Towards consensus in defining and handling contextual factors within rheumatology trials

An initial qualitative study from an OMERACT working group

Nielsen, Sabrina Mai; Rasmussen, Marianne Uggen; Boers, Maarten; van der Windt, Danielle A.; de Wit, Maarten; Woodworth, Thasia G.; Flurey, Caroline A.; Beaton, Dorcas; Shea, Beverley; Escorpizo, Reuben; Furst, Daniel E.; Smolen, Josef S.; Toupin-April, Karine; Boonen, Annelies; Voshaar, Marieke; Ellingsen, Torkell; Wells, George A.; Reeves, Barnaby C.; March, Lyn; Tugwell, Peter; Christensen, Robin

Published in:
Annals of the rheumatic diseases

DOI:
10.1136/annrheumdis-2020-217895

Publication date:
2021

Document version:
Accepted manuscript

Citation for pulished version (APA):
Nielsen, S. M., Rasmussen, M. U., Boers, M., van der Windt, D. A., de Wit, M., Woodworth, T. G., Flurey, C. A., Beaton, D., Shea, B., Escorpizo, R., Furst, D. E., Smolen, J. S., Toupin-April, K., Boonen, A., Voshaar, M., Ellingsen, T., Wells, G. A., Reeves, B. C., March, L., ... Christensen, R. (2021). Towards consensus in defining and handling contextual factors within rheumatology trials: An initial qualitative study from an OMERACT working group. *Annals of the rheumatic diseases*, *80*(2), 242-249. https://doi.org/10.1136/annrheumdis-2020-217895

Toward consensus in defining and handling contextual factors within rheumatology trials: an initial qualitative study from an OMERACT Working Group

Authors

Sabrina Mai Nielsen[1,2], Marianne Uggen Rasmussen[1], Maarten Boers[3], Danielle A. van der Windt[4], Maarten de Wit[5], Thasia G. Woodworth[6], Caroline Flurey[7], Dorcas Beaton[8], Beverly Shea[9], Reuben Escorpizo[10], Daniel E. Furst[11,12,13], Josef S. Smolen[14], Karine Toupin April[15], Annelies Boonen[16], Marieke Voshaar[17], Torkell Ellingsen[2], George A. Wells[18], Barnaby C. Reeves[19], Lyn March[20], Peter Tugwell[21], Robin Christensen[1,2]

Affiliations

1 Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark
2 Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark
3 Department of Epidemiology & Biostatistics; and Amsterdam Rheumatology and Immunology Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands.
4 School of Primary, Community, and Social Care; Primary Care Centre Versus Arthritis; and Centre for Prognosis Research, Faculty of Medicine and Health Sciences, Keele University, Staffordshire, UK.
5 OMERACT Patient Research Partner, Amsterdam, the Netherlands.
6 Division of Rheumatology, David Geffen School of Medicine, University of California, Los Angeles, USA.
7 Department of Health and Social Sciences, Faculty of Health and Applied Sciences, University of the West of England, Bristol, UK.
8 Musculoskeletal Health and Outcomes Research, St. Michael's Hospital, and Institute for Work and Health, and Department of Occupational Science and Occupational Therapy, Rehabilitation Sciences Institute and the Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.
9 Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada.
10 Department of Rehabilitation and Movement Science, College of Nursing and Health Sciences, University of Vermont, Burlington, VT, USA.
11 University of California in Los Angeles, USA.
12 University of Washington, Seattle Washington, USA.
13 University of Florence, Florence, Italy.
14 Division of Rheumatology, Department of Medicine 3, Medical University of Vienna, Vienna, Austria.
15 Children's Hospital of Eastern Ontario Research Institute; Department of Pediatrics and School of Rehabilitation Sciences, University of Ottawa, Ottawa, Ottawa, ON, Canada.
16 Care and Public Health Research Institute (CAPHRI), 6229 ER Maastricht University, Maastricht, The Netherlands; Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre+, 6202 AZ Maastricht, The Netherlands.
17 University of Twente, Enschede, Faculty of Behavioural Management and Social sciences, Department Psychology, Health and Technology, University of Twente, the Netherlands.
18 School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.
19 Bristol Trials Centre, Bristol Medical School, University of Bristol, Bristol, UK.
20 Florance and Cope Professorial Department of Rheumatology, Royal North Shore Hospital and Institute of Bone and Joint Research, Faculty of Medicine and Health, University of Sydney, Australia.
21 Department of Medicine, University of Ottawa, Canada.

Corresponding Author:

Dr Robin Christensen, MSc, PhD; Professor of Biostatistics and Clinical Epidemiology
Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital & The Rheumatology Research Unit, Department of Rheumatology, Odense University Hospital and University of Southern Denmark, Denmark.
The Parker Institute, Bispebjerg and Frederiksberg Hospital
Nordre Fasanvej 57, DK-2000 Copenhagen Frederiksberg, Denmark
Phone: +45 3816 4165/E-mail: robin.christensen@regionh.dk

Word count: 3,000

ABSTRACT

Objectives: The Outcome Measures in Rheumatology (OMERACT) Initiative established the Contextual Factors Working Group (CFWG) to guide the understanding, identification, and handling of contextual factors for clinical trials. In clinical research, different uses of the term 'contextual factors' exist. This study explores the perspectives of researchers (incl. clinicians) and patients in defining 'contextual factor' and its related terminology, identifying such factors, and accounting for them in trials across rheumatology.

Methods: We conducted individual semi-structured interviews with researchers (incl. clinicians) who have experience within the field of contextual factors in clinical trials or other potentially relevant areas, and small focus group interviews with patients with rheumatic conditions. We transcribed the interviews and applied qualitative content analysis.

Results: We interviewed 12 researchers and 7 patients. Researcher and patient descriptions of contextual factors were categorised into two broad themes, each comprising two contextual factors types. The 'treatment effect' theme focused on factors explaining variations in treatment effects a) among patients, and b) among studies. The 'outcome measurement' theme focused on factors that explain c) variations in the measurement result itself (apart from actual changes/differences in the outcome); and d) variations in the outcome itself (beside treatment of interest). Methods for identifying and handling contextual factors differed among these themes and types.

Conclusions: Two main themes for contextual factors with four types of contextual factors were identified based on input from researchers and patients. This will guide operationalisation of contextual factors. Further research should refine our findings and establish consensus among relevant stakeholders.

Keywords

Arthritis; Outcome Assessment, Health Care; Patient Reported Outcome Measures; Qualitative research

What is already known about this subject?

- Contextual factors should be considered when developing core outcome sets. Guidance and operationalisation of the current definition are needed to ensure consistency in understanding, approaching, and identifying contextual factors.
- Within OMERACT, the Contextual Factors Working Group (CFWG) was formed to develop guidance on how to address contextual factors in clinical trials.

What does this study add?

- This qualitative study, using semi-structured interviews with researchers and small focus group interviews with patients, suggests that contextual factors can be grouped into two broad themes: 'treatment effect' and 'outcome measurement.' The 'treatment effect' theme comprises two types of contextual factors: a) 'effect modifying' (pertaining to effect variations among patients); and b) 'meta confounding' (pertaining to effect variations among studies). The 'outcome measurement' theme also comprises two types of contextual factors: c) 'measurement affecting' (pertaining to variations in measurement results); and d) 'outcome explaining' (pertaining to variations in the outcome itself).

How might this impact on clinical practice or future developments?

- This study provides a foundation for developing a consensus-based operational definition of contextual factors, which may specify relevant contextual factor types and include guidance on how to identify such factors and take them into account to ensure proper interpretation of clinical trial findings.

INTRODUCTION

A 'core outcome measurement set' is a minimum consensus-based set of outcome domains and instruments that should be measured and reported in clinical trials for a specific health condition and/or intervention.[1] Since 1992, the Outcome Measures in Rheumatology (OMERACT) initiative has successfully developed core sets for many rheumatologic conditions[2] and kept patients actively involved since 2002.[3]

In 2012, the concept of contextual factors was introduced in the OMERACT process. In clinical research, different uses of the term 'contextual factors' exist, describing different concepts.[4-7] Within OMERACT, a contextual factor is defined as a "*variable that is not an outcome of the study, but needs to be recognized (and measured) to understand the study results. This includes potential confounders and effect modifiers*".[3] Core set developers need to consider if there are contextual factors that should be measured in all trials. However, the research presented at the OMERACT meeting in 2014 revealed much heterogeneity in understanding, approaching, and identifying contextual factors.[8] To address this, the Contextual Factors Working Group (CFWG) was formed to develop guidance on how to address contextual factors in clinical trials.[8]

In 2018, the CFWG presented a research plan: initially it would collect 'case scenarios' involving 'contextual factors' from OMERACT working groups; then develop an operational definition and guidance on how to address contextual factors in rheumatology trials when developing core outcome measurement sets; and ultimately develop a generic set of important contextual factors (i.e., important across all rheumatic diseases) that should always be considered in rheumatology trials based on empirical evidence and consensus.[9] To operationalise the definition of contextual factors, an expert-driven approach, including qualitative data collection with a subsequent consensus process among important stakeholders, was proposed.

The objective of the current study is to explore the perspectives of researchers (incl. clinicians) and patients in defining 'contextual factor' and its related terminology, identifying such factors, and accounting for them in trials across rheumatology (i.e., across different OMERACT working groups).

METHODS

Design

In this qualitative study, we conducted semi-structured interviews with researchers and small focus group interviews with 2-3 patients, and applied qualitative content analysis.[10 11] As a research method, qualitative content analysis aims "to provide knowledge and understanding of the phenomenon under study."[12] We published a protocol online prior to conducting any interviews (**online supplementary file 1** and www.parkerinst.dk). The Danish Data Protection Agency approved the study (ID 06081, BFH-2017-127), and the study was carried out in accordance with the Declaration of Helsinki.

Participants and setting

Individually interviewed participants were required to be researchers (e.g., statisticians, methodologists, trialists, including clinicians) who have experience with the field of contextual factors in clinical trials or other potentially relevant areas, such as predictive/prognostic factors, effect modification, subgroup effects, stratified analyses, or equity efforts (i.e., initiatives centred on factors of social inequity). We used purposive sampling to maximise variation of disciplines and sex and geographical representation, and expanded our sample by snowball sampling (i.e., asking each participant to suggest additional researchers).[13] We initially identified participants among our co-authors, the OMERACT Executive board, and authors of relevant empirical studies and known guidance documents. We selected patients from the patient research partners (PRPs) of the CFWG. The main interviewer (SMN) determined the sample size by theoretical saturation, defined as the size where subsequent interviews contribute no more new data.[14]

We approached potential participants by e-mail invitation. Upon their acceptance of participation, we provided an overview of the interview content, the research protocol, and case scenarios involving contextual factors previously collected from OMERACT working groups.[9]

Data collection

From April through July of 2018, one investigator (SMN) interviewed the researchers individually (average 47 min) and interviewed the focus groups (2-3 patients) supported by 1-2 co-investigators (TW and CF; average 1 hour and 21 min). We conducted the interviews online or face-to-face. We conducted all interviews in English, using a predefined semi-structured interview guide (**online supplementary file 1**) and probing questions, allowing relevant statements to be explored in more depth. Patients were interviewed using an adapted interview guide (i.e., reformulated using lay terms in collaboration with a PRP, MdW). We audio recorded the interviews, transcribed verbatim, returned the transcripts to the participants for comments and/or corrections, and collected demographic data.

Data analysis

One investigator (SMN, supported by MUR) conducted qualitative content analysis[10 11] (investigator characteristics in **online supplementary file 1**) using NVIVO (version 12 Pro). We generated the coding frame by initially creating main categories in a concept-driven way based on the structure of the interview guide, and adding subcategories in an inductive, data-driven way with open coding based on 'successive summarising'. This method involved paraphrasing relevant passages while removing unnecessary parts. We

revised the coding frame, added explanations and supporting quotes, and subsequently conducted further data exploration to search for patterns and co-occurrences of selected categories.[10]

We ensured rigor and credibility by discussing key findings at CFWG meetings, and sharing a draft of the findings with some of the interviewees to ensure viewpoints were appropriately interpreted and the account made sense to other researchers and patients (i.e., 'member checking').[15] We ensured comprehensive reporting by following the Consolidated Criteria for Reporting Qualitative Studies (COREQ)[16] and the Standards for Reporting Qualitative Research (SRQR).[17]

Patient involvement

During the whole process we involved two PRPs who are familiar with the research topic. These and five additional PRPs with experience of living with rheumatic conditions were involved as participants in the interviews.

RESULTS

Participant characteristics

A total of 16 researchers were invited; 4 (25%) did not respond and 12 (75%) agreed to participate. All seven (100%) invited patients agreed to participate. The researchers represented several stakeholder groups, and half were involved in patient care (**Table 1**). The patients represented three rheumatic conditions.

Table 1: Characteristics of the interviewed researchers and patients

	Researchers (n=12)	**Patients** (n=7)
Females	6	5
Age, years, mean (SD)	58 (8)*	55 (8)
Continent		
North America	3	2
Europe	8	3
Australia	1	2
Involved in OMERACT		
Currently involved	11	7
Never involved	1	0
Organisation		
Academic	11	-
Healthcare	1	-
Main role providing CF experience		
Rheumatologist	5	-
Statistician	2	-
Epidemiologist	2	-
Methodologist	1	-
Occupational therapist	1	-
ICF expert	1	-
Involved in patient care		
Currently	6	-
Previously	3	-
Never	3	-
Rheumatic condition		
Rheumatoid arthritis	-	4
Psoriatic arthritis	-	2
Bechet's Syndrome	-	1
Research experience beside PRP role		
Yes	-	6
No	-	1

Values are the number of patients, unless indicated otherwise.
CF, contextual factor; ICF, The International Classification of Functioning, Disability and Health; PRP, patient research partners; SD, standard deviation.
*Data on age were missing for three researchers.

Reflections on the current OMERACT definition

Only a minority of the participants found the current OMERACT definition of a contextual factor as a "*variable that is not an outcome of the study, but needs to be recognized (and measured) to understand the study results. This includes potential confounders and effect modifiers*"[3] to be clear and understandable (A.1 in **online supplementary table 1**). Some thought the term 'contextual factor' was too broad and confusing. Some researchers discussed whether 'confounder' should be part of the definition because it may be less relevant in randomised trials. In contrast to the definition's first part, many considered the outcome itself at baseline to be a possible contextual factor (e.g., the level of pain at baseline may be important when interpreting the changes in pain at follow-up in a trial). Overall, the patients had difficulty understanding the definition, mainly due to the terms used:

"I still find it quite difficult to understand. I think I have an idea of what a contextual factor is.

I'm not sure that I know exactly the difference between a confounder or an effect modifier. Do we really need these terms? (…) I think, it's a definition for researchers, but it's not a definition for patient research partners." (Patient 3)

Participants' own description of contextual factors

The participants' own descriptions of contextual factors revealed two broad themes, each comprising two types of contextual factors. The first theme, 'treatment effect', focused on factors that explain variations in treatment effects a) among patients (or groups of patients), and b) among studies. The second theme, 'outcome measurement', focused on factors that explain c) variations in the measurement result itself (apart from actual changes/differences in the outcome), and d) variations in the outcome itself (apart from the treatment of interest). These four types may be termed 'effect modifying', 'meta-confounding', 'measurement affecting', and 'outcome explaining' contextual factors, respectively (**Figure 1**, **Table 2** and part A.2 in **online supplementary table 1**). Specific examples of factors may fit within more than one contextual factor type.

[Figure 1 here]

Few researchers recognised that both themes exist; most emphasised only one of them. The patients mostly focused on what (besides treatment) affects their condition, their lives with the condition, and how symptoms are perceived—which in turn also affects their lives (these considerations relate to the outcome measurement theme). Several patients emphasised that contextual factors are inherently patient centric:

"In terms of what you were doing here is patient centric, in terms of the contextual factors, because we're the only ones who really know what they are." (Patient 7)

225 **Table 2:** The two themes for contextual factors, each describing two types of contextual factors

Theme	Treatment effect theme	Outcome measurement theme
Description	Factors that influence (or are associated with or predict) the treatment effects.	Factors that influence the outcome measurement.
Rationale	To understand the study results in terms of for whom and/or in which settings a treatment shows an effect, and to assess the external validity/generalisability of a study, which relates to stratification/precision medicine.	To understand the study results in terms of what influences the outcome measurement (beside the treatment of interest), and to understand 'what is behind the numbers' of a measurement.
Types	**A) Effect modifying** factors are effect modifiers and explain the variability in treatment effect among patients according to characteristics, and may guide treatment decisions (stratified medicine). **B) Meta-confounding** factors relate to the interpretation of the results of a trial when comparing with other trials (e.g., in meta-analysis), and explain inherent variations in treatment effects among trials according to trial-level characteristics.	**C) Measurement affecting** factors explain the variability in the measurement itself, and relate to the difficulty or inability to measure an outcome (validity/reliability), and may impact our ability to see a treatment response. **D) Outcome explaining** factors (besides treatment of interest) affect the outcome; they may be prognostic factors† and may explain different impact of symptoms or perceptions of a response, and may confound group trial results. Such factors may follow the ICF framework.[4]
Lay terms to a patient*	Factors that may predict how well you will benefit from a treatment. Factors that we need to know in a study to know whether the findings can be applied to a particular situation.	Clinicians and researchers need to know what affects your assessment (e.g., of pain), so they can understand the numbers. When they ask you about your scores, you may say "Well, it depends (…)". Factors that influence your condition and your life with the condition, besides the treatment you are getting.
Examples of evidence	***A) Disease duration:*** Rheumatoid arthritis (RA) patients, with a history of responding inadequately to bDMARDs, tend to have a higher chance of responding to Baricitinib compared to placebo if they had RA for ≥10 years.[18] (i.e., disease duration modifies the effect of Baricitinib). ***B) Study year (capturing disease severity):*** Over time, disease characteristics of RA patients in trials on TNFα inhibitors have generally become less severe. This may be due to a change in standard of care, trial site location, trends in inclusion criteria, etc.[19] (i.e., study year may capture inherent CFs that are important when interpreting study results, such as in a meta-analysis).	***C) A. Literacy when assessing reliability of joint pain measurement instruments:*** In RA patients, VAS pain assessments are less reliable in illiterate patients compared to literate patients.[20] ***D) Weather:*** In knee OA patients, reporting more severe pain (WOMAC pain) was associated with lower ambient temperature and higher change in barometric pressure.[21] ***D) CFs for worker productivity:*** OMERACT members (incl. PRPs, HCPs, etc.) were asked to propose and rank CFs affecting WP in arthritis patients. Key CFs identified were type of job, personal factors, disease status, financial need, societal incentive, and age, and should be considered when interpreting WP measurements.
Suggested criteria for important CFs*	Strong suspicion until evidence exists, evidence for statistical interaction and important variability in effect across subgroups. ***For generic factors***, criteria for strong and consistent evidence across rheumatology.	Factors that patients frequently consider important for interpreting outcome measurements, or for their condition/life with their condition. ***For generic factors,*** need to be relevant across countries and conditions.
Suggested methods for identifying important CFs*	Investigate CFs in existing data sets, request trialists to measure CFs and provide stratified analyses as supplement, conduct systematic review. Use existing guidelines on investigating subgroup effects. ***For generic factors***, investigation of effect modifiers in IPD meta-analysis, literature review and/or seeking expert/stakeholder opinion, use CFs identified in OMERACT disease working groups.	Ask patients and/or clinicians directly or do a systematic review.

226 CFs, contextual factors; bDMARDs, biologic disease-modifying anti-rheumatic drugs; HCPs, health care professionals; ICF, The
227 International Classification of Functioning, Disability and Health; IPD, individual patient data; OA, osteoarthritis; PRPs, patient
228 research partners; RA, rheumatoid arthritis; VAS, visual analogue scale; WOMAC pain, Western Ontario and McMaster Universities
229 Arthritis Index pain subscale; WP, worker productivity.
230 *Descriptions mainly relate to only two types of contextual factors (i.e., the 'effect modifying' – and 'outcome explaining'
231 contextual factors, respectively), due to lack of data on the two remaining types ('meta-confounding' and 'measurement affecting').
232 † Prognostic factors are factors predicting the outcome or course of a patient's condition, regardless of treatment.[22]
233

Some researchers initially considered contextual factors to be measured at baseline (A.4 in **online supplementary table 1**), and hence, fixed, but later acknowledged that some may be time-varying:

> *"I have to admit that I usually think of contextual factors as being fixed, but I can't see why they can't be time-varying"* (Researcher 7)

However, allowing contextual factors to vary over time adds complexity, and several researchers recommended focusing only on contextual factors measured at baseline. One researcher termed time-varying contextual factors 'mediators', which may explain why a treatment works in terms of working mechanism, and the researcher mentioned adherence to a regimen and patient-therapist relationships as examples. However, the researcher pointed out that 'mediators' are not mentioned in the current definition.

Explaining contextual factors in lay terms to a patient

When researchers were asked how to explain contextual factors in lay terms to patients (A.3 in **online supplementary table 1**, and **Table 2**), within the treatment effect theme, contextual factors were often explained as factors that may determine which patients experience an effect. Within the outcome measurement theme, one researcher suggested explaining contextual factors within the International Classification of Functioning, Disability and Health (ICF) framework and providing examples for specific outcomes. The patients themselves repeatedly expressed that the terms 'confounder' and 'effect modifier' were problematic and that examples are needed:

> *"(…) it would be good if you could find an example within rheumatology (…) And I think that would be very helpful if you also could find an example of a contextual factor that has been studied, and for which we have some data to show how it influences."* (Patient 3)

Terminology

For the treatment effect theme, researchers often considered contextual factors to be related with the terms 'effect modifiers' (i.e., factors modifying the effects of a treatment), and 'predictive factors' (i.e., factors predicting the effects of a treatment) and used terms such as 'baseline covariate', 'Table 1 factors', 'subgroup effects' and 'baseline covariance'. For the outcome measurement theme, a researcher explained that the contextual factors are not required to predict treatment response (A.5 in **online supplementary table 1**).

> *"(…) all of these are contextual factors, irrespective of their role as a predictive factor or not."* (Researcher 3)

Examples of contextual factors

Examples of contextual factors mentioned by at least five participants were age, sex, place of residence, socioeconomic status, disease duration, healthcare system, adherence and support, (**online supplementary figure 1**, and part A.6 in **online supplementary table 1**). These were mostly 'effect modifying' contextual factors as most of the interviews concerned those. Some factors were sometimes considered specific to disease, outcome, or treatment. Within the outcome measurement theme, the contextual factors mentioned related often to specific outcomes, such as joint pain (**Figure 1**). Consistent with the ICF, some researchers only considered two categories of factors (e.g., personal and environmental factors). Examples of factor categories that some researchers intuitively did not consider contextual factors included disease-, intervention-, and measurement-related factors (e.g., how the questionnaire was administered), baseline status of outcome of interest, and factors relating to study design.

Identifying important contextual factors

For considering contextual factors to be important (B.1 in **online supplementary table 1**) within the treatment effect theme, a researcher suggested that a strong suspicion—based on expert consensus—be required until evidence exists of a statistically significant interaction between the contextual factor and intervention, with important effect size (i.e., important variability in effect size among subgroups or settings). For generic (across diseases) contextual factors, researchers suggested that sufficient (meaning strong and convincing) and consistent evidence across rheumatologic conditions should be present. It was further emphasised that the criteria need to be strict and that there be consensus about them:

> *"There should be some very, very strict criteria, before we as OMERACT, can say, this is core and we mandate everybody to measure this always. (…) and then you'd have to have some sort of consensus exercise to say, well, we're only going to name it 'core' if we can show in at least three rheumatology conditions that it makes a difference, something like that."*
> (Researcher 1)

Researchers suggested several different methods for identifying important contextual factors (**Table 2**).

Contextual factors in future research

Within the treatment effect theme, researchers provided many different suggestions on how future trials can take contextual factors into account in their design, analysis, and reporting (**Table 3** and part B.2 in **online supplementary table 1**). Participants emphasised that a list of important contextual factors should be available when designing trials. The suggested analysis methods and reporting depended to some extent on the participant's discipline and on the terms (e.g., confounders, prognostic factors, effect modifiers)

with which they associated contextual factors. Several participants suggested that the analyses had to be pre-specified and that stratified results according to contextual factors should be presented. Within the outcome measurement theme, fewer and less statistical approaches were suggested.

Table 3: Suggestions on how to take contextual factors into account in future research

Theme	Treatment effect theme*	Outcome measurement theme*
Designing	● Measure CFs according to evidence-based and/or consensus-based CF list available for investigators and regulators ● Design trials so confounding is avoided (e.g., by excluding specific types of patients) ● Ensure balance of CFs among the treatment groups ● Ensure sufficient variation within CFs in the trial population ● Require that some CFs be investigated in meta-research	● Measure CFs relevant for outcome of interest ● Allow flexibility to deviate from CF list ● Avoid influence from CFs by measuring outcomes as consistently as possible (e.g., at same time each day)
Analysing	● Adjust for CFs (for confounders) ● Stratify analyses for CFs (effect modifiers) ● Conduct proper analysis for effect modifiers (i.e., test for interaction and present stratified results) ● Pre-specify analyses in an analysis plan and specify whether they are exploratory or confirmatory (most trials are not powered to detect subgroup effects) ● Aggregate data from several trials and stratify	● Conceptually adjust the outcome measurements for relevant/influential CFs
Reporting	● Require stringent reporting of CF data (measure of variability, amount of missing data, how it was measured, and how well it was measured) ● Require CFs be in reporting guideline for rheumatology trials ● Report CFs (prognostic factors) as part of extensive baseline table ● Stratify results by CFs (predictive factors) (e.g., as appendix) ● Account for CFs when interpreting results (with respect to generalisability, differing results according to levels of CFs, or explaining skewed results from imbalances among groups) ● Ask stakeholders how they prefer CFs to be reported	● Account for CFs when interpreting results (in terms of what the numbers mean) and co-report relevant/influential CFs

CFs, contextual factors.
*Suggestions within these themes mainly relate to only two types of contextual factors (i.e., the 'effect modifying' – and 'outcome explaining' contextual factors, respectively), due to lack of data on the two remaining types.

Further comments and suggestions

The participants acknowledged the importance of the effort of the OMERACT CFWG (B.3 in **online supplementary table 1**) but raised concerns on several potential issues: whether a generic set of contextual factors can be developed; how to deal with factors that are not feasible to measure (e.g., due to cost or causing delays in trials); and how to determine something is so important that everybody needs to measure it, if robust evidence is lacking to make that call.

One researcher advocated that OMERACT should focus on factors within the outcome measurement theme and argued that this should be the niche of OMERACT, as others are already looking into factors within the treatment effect theme:

"I think those are maybe of primary importance to OMERACT, the ones that are influencing the very meaning of the results of what those numbers mean, how we should be interpreting these numbers. (…) but does OMERACT need to have a special little niche where it talks about the outcomes and what you need to do to measure outcomes well, which nobody else is doing? Nobody else is picking up the contextual factors that you need to be able to perfect your outcome measurements." (Researcher 11)

Other suggestions included: to ensure the operational definition can be understood both by people who are familiar with statistics and those who are not; to pass measures of contextual factors through the OMERACT instrument filter; and to use the term 'important contextual factors' rather than 'core contextual factors' until sufficient evidence is present. Two researchers even suggested not using the term 'contextual factors' altogether. Further, comments included that differences between sexes are neglected in trials, and study year and type of placebo are neglected in systematic reviews. Also, a researcher commented that contextual factors may be PICOT- (acronym for population, intervention, comparison, outcome, and time) specific.

DISCUSSION

This study found that contextual factors overall may be described within two broad themes: those relating to the 'treatment effect' and those relating to the 'outcome measurement.' Each theme, in turn, comprised two types of contextual factors, thus making four types of contextual factors. The descriptions of the contextual factor types should not be considered final, but rather first step in approaching a complex concept, and is intended to stir up debates regarding improving interpretation of trial results, and eventually lead to an consensus-based operational definition.

Most participants in this study recognised only one type of contextual factors, indicating that efforts are needed to facilitate understanding of all four types when describing contextual factors. This finding may explain the heterogeneity in understanding and identifying contextual factors within (and outside) OMERACT. This study provides a foundation for designing a Delphi study to reach consensus on an operational definition of contextual factors. As OMERACT mainly focuses on clinical trials, 'meta-confounding' contextual factors may be considered outside the scope of such effort.

Operationalising contextual factors will include refining the descriptions of each contextual factors type and developing guidance for each of them (i.e., how to identify and account for them in trials). Guidance for 'effect modifying' contextual factors may already exist, related to investigating,[23-25] reporting,[24 26] and evaluating the credibility[27] of subgroup effects in trials, and for systematic reviews.[28-30] For sex/gender specifically, the Sex and Gender Equity in Research (SAGER) guideline[31] recommends that

results are presented disaggregated by sex. Guidance from regulators, such as European Medicines Agency (EMA)[32] and U.S. Food and Drug Administration (FDA) also exists.[33-35] The 'outcome explaining' contextual factors may relate to so-called 'intercurrent events'.[36-38] Many potentially relevant efforts may provide inspiration when developing guidance (**Box 1**).

Box 1: Efforts potentially related to contextual factors

- Recommendations on subgroup effects, *including investigating,[23-25] reporting[24 26] and evaluating the credibility[27] of subgroup effects in trials, but also in systematic reviews[28-30] from various research groups as well as* regulators*, such as European Medicines Agency (EMA)[32] and U.S. Food and Drug Administration (FDA)[33-35]*
- Efforts aimed at equity,[39-43] *centered on factors of social inequity, represented by the acronym PROGRESS-Plus (i.e., place of residence, race/ethnicity/culture/language, occupation, sex/gender, religion, education, socioeconomic status, social capital, and other characteristics, such as age, disability, sexual orientation, time-dependent situations, and relationships)*
- The PROGnosis RESearch Strategy (PROGRESS) framework,[22 25] *including guidelines for prognostic factors and factors predictive of treatment effect*
- The Context and Implementation of Complex Interventions (CICI) framework,[44] *including context separated into seven domains (i.e., geographical, epidemiological, socio-cultural, socio-economic, ethical, legal, political)*
- The International Consortium for Health Outcomes Measurement (ICHOM),[45 46] *including so-called 'case mix variables' (i.e., risk-adjustment variables) for the outcome set developed*
- The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN),[47 48] *including guidelines for assessing 'cross-cultural validity' and 'inconstancy' in systematic reviews of patient-reported outcome measures*
- The ICF framework,[4] *including so-called personal and environmental contextual factors*
- The Grading of Recommendations, Assessment, Development and Evaluations (GRADE),[49] *including recommendations for assessing inconsistency and applicability in systematic reviews*
- The Cochrane Collaboration's revised tool for assessing risk of bias in randomised trials (RoB 2.0),[50] *including risk of bias in measurement of the outcome*
- Efforts on estimands and sensitivity analysis in clinical trials by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH),[36] FDA[37] and EMA,[38] *including descriptions of 'intercurrent events'*
- Efforts investigating placebo effects,[5 6] *using the terms 'contextual effect' or 'context effect' (and 'context factors' and 'contextual factors')*
- Efforts investigating the active use of patient's context in patient care,[7] *using the term 'contextualization' of patient care referring to the process of identifying the context (circumstances) of individual patients and, if necessary, adapting the plan of care*

One limitation of the study is the absence of investigator triangulation (i.e., corroboration of key findings through analysis by several investigators and subsequent consensus). Member checking (i.e., sharing a draft of the findings and inquiring whether viewpoints were faithfully interpreted)[15] was conducted for only some of the participants. As we used purposive sampling, the participants may not be representative of all relevant experts. The term 'contextual factor' has been used to describe different concepts in clinical research.[4-7] These other concepts could potentially have appeared more strongly in the interviews if a different sample of experts had been included. Most participants focused on 'effect

modifying' or 'outcome explaining' contextual factors; little data was available for the two other types, making the findings less conclusive and leaving more to be clarified during a subsequent consensus process. Furthermore, this study did not address how to measure contextual factors.

To conclude, this qualitative study found that contextual factors overall may be described in two broad themes, 'treatment effect' and 'outcome measurement', with each theme comprising two types of contextual factors. The methods for identifying and handling contextual factors differ between the types, so an operational definition of contextual factors may need to specify these types, and include guidance on how to identify such factors and take them into account. Further research should refine our findings and establish consensus.

Acknowledgements

We gratefully acknowledge all the researchers and patients interviewed for this study. We acknowledge the researchers Francis Guillemin (Université de Lorraine, APEMAC, Nancy, France), Irene van der Horst-Bruinsma (Amsterdam UMC, location Vumc), Melissa Selb (ICF Research Branch, a cooperation partner within the WHO Collaboration Centre in Germany, at DIMDI), Philip Mease (Division of Rheumatology Research, Swedish Medical Center and University of Washington), and Sarah Mackie (Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust). We acknowledge the patients Esen Cam (Phd Candidate), Pamela Richards (University of Bristol), Niti Goel (Duke University School of Medicine), Amye Leong (President & CEO, Healthy Motivation), and Michael Gill (Dragon Claw Charity Ltd). We thank Mr. Tudor Stefan Dragan for his assistance during the transcription process.

Contributors

RC and SMN conceived the study and developed the protocol. SMN collected the data, supported by TW and CF. SMN did the analysis and interpreted the analysis, supported by MUR. SMN drafted the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final version of the manuscript. RC, TE and SMN obtained funding. SMN and RC are the guarantors.

Funding

The Parker Institute is grateful for the financial support received from public and private foundations, companies, and private individuals over the years. The Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL); 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. SMN has received PhD Scholarships from the Faculty of Health

Sciences, University of Southern Denmark, and Odense University Hospital, and an introductory scholarship from the BFH Research Foundation. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all data of the study and had final responsibility for the decision to submit for publication.

Competing interests

AB received a research grant from Abbvie, and an honorarium for lecture or advisory boards from UCB, Galapagos and Lilly; all financial support was paid to her department. **DB** reports that she is on the executive of OMERACT (Outcome Measurement in Rheumatological Research) and chair its methodology group; this did not influence this current written work. **DEF** reports grant/research support from Corbus, Galapagos Golead, GSK, NIH, Pfizer, Talaris, CSL Behring, as well as consultant for Abbvie, Amgen, Corbus, Galapagos, Gilead, Novartis, Pfizer, Roche/Genentech, Talaris, CSL Behring, Boehringer Ingelheim. **JS** reports grants from Abbvie, AstraZeneca, Janssen, Lilly, Merck Sharpe & Dohme, Pfizer, and Roche, from AbbVie, Amgen, AstraZeneca, Astro, Bristol-Myers Squibb, Celgene, Celltrion, Chugai, Gilead, ILTOO Pharma, Janssen, Lilly, Merck Sharp & Dohme, Novartis-Sandoz, Pfizer, Roche, Samsung, Sanofi, and UCB., outside the submitted work. **RC** reports non-financial support from Board membership, grants from Consultancy (AbbVie, Amgen, Axellus A/S, Biogen, Bristol-Myers Squibb, Cambridge Weight Plan, Celgene, Eli Lilly, Hospira, MSD, Norpharma, Novartis, Orkla Health, Pfizer, Roche, Sobi, Takeda), personal fees from Employment (Research Unit for Musculoskeletal Function and Physiotherapy, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark), non-financial support from Expert testimony, grants from Grants/grants pending (Axellus A/S, AbbVie, Cambridge Weight Plan, Janssen, MSD, Mundipharma, Novartis, and Roche), grants from Payment for lectures including service on speakers bureaus (Abbott, Amgen, Axellus, Bayer HealthCare Pharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Cambridge Weight Plan, Ipsen, Janssen, Laboratoires Expanscience, MSD, Mundipharma, Norpharma, Novartis, Pfizer, Roche, Rottapharm-Madaus, Sobi, and Wyeth), grants from Payment for manuscript preparation (Axellus, Bristol-Myers Squibb, and Cambridge Weight Plan, Aleris-Hamlet (via Norpharma)), non-financial support from Patents (planned, pending or issued), non-financial support from Royalties, grants from Payment for development of educational presentations (Bristol-Myers Squibb, MSD, Pfizer), non-financial support from Stock/stock options, grants from Travel/accommodations/meeting expenses unrelated to activities listed (Abbott, AbbVie, Axellus, Biogen, Bristol-Myers Squibb, Cambridge Weight Plan, Celgene, Laboratoires Expanscience, Norpharma, Novartis, Pfizer, Roche, Rottapharm-Madaus, and Wyeth), non-financial support from Other (err on the side of full disclosure), outside the submitted work; and he is involved in many health-care initiatives and research that could benefit from wide uptake of this publication (including Cochrane, OMERACT, IDEOM, RADS, and the GRADE Working Group);

Musculoskeletal Statistics Unit, The Parker Institute is grateful for the financial support received from public and private foundations, companies and private individuals over the years. **PT** reports other from Amgen, Astra Zeneca, Bristol Myers Squibb, Celgene, Eli Lilly, Genentech/ Roche, Genzyme / Sanofi, Horizon Pharma Inc., Merck, Novartis , Pfizer, PPD, Quintiles, Regeneron, Savient, Takeda Pharmaceutical, UCB Group, Vertex, Forest, Bioiberica, personal fees from UCB Biopharma GmbH & SPRL, Parexel International, Prahealth Sciences, personal fees from CHEOR Solutions (Canada) Ltd., Innovative Science Solutions LLC, Reformulary Group, other from Elsevier, Little Brown, Wolters Kluwer Ltd. and John Wiley & Sons Ltd, other from Abbott, Roche, Schering Plough/Merck, UCB, BMS, outside the submitted work; and Peter Tugwell is the recipient of Canada Research Chair in Health Equity (Tier 1 - 2016 to 2024) from the Canadian Institutes of Health Research.

Figure legends

Figure 1: Illustration of the two themes for contextual factors, each describing two types of contextual factors. Specific examples of factors may fit within more than one contextual factor type. Meta-confounding contextual factors (marked with dotted lines) are factors that can be investigated only across trials (on a trial-level) and are therefore not relevant within a single trial. The meta-confounding factor study year may capture different important aspects to consider between studies, such as different therapeutic trends of the time and, hence, typical treatment history of patients, as well as trends in exclusion criteria (e.g., TB screening).

Online supplementary material

Online supplementary file 1: Protocol

Online supplementary table 1: Coding frame with explanations and supporting quotes 2020.03.31

Online supplementary figure 1: Contextual factor domains mentioned in the interviews

REFERENCES

1. Boers M, Beaton DE, Shea BJ, et al. OMERACT Filter 2.1: Elaboration of the Conceptual Framework for Outcome Measurement in Health Intervention Studies. *The Journal of rheumatology* 2019;46(8):1021-27. doi: 10.3899/jrheum.181096 [published Online First: 2019/02/17]
2. Maxwell LJ, Beaton DE, Shea BJ, et al. Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology. *The Journal of rheumatology* 2019;46(8):1014-20. doi: 10.3899/jrheum.181097 [published Online First: 2019/02/17]
3. Boers M, Kirwan JR, Wells G, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *Journal of clinical epidemiology* 2014;67(7):745-53. doi: 10.1016/j.jclinepi.2013.11.013 [published Online First: 2014/03/04]

4. WHO. International classification of functioning, disability and health: ICF. Geneva: WHO, 2001.
5. Di Blasi Z, Harkness E, Ernst E, et al. Influence of context effects on health outcomes: a systematic review. *Lancet (London, England)* 2001;357(9258):757-62. doi: 10.1016/s0140-6736(00)04169-6 [published Online First: 2001/03/20]
6. Zou K, Wong J, Abdullah N, et al. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: meta-analysis of randomised controlled trials. *Annals of the rheumatic diseases* 2016;75(11):1964-70. doi: 10.1136/annrheumdis-2015-208387 [published Online First: 2016/02/18]
7. Schwartz A, Weiner SJ, Harris IB, et al. An educational intervention for contextualizing patient care and medical students' abilities to probe for contextual issues in simulated patients. *Jama* 2010;304(11):1191-7. doi: 10.1001/jama.2010.1297 [published Online First: 2010/09/16]
8. Finger ME, Boonen A, Woodworth TG, et al. An OMERACT Initiative Toward Consensus to Identify and Characterize Candidate Contextual Factors: Report from the Contextual Factors Working Group. *The Journal of rheumatology* 2017;44(11):1734-39. doi: 10.3899/jrheum.161200 [published Online First: 2017/05/04]
9. Nielsen SM, Tugwell P, de Wit MPT, et al. Identifying Provisional Generic Contextual Factor Domains for Clinical Trials in Rheumatology: Results from an OMERACT Initiative. *The Journal of rheumatology* 2019;46(9):1159-63. doi: 10.3899/jrheum.181081 [published Online First: 2019/01/17]
10. Flick U. Chapter 12 - Qualitative Content Analysis. The SAGE Handbook of Qualitative Data Analysis. London: SAGE Publications, Inc. 2014.
11. Elo S, Kyngas H. The qualitative content analysis process. *Journal of advanced nursing* 2008;62(1):107-15. doi: 10.1111/j.1365-2648.2007.04569.x [published Online First: 2008/03/21]
12. Downe-Wamboldt B. Content analysis: method, applications, and issues. *Health care for women international* 1992;13(3):313-21. doi: 10.1080/07399339209516006 [published Online First: 1992/07/01]
13. Marshall MN. Sampling for qualitative research. *Family practice* 1996;13(6):522-5. [published Online First: 1996/12/01]
14. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology & health* 2010;25(10):1229-45. doi: 10.1080/08870440903194015 [published Online First: 2010/03/06]
15. Giacomini MK, Cook DJ. Users' guides to the medical literature: XXIII. Qualitative research in health care A. Are the results of the study valid? Evidence-Based Medicine Working Group. *Jama* 2000;284(3):357-62. doi: 10.1001/jama.284.3.357 [published Online First: 2000/07/13]
16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care : journal of the International Society for Quality in Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042 [published Online First: 2007/09/18]
17. O'Brien BC, Harris IB, Beckman TJ, et al. Standards for reporting qualitative research: a synthesis of recommendations. *Academic medicine : journal of the Association of American Medical Colleges* 2014;89(9):1245-51. doi: 10.1097/acm.0000000000000388 [published Online First: 2014/07/01]
18. Genovese MC, Kremer JM, Kartman CE, et al. Response to baricitinib based on prior biologic use in patients with refractory rheumatoid arthritis. *Rheumatology (Oxford, England)* 2018;57(5):900-08. doi: 10.1093/rheumatology/kex489 [published Online First: 2018/02/08]
19. Rahman MU, Buchanan J, Doyle MK, et al. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. *Annals of the rheumatic diseases* 2011;70(9):1631-40. doi: 10.1136/ard.2010.146043 [published Online First: 2011/06/29]
20. Ferraz MB, Quaresma MR, Aquino LR, et al. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *The Journal of rheumatology* 1990;17(8):1022-4. [published Online First: 1990/08/01]

21. McAlindon T, Formica M, Schmid CH, et al. Changes in barometric pressure and ambient temperature influence osteoarthritis pain. *The American journal of medicine* 2007;120(5):429-34. doi: 10.1016/j.amjmed.2006.07.036 [published Online First: 2007/05/01]
22. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ (Clinical research ed)* 2013;346:e5595. doi: 10.1136/bmj.e5595 [published Online First: 2013/02/07]
23. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ (Clinical research ed)* 2003;326(7382):219. [published Online First: 2003/01/25]
24. Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010;11:85. doi: 10.1186/1745-6215-11-85 [published Online First: 2010/08/14]
25. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ (Clinical research ed)* 2013;346:e5793. doi: 10.1136/bmj.e5793 [published Online First: 2013/02/07]
26. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine--reporting of subgroup analyses in clinical trials. *The New England journal of medicine* 2007;357(21):2189-94. doi: 10.1056/NEJMsr077003 [published Online First: 2007/11/23]
27. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ (Clinical research ed)* 2012;344:e1553. doi: 10.1136/bmj.e1553 [published Online First: 2012/03/17]
28. Gagnier JJ, Morgenstern H, Altman DG, et al. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. *BMC medical research methodology* 2013;13:106. doi: 10.1186/1471-2288-13-106 [published Online First: 2013/09/06]
29. Pincus T, Miles C, Froud R, et al. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology* 2011;11:14. doi: 10.1186/1471-2288-11-14 [published Online First: 2011/02/02]
30. Fisher DJ, Carpenter JR, Morris TP, et al. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ (Clinical research ed)* 2017;356:j573. doi: 10.1136/bmj.j573 [published Online First: 2017/03/05]
31. Heidari S, Babor TF, De Castro P, et al. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Research integrity and peer review* 2016;1:2. doi: 10.1186/s41073-016-0007-6 [published Online First: 2016/05/03]
32. European_Medicines_Agency_(EMA). Guideline on the investigation of subgroups in confirmatory clinical trials. London, United Kingdom: European Medicines Agency (EMA), 2019.
33. Food_and_Drug_Administration_(FDA). Integrated Summary of Effectiveness: Guidance for Industry. Maryland, United States: Food and Drug Administration (FDA), 2015.
34. Food_and_Drug_Administration_(FDA). Evaluation of Sex-Specific Data in Medical Device Clinical Studies Guidance for Industry and Food and Drug Administration Staff. Maryland, United States: Food and Drug Administration (FDA), 2014.
35. Food_and_Drug_Administration_(FDA). Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies. In: Maryland US, ed.: Food and Drug Administration (FDA), 2017.
36. ICH. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). November 2019. https://www.ich.org/page/efficacy-guidelines.
37. U.S._Food_and_Drug_Administration. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. October 2017
38. EMA. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials Step 5. February 2020. https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials.

39. Welch VA, Norheim OF, Jull J, et al. CONSORT-Equity 2017 extension and elaboration for better reporting of health equity in randomised trials. *BMJ (Clinical research ed)* 2017;359:j5085. doi: 10.1136/bmj.j5085 [published Online First: 2017/11/25]
40. Welch VA, Akl EA, Guyatt G, et al. GRADE equity guidelines 1: considering health equity in GRADE guideline development: introduction and rationale. *Journal of clinical epidemiology* 2017;90:59-67. doi: 10.1016/j.jclinepi.2017.01.014 [published Online First: 2017/04/17]
41. Petkovic J, Barton JL, Flurey C, et al. Health Equity Considerations for Developing and Reporting Patient-reported Outcomes in Clinical Trials: A Report from the OMERACT Equity Special Interest Group. *The Journal of rheumatology* 2017;44(11):1727-33. doi: 10.3899/jrheum.160975 [published Online First: 2017/02/17]
42. Welch V, Petticrew M, Tugwell P, et al. PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. *PLoS medicine* 2012;9(10):e1001333. doi: 10.1371/journal.pmed.1001333 [published Online First: 2012/12/12]
43. O'Neill J, Tabish H, Welch V, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of clinical epidemiology* 2014;67(1):56-64. doi: 10.1016/j.jclinepi.2013.08.005 [published Online First: 2013/11/06]
44. Pfadenhauer LM, Gerhardus A, Mozygemba K, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implementation science : IS* 2017;12(1):21. doi: 10.1186/s13012-017-0552-5 [published Online First: 2017/02/17]
45. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. *The New England journal of medicine* 2016;374(6):504-6. doi: 10.1056/NEJMp1511701 [published Online First: 2016/02/11]
46. Oude Voshaar MAH, Das Gupta Z, Bijlsma JWJ, et al. The International Consortium for Health Outcome Measurement (ICHOM)Set of Outcomes that Matter to People Living with Inflammatory Arthritis Consensus from an international Working Group. *Arthritis care & research* 2018 doi: 10.1002/acr.23799 [published Online First: 2018/10/26]
47. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018;27(5):1171-79. doi: 10.1007/s11136-017-1765-4 [published Online First: 2017/12/21]
48. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018;27(5):1147-57. doi: 10.1007/s11136-018-1798-3 [published Online First: 2018/02/13]
49. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *Journal of clinical epidemiology* 2011;64(12):1294-302. doi: 10.1016/j.jclinepi.2011.03.017 [published Online First: 2011/08/02]
50. Sterne JAC, Savovic J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed)* 2019;366:l4898. doi: 10.1136/bmj.l4898 [published Online First: 2019/08/30]