen KM, Furst DE, Devoe D, et al. Identifying possible outcome domains from existing outcome measures to inform an OMERACT core domain set for safety in rheumatology trials. J Rheumatol 2019;46(9):1173–8.
[22] Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, et al. Development of the national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 2014;106(9).
[23] OMERACT Handbook [Internet]. OMERACT Handbook. [cited 2020 Apr 6]. Available from: https://omeracthandbook.org/
[24] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ (Clinical research ed) 2009;339:b2700.
[25] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467.
[26] Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med 2007;4(3):e78.
[27] Tarp S, Eric Furst D, Boers M, Luta G, Bliddal H, Tarp U, et al. Risk of serious adverse effects of biological and targeted drugs in patients with rheumatoid arthritis: a systematic review meta-analysis. Rheumatology (Oxford) 2017;56(3):417–25.
[28] Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64(4):465–74.
[29] Smolen JS, Landewe R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis 2017;76(6):960–77.
[30] Singh JA, Saag KG, Bridges SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American college of rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol (Hoboken, NJ) 2016;68(1):1–26.
[31] Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Hauser W, Fluss E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017;76(2):318–28.
[32] Khanna D, Fitzgerald JD, Khanna PP, Bae S, Singh MK, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricemia. Arthritis Care Res (Hoboken) 2012;64(10):1431–46.
[33] Khanna D, Khanna PP, Fitzgerald JD, Singh MK, Bae S, Neogi T, et al. 2012 American College of Rheumatology guidelines for management of gout. Part 2: therapy

and antiinflammatory prophylaxis of acute gouty arthritis. Arthritis Care Res (Hoboken) 2012;64(10):1447–61.
[34] Richette P, Doherty M, Pascual E, Barskova V, Becce F, Castaneda-Sanabria J, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Ann Rheum Dis 2017;76(1):29–42.
[35] Hahn BH, McMahon MA, Wilkinson A, Wallace WD, Daikh DI, Fitzgerald JD, et al. American college of rheumatology guidelines for screening, treatment, and management of lupus nephritis. Arthritis Care Res (Hoboken) 2012;64(6):797–808.
[36] Bertsias GK, Tektonidou M, Amoura Z, Aringer M, Bajema I, Berden JH, et al. Joint European league against rheumatism and European renal association-european dialysis and transplant association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis. Ann Rheum Dis 2012;71(11):1771–82.
[37] Lems WF, Dreinhofer KE, Bischoff-Ferrari H, Blauth M, Czerwinski E, da Silva J, et al. EULAR/EFORT recommendations for management of patients older than 50 years with a fragility fracture and prevention of subsequent fractures. Ann Rheum Dis 2017;76(5):802–10.
[38] van der Heijde D, Ramiro S, Landewe R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017;76(6):978–91.
[39] Kowal-Bielecka O, Fransen J, Avouac J, Becker M, Kulak A, Allanore Y, et al. Update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis 2017;76(8):1327–39.
[40] Post Approval Safety Data Management. Note for guidance on definitions and standards for expedited reporting E2D. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-12.pdf Accessed April 8, 2020.
[41] U.S. Department of Health and Human Services: Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0 (v4.03: June 14, 2010): Available at https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf. Accessed October 28, 2017.
[42] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
[43] Berthelsen DB, Ginnerup-Nielsen E, Juhl C, Lund H, Henriksen M, Hróbjartsson A, et al. Treatment effect sizes vary in randomized trials depending on type of outcome measure. J Clin Epidemiol 2020(123):27–38.
[44] Basch E. The missing patient voice in drug-safety reporting. N Engl J Med 2010;362(10):865–9.
[45] Schroll JB, Maund E, Gøtzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. PLoS ONE 2012;7(7):e41174.
[46] Introductory Guide MedDRA version 22.1.pdf. [Internet]. [cited 2019 Nov 18]. Available from: https://www.meddra.org/sites/default/files/guidance/file/000354_intguide_22.1.pdf
[47] Stach CM, Sloan VS, Woodworth TG, Kilgallen B, Furst DE. Rheumatology common toxicity criteria (RCTC): an update reflecting real-world use. Drug Saf 2019;42(12):1499–506.
[48] Eypasch E, Lefering R, Kum CK, Troidl H. Probability of adverse events that have not yet occurred: a statistical reminder. BMJ 1995;311(7005):619–20.