Reporting of effect modifying contextual factors in trials: A protocol for an overview and tutorial

Collaborators/authors

Sabrina Mai Nielsen[1,2], Henning Bliddal[1], Marius Henriksen[1], Kirstine Amris[1], Eva E. Wæhrens[1,3], Danielle A. van der Windt[4], Maarten Boers[5], Annelies Boonen[6,7], Ernest H. Choy[8], Joel Gagnier[9,10], Francis Guillemin[11], Philip Mease[12], Barnaby C. Reeves[13], Josef S. Smolen[14], Vibeke Strand[15], Peter C. Taylor[16], Karine Toupin April[17], Maarten de Wit[18], Marieke Voshaar[19], Lyn March[20], George A. Wells[21], Peter Tugwell[22], Torkell Ellingsen[2], Robin Christensen[1,2]

Affiliations

1: 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: Occupational Science & Occupational Therapy, User Perspectives and Community-based Research, Department of Public Health, University of Southern Denmark, Odense, Denmark.

4: School of Medicine, Keele University, Staffordshire, UK.

5: Department of Epidemiology & Data Science; and Amsterdam Rheumatology and Immunology Center, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, The Netherlands.

6: Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre+, Maastricht, The Netherlands.

7: Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands.

8: Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, United Kingdom.

9: Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA.

10: Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA.

11: Université de Lorraine, APEMAC, 54000 Nancy, France.

12: Division of Rheumatology, Swedish Medical Center/Providence St. Joseph Health and University of Washington, Seattle, Washington, USA.

13: Bristol Trials Centre, Bristol Medical School, University of Bristol, Bristol, UK.

14: Division of Rheumatology, Department of Medicine 3, Medical University of Vienna, Vienna, Austria.

15: Division of Immunology/Rheumatology, Stanford University, Palo Alto, CA.

16: Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Headington, Oxford, UK.

17: School of Rehabilitation Sciences and Department of Pediatrics, University of Ottawa; Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.

18: OMERACT Patient Research Partner, Amsterdam, The Netherlands.

19: Department of Pharmacy, Sint Maartenskliniek, The Netherlands. Department of Pharmacy, Radboudumc, Nijmegen, The Netherlands.

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, Sydney, New South Wales, Australia.

21: School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

22: Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Corresponding Author

Dr Robin Christensen, MSc, PhD; Professor of Biostatistics and Clinical Epidemiology
Section for Biostatistics and Evidence-Based Research, 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

ABSTRACT

Background and Objectives: The average treatment effect estimate from a randomized controlled trial (RCT) may not be valid for deciding what treatment is best for individual patients with specific characteristics. Characteristics may include contextual factors related to the person (e.g. age), disease (e.g. disease duration), or environmental (e.g. healthcare system). However, individual RCTs rarely report results stratified by potentially important contextual factors in a consistent manner. In this study we will summarize existing guidance from key organizations and other research initiatives, and subsequently demonstrate how subgroup results from RCTs can be presented in a consistent manner that enables future evidence synthesis.

Study Design and Setting: We will use a tutorial-type approach to explain and demonstrate the recommended presentation of subgroup results from RCTs. In part 1, we will systematically review papers and official documents. Relevant guidance will be identified via a broad search strategy, including searching MEDLINE (via PubMed) and Web of Science, reviewing official websites of FDA, EMA, ICH, Equator Network etc., 'asking around', and snowball methods. Recommendations will be extracted and summarized. In part 2, we will demonstrate the recommendations in practice. We will use data from four previously published RCTs of rheumatic and musculoskeletal conditions from the Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark. For each RCT, we will reanalyze the data by stratifying according to proposed potential candidates of effect modifying contextual factors and apply the reporting guidance described in part 1.

Dissemination and perspectives: This study will be disseminated in a peer-reviewed publication and conference presentations. It is anticipated that this tutorial can guide researchers in reporting trial results stratified by effect modifying contextual factors in a consistent manner, allowing for investigation of subgroup effects in future meta-analyses, potentially leading to more patient-specific evidence-based management of patients with rheumatic conditions.

INTRODUCTION

Evidence-based medicine is essential for developing clinical guidelines, and, hence, shapes clinical practice. However, an average treatment effect estimate from a clinical trial may not be valid for deciding what treatment is best for an individual patient [1-3]. Using the treatment that *on average* is the most effective will do more good than harm, but it is not necessarily cost-effective [1] and it may prevent the use of alternatives that could be more effective for individual patients [4]. In fact, most common drugs have incomplete efficacy, and it has been estimated that the top ten grossing drugs in the United States improve the conditions of only 4-25% of the patients who take them [5]. This leaves a potential for exploring what characterizes the patients who benefit from a treatment and those who do not, and subsequently match patient subgroups with the treatment that best suits them. Subgroups can be defined based on dichotomous (e.g. male/female), categorical (e.g. region), ordered categorical (e.g. disease severity at baseline) or continuous (e.g. age) factors. If certain patient characteristics modify the effects of treatments, these are called 'effect modifiers', or 'predictors of treatment effect'. Designing interventions to target treatment to those who are most likely to benefit (or least likely to experience harm) by identifying effect modifiers, is termed stratified medicine (when considering subgroups) [6] or personalized medicine (when individualizing treatment decisions).

Key stakeholder organizations, such as regulatory authorities approving new medical treatments, are increasingly focusing on the issue of variation of treatment response among patients, but with varying levels of detail of their guidance [7]. The European Medicines Agency (EMA) states "*It is not acceptable to assume consistent effects across important subgroups without further investigation or discussion*". The EMA has published a dedicated guideline on the investigation of subgroups in confirmatory clinical trials [8]. In the guideline, EMA focuses on quantifying subgroup effects as part of checking that the estimated overall effect is broadly applicable to relevant subgroups of the target population. The EMA does not seem to require specific subgroup analyses, but suggests that some of the variability in treatment response between patients is caused by demographics, environmental factors, genomics and/or disease characteristics, co-morbidities, etc. The U.S. Food and Drug Administration (FDA) specifically requires that effectiveness data are analyzed by sex, age, and racial subgroups, to identify any modifications of dosing for specific subgroups [9].

Statistically significant effect modification is detected by estimating the interaction of the intervention being evaluated with the putative effect modifier. Individual RCTs are almost always designed to have adequate power to detect a difference greater than or equal to prespecified target difference between intervention and control groups. Such trials will not have the adequate power to detect the same target difference between the effect of the treatment and control in people in a subgroup compared to people in another subgroup (e.g. males vs females, or short vs long disease duration) , as a four-fold greater sample size is usually needed for an interaction to be statistically significant [4]. Designing RCTs with such large sample sizes is both costly and likely not feasible in practice. Instead, estimating subgroup effects by pooling individual RCTs in a meta-analysis can help to achieve satisfactory power [10 11], providing the subgroup effects are consistently reported in the individual RCTs (if subgroups effects are not consistently reported, individual patient data meta-analyses are required [12]). The method relies on sufficient and consistent reporting of data on treatment effects stratified by suspected effect modifiers from each trial. However, such detailed data is currently not available from most trial reports [13-17] – even baseline data for many suspected important effect modifiers are often lacking [18]. The generally accepted guideline for the reporting of RCTs, CONSORT, only mentions subgroup analyses briefly as part of potential ancillary analyses [19]. The lack of sufficient reporting in trial reports prevents any future evidence synthesis exploring possible subgroup effects. Therefore, efforts are needed to improve the reporting of subgroup analyses in trial reports – preferably, the reporting should be consistent and based on a consensus-based list of potentially important effect modifiers of high priority.

The Outcome Measures in Rheumatology (OMERACT) initiative established the Contextual Factors Working Group (CFWG) to guide the understanding, identification, and handling of so-called 'contextual factors' for RCTs in rheumatology. As part of this effort, the working group aims to list important 'effect modifying contextual factors' to consider when designing, analyzing and reporting trials. Factors of interest have been narrowed down to personal factors (e.g. age, sex, race), environmental factors (e.g. health care system, place of residence), and disease-related factors (e.g. disease duration, treatment history) [20]. Relevant candidates to consider may include the factors initially suggested by the group in 2018 [21], the factors used by the FDA (i.e. sex, age, and race) [9], and

factors of social inequity described by the acronym 'PROGRESS-PLUS' [22]. The importance of these factors must be confirmed by evidence from trials to support the future development of a consensus-based generic list of (ideally, a limited number of) important contextual factors that should always be considered in trials.

In this study we will summarize existing guidance from key organizations and other research initiatives, and subsequently demonstrate how subgroup results from RCTs can be presented in a consistent manner that enables future evidence synthesis.

METHODS

Adopting a tutorial-type approach, in part 1 we will explain recommended presentation of subgroup results from RCTs based on a review of existing guidance from key organizations and research initiatives. To illustrate this in practice, we will introduce and reanalyze data from published RCTs within rheumatic and musculoskeletal conditions according to potential candidates of effect modifying contextual factors. We developed the present protocol prior to conducting the search or any subgroup analyses. The protocol will be available online (http://www.parkerinst.dk/).

Part 1. Guidance from key organizations and research initiatives on reporting subgroup effects

Initially we will use a systematic search strategy to determine and evaluate the current guidance from key organizations and research initiatives for reporting subgroup analyses.

Eligibility criteria

Any paper or official document (e.g. by industry, health technology assessment (HTA) organizations, academic/non-profit research organizations, or regulatory bodies) that includes research guidance on reporting subgroup analyses for RCTs. Inspired by Wijn et al. [7], terms analogous to subgroup research may include (but are not limited to) the following: investigate treatment effect according to baseline characteristics, predictors of response, covariate treatment interaction, heterogeneity, effect modifying factors, stratification, moderators, split the population into

subgroups, or explicitly mention particular characteristics that should be analyzed or presented separately. The papers or official documents need to be available in English language and as full text.

Information sources

We will conduct a broad search of a range of different sources, including: MEDLINE (via PubMed) and Web of Science; websites of key organizations (e.g. Equator Network [www.equator-network.org], the U.S. Food and Drug Administration [FDA], the European Medicines Agency [EMA], and The International Council for Harmonisation [ICH]); 'snowball' methods including screening reference lists of relevant articles and of overviews on the topic, and using the option 'related articles' function in PubMed; call of key experts in the field; and look through relevant articles collected within the CFWG over the years, such as the scoping review by Wijn et al. [7].

Search

Our search terms are inspired by Wijn et al. [7].

PubMed search (MEDLINE):

(subgroup analysis[tiab] OR subgroup analyses[tiab] OR subgroup effects[tiab] OR heterogeneity[tiab] OR effect modification[tiab] OR differential effect[tiab] OR moderator[tiab] OR moderation[tiab] OR subgrouping variable[tiab] OR predictor of treatment response[tiab] OR predictors of response[tiab] OR predictor of treatment effect[tiab] OR responder analysis[tiab] OR effect modifier[tiab] OR treatment-covariate interaction[tiab] OR statistical interaction[tiab] OR stratified[tiab])

AND

(guidance[ti] OR guideline[ti] OR recommendation[ti] OR reporting[ti])

AND

(intervention research OR randomized controlled trials OR trials)

Web of Science search (using 'Web of Science Core Collection'):

# 1	TI=(subgroup analysis OR subgroup analyses OR subgroup effects OR heterogeneity OR effect modification OR differential effect OR moderator OR moderation OR subgrouping variable OR predictor of treatment response OR predictors of response OR predictor of treatment effect OR responder analysis OR effect modifier OR treatment-covariate interaction OR statistical interaction OR stratified)

# 2	AB=(subgroup analysis OR subgroup analyses OR subgroup effects OR heterogeneity OR effect modification OR differential effect OR moderator OR moderation OR subgrouping variable OR predictor of treatment response OR predictors of response OR predictor of treatment effect OR responder analysis OR effect modifier OR treatment-covariate interaction OR statistical interaction OR stratified)
# 3	TI=(guidance OR guideline OR recommendation OR reporting)
# 4	AB=(intervention research OR randomized controlled trials OR trials)
# 5	#1 OR #2
# 6	#3 AND #4 AND #5

Selection process

One researcher (SMN) will review the identified papers and documents for eligibility, supported by a senior researcher (RC).

Data collection and data items

The publication year, author/organization, title, and intended area of application will be extracted, as well as all text pieces/paragraphs involving recommendations for reporting of subgroup analyses together with the page number and/or section heading where this is located (**Table 1**). Each paper/document will be given an ID number.

ID	**Year**	**Author/ organization**	**Title**	**Area of application**	**Page/ heading**	**Recommendations**
001						
002						
…	…	…		…	…	…

Table 1: Anticipated structure of the data extraction form

Synthesis

All extracted text sections will be analyzed and categorized, and the recommendations for reporting will be compared across papers/documents. Cases of disagreement between recommendations will

be identified and discussed within the author group. In the manuscript text for publication, we will explain the overall recommended presentation of subgroup results from RCTs based on our findings. The detailed review results will be provided as supporting information in an appendix.

Part 2. Data acquisition and reanalysis of existing trials

To demonstrate the reporting of subgroup results, we will apply the recommendations in practice by reanalyzing published RCTs within rheumatic and musculoskeletal conditions according to potential candidates of effect modifying contextual factors.

Data set

We have selected four previously published RCTs with readily available data from the Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark [23-26]. The selection of trials was aimed at covering different rheumatic and musculoskeletal conditions, different interventions, and outcomes. The trials provide data on the effects of 1) weight loss intervention, 2) spiritual healing, 3) group-based multicomponent treatment course, and 4) intra-articular injection prior to exercise therapy, for either knee osteoarthritis, rheumatoid arthritis or chronic widespread pain.

The first trial, 'Diet#1' [23], randomized 96 patients with knee osteoarthritis to a weight loss group, receiving an intensive low-energy diet [LED] program, or minimal attention control group. The LED program consisted of dietary consultations and two periods of low-calorie diet during week 0-8 and week 32-36. The control group received dietary instructions and attention for two hours at baseline, and at weeks 8, 32, 36 and 52. The primary outcome was change in total Western Ontario and McMaster Universities Osteoarthritis index (WOMAC) measured at 1-year follow-up. A total of 89 were included in the modified intention-to-treat (ITT) analysis, which showed that despite a significantly larger weight loss obtained in the treatment group (-10.9 kg [SE 0.8] vs. -3.6 kg [SE 0.8]), there was no significant difference between the groups regarding change in total WOMAC (-7.3 [SE 1.9] vs. -3.0 [SE 1.9], difference of 4.3 [95%CI -0.9 to 9.6; P = 0.107]). *For our study, we will use data on the WOMAC outcome.*

The second trial, 'Spiritual Healing' (NCT00967395) [24], randomized 96 women with rheumatoid arthritis to three groups, receiving 'active healing', 'sham healing', or 'no healing'. Active healing and sham healing consisted of 8 sessions over 21 weeks. The trial was participant-, and outcome assessor–blinded. Co-primary outcomes were change in Disease Activity Score for 28 joints based on C-reactive protein (DAS28-CRP) and change in Doppler ultrasound measured at 29-weeks follow-up. All 96 participants were included in the ITT analysis, using simple non-responder imputation with baseline observation carried forward (BOCF) for missing data. The analysis showed that active healing led to a larger decrease (improvement) in DAS28-CRP compared to sham healing (−0.525 [95%CI −0.871 to −0.179] vs. 0.092 [95%CI −0.254 to 0.438], difference of -0.62 [95%CI 0.13 to 1.11; P = 0.046]). No differences between groups were found for Doppler ultrasound. *For our study, we will use data on the DAS28-CRP outcome and only for the groups receiving active healing and sham healing (i.e. leaving out the group receiving 'no healing').*

The third trial, 'IMPROvE' (NCT01352052) [25], randomized 191 women with chronic widespread pain to a 2-week group-based multicomponent treatment course or to waiting list. The course was conducted by a multidisciplinary team consisting of a rheumatologist, a psychologist, a nurse, and occupational and physical therapists, and included a daily scheduled program (35 hours in total) aimed at increasing participants' functional ability and pain coping through patient education and adjustment in everyday life. Further details on the components of the program can be found in the main publication by Amris et al. [25]. Co-primary outcomes were change in Assessment of Motor and Process Skills (AMPS) ADL motor ability, change in AMPS ADL process ability, and change in SF-36 MCS, measured at 6-months follow-up. AMPS is an individualized, observation-based evaluation of an individual's ability to perform and complete activities of daily living (ADL), where higher measures mean more ADL ability. All 191 participants were included in the ITT analysis, using non-responder imputation with BOCF for missing data. The analyses showed more improvement in the treatment group for both AMPS ADL motor and AMPS ADL process ability. For SF-36 MCS, no difference was seen between the groups. Furthermore, responder analyses were carried out showing significant group differences, e.g. 36% in the intervention group vs. 25% in the control group experienced a clinically relevant improvement (at least 0.3 logits) in the AMPS ADL motor ability measure (group

difference of 11.2%-points [95%CI 0.4% to 22.0%; P = 0.041]). *For our study, we will use data for the outcome AMPS ADL motor ability response.*

The fourth trial, 'PEX-4' (NCT01945749) [26], randomized 100 patients with knee osteoarthritis to receive either an intra-articular corticosteroid injection or a placebo isotonic saline injection. The trial was participant-, practitioner-, and outcome assessor–blinded. After two weeks, all participants started a 12-week supervised exercise program. The primary outcome was change in the pain subscale of the Knee Injury and Osteoarthritis Outcome Score (KOOS) ranging from 0-100, where higher scores indicate grater improvement. At 14-weeks follow-up, no difference in KOOS pain change was seen between the groups (13.6 [SE 1.8] vs 14.8 [SE 1.8], difference of 1.2 points [95%CI −3.8 to 6.2; P = 0.64]). *For our study, we will use data on the KOOS pain outcome.*

Effect modifying contextual factors

The potential candidates of effect modifying contextual factors measured at baseline will include the 12 factors initially suggested by the CFWG in 2018 [21], the factors used by FDA (i.e. sex, age, and race) [9], and factors of social inequity described by the acronym 'PROGRESS-PLUS' [22]. The PROGRESS factors comprise place of residence, race/ethnicity/culture/language, occupation, sex/gender, religion, education, socioeconomic status, social capital, while '-Plus' refers to additional personal characteristics associated with discriminations (e.g. age, disability), features of relationships (e.g. smoking parents, excluded from school), time-dependent relationships (e.g. leaving the hospital, respite care, other instances where a person may be temporarily at a disadvantage) [22]. Preliminary suggested subgroupings are indicated in parentheses. Further information on the individual factors including initial working definitions can be found in the appendix of Nielsen et al. 2020 [18]. The exact measures and subgroupings to use for each contextual factor will be decided based on ad hoc expert input and data availability, prior to conducting any analyses (i.e., no iterative repeated analyses will be allowed).

1. **Sex** (female/male)
2. **Comorbidity***
3. **Healthcare system***

4. **Psychological wellbeing** (<median/≥median score or above/below certain threshold for e.g. SF-36 mental health)
5. **Adherence to treatment at baseline***
6. **Age** (<median/≥median age)
7. **Previous exposure to drugs** (Exposed/Not exposed to specific relevant drugs/biological therapies)
8. **Patient education/health literacy***
9. **Disease duration** (<median/≥median duration)
10. **Race***
11. **Smoking***
12. **Pain sensitization** (present/absent based on the painDETECT questionnaire [PDQ])
13. **Support at work, family, friends** (living alone/living with someone)
14. **Socio-economic status***
15. **Occupation** (in job/not in job)
16. **Religion** (Christianity, Islam, Jewish, Muslim, Buddhist, Hindu, Other, unaffiliated)
17. **Economic status** (…)
18. **Functioning** <median/≥median score or above/below certain threshold for e.g. SF-36 physical function)
19. **Education** (short/longer education)
20. **Place of residence***

*Since the trials were not designed with all the above potential effect modifying contextual factors in mind, we do not anticipate having data available for these factors. Furthermore, sufficient heterogeneity may not be present for all the remaining factors.

Statistical methods

For each trial, we will stratify the results by each of the contextual factors. End-of-treatment data will be preferred over latest follow-up data. We will present examples of both binary and continuous outcomes. As indicated in the 'Data set' section above, we will analyze the outcomes: change in total

WOMAC index, change in DAS28-CRP, AMPS ADL motor ability response, and change in KOOS pain. We will estimate the within-subgroup differences between the intervention and comparator with corresponding 95% confidence intervals [27-29]:

For continuous outcomes

For each level of the contextual factor (within-subgroup), an effect estimate, E, a standard error, SE, and a confidence interval, 95%CI, will be calculated.

$$E_1 = mean\ change_{I,subgroup1} - mean\ change_{C,subgroup1}$$

$$SE_{E1} = \sqrt{SE^2_{mean\ change,I,subgroup1} + SE^2_{mean\ change,C,subgroup1}}$$

$$95\%CI_1 = E_1 \pm 1.96 \cdot SE_{E1}$$

For subgroup 2, the estimated E_2, SE_2, and 95%CI_2 are calculated using the same method as described above.

For binary outcomes:

For binary outcomes, similar estimates can be calculated, using risk differences (RD) based on proportion with events (risk) within the intervention group and comparator group.

$$E_1 = risk_{I,subgroup1} - risk_{C,subgroup1}$$

$$SE_{E1} = \sqrt{SE^2_{risk,I,subgroup1} + SE^2_{risk,C,subgroup1}}$$

$$95\%CI_1 = E_1 \pm 1.96 \cdot SE_{E1}$$

For subgroup 2, the estimated E_2, SE_2, and 95%CI_2 are calculated using the same method as described above.

In case of two subgroup levels, a simple test of interaction can be carried out [27 28], based on calculating a z-value from which a p-value can be found by referring to the normal distribution:

For continuous and binary outcomes

$z\ value = \frac{E_1 - E_2}{SE_{Diff.\ in\ E1\ and\ E2}}$, where $SE_{Diff.\ in\ E1\ and\ E2} = \sqrt{SE^2_{E1} + SE^2_{E2}}$

The same number of patients used for calculating the overall treatment effect in the original trial publications, will be used for the subgroup analyses to the extent data is available for the individual contextual factors.

The presentation of the results will be guided by the summary of main recommendations for reporting subgroup results, described in the previous section, and will likely consist of a forest plot for each trial.

We will use the statistical software R [30] for the analyses and develop generic codes for graphical presentation that will be available as an appendix.

Patient involvement

Two patient research partners (PRPs) are already involved in this project. We are following the recommendations of European League Against Rheumatism (EULAR) [31] and OMERACT [32], and we will take into account experiences with these recommendations [33]. We have chosen to involve two PRPs (MdW and MV), who are familiar with the topic, and who will be involved throughout the full process, including to review the protocol draft, contribute to resolving any unexpected serious issues along the way, contribute to interpretation of the results and reviewing the manuscript drafts. A total of 12 PRPs (i.e. the PRPs already involved in our working group) will be invited to comment on the study findings, whether they find the tutorial understandable, logical and easy to follow, or whether certain sections need reformulation.

Discussion

The proposed study will synthesize recommendations from key organizations and other research initiatives for reporting subgroup results and demonstrate how this can be done in practice for individual RCTs, using rheumatology as an example, to enable future meta-analyses investigating effect modifying contextual factors.

In contrast to our previous study that was investigating trial-level data [18], in this study we will promote making subgroup data from trials available, allowing future meta-analyses to investigate

patient-level contextual factors while potentially avoiding the risk of ecological bias from meta-regression using aggregated data [34].

It should be emphasized that most trials are unlikely to provide sufficient power to investigate subgroup effects and interpretation of subgroup results from most individual trials alone should therefore be avoided. Instead, the results should simply be made available for future meta-analyses of subgroup differences. In order to be able to interpret subgroup results from individual RCTs, the trials need to be sufficiently powered, the subgroup analyses should be prespecified and plausible, and the potential issue of multiple testing needs to be considered. Further guidelines on conducting and assessing the credibility of subgroup analyses from individual RCTs exist [35 36 37], as well as in systematic reviews [38 39], and will be beyond the scope of this overview and tutorial.

Statistical interactions are scale dependent, meaning that the choice of reporting an effect measure on either an additive scale (such as risk difference) or on a multiplicative scale (such as risk ratio [RR]) can impact the ability to detect a statistical interaction, or even lead to opposite conclusions [40]. Recommendations regarding the choice of scale tend to vary. EMA states: "*It is recommended that the exploration of interactions and effects in subgroups proceeds first on the scale on which the endpoint is commonly analysed. It might be useful to present supplementary analyses on the complementary scale for those covariates or subgroups where inconsistency is observed*" [8]. In contrast, based on a literature review and expert opinion, Lesko and colleagues recommend reporting effect measures on the additive scale due to ease of interpretation [41]. In our study we anticipate reporting RDs for dichotomous outcomes, as this is of interest when the aim is to identify which subgroup would benefit more from a treatment in cases where the baseline risk varies [40]. Other arguments include that causal effect modification will always be evident on an additive scale but not necessarily on a multiplicative scale, and that the additive scale is most relevant for identifying targets for intervention as it indicates absolute benefit [42]. However, reporting the number of events in all groups ensures that data will be available to calculate other effect sizes for future systematic reviews and meta-analyses.

The contextual factors investigated in this study are based on preliminary work from the CFWG, and more in-depth work is needed before a final set of contextual factors that should always

be considered in individual trials will be developed. This set should ideally include a limited number of factors to ensure feasibility and, hence, not include all 20 factors we have listed for this study. Also, factors outside the ones we have listed may be important, such as BMI and disease activity. In addition to such a set, disease- or study-specific contextual factors may be investigated as well (e.g. based on consensus within OMERACT working groups).

In our study, we will be using pragmatic cut-off points for continuous variables to form subgroups (e.g. median age and disease duration) as no standard exists, and meaningful cut-off points will likely vary depending on the disease of interest. Dichotomizing continuous variables reduces the power and may affect the magnitude of statistical significance of the results [43]. Overcoming this issue would require individual patient data to be provided from individual trials and/or results from advanced modelling of continuous variable [43] to be reported in trial reports. Instead, to make future subgroup investigations at meta-level feasible (as well as the application of the results), the pragmatic approach of forming subgroups based on continuous variables is necessary. Future evidence- and consensus-based research will be needed for determining standard cut-offs for continuous outcomes to form meaningful subgroups that can be compared across studies.

Finally, it should be noted that despite analyzing RCTs, the investigation of subgroup effects is observational, and the results may be explained by other factors associated with the contextual factors investigated.

Despite above-described limitation and identification of areas where further work is needed, we believe that the proposed study will have important implications, as it is currently common not to report stratified results by contextual factors in individual trials in a consistent manner.

Dissemination and perspectives

This study will be disseminated in a peer-reviewed publication and conference presentations. It is anticipated that this tutorial can guide researchers in reporting trial results stratified by effect modifying contextual factors in a consistent manner, allowing for investigation of subgroup effects in future meta-analyses, potentially leading to more patient-specific evidence-based management of patients with rheumatic conditions.

Contributions

SMN and RC conceived the study and developed the initial draft of the protocol.

All authors critically revised the protocol for important intellectual content and approved the final version.

RC, TE and SMN obtained funding.

SMN and RC are the guarantors.

Support

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 organizations that, since its establishment in 1983, has given grants to not-for-profit organizations 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 or writing of the protocol, and will have no role in the data collection, data analysis, data interpretation, or writing of the final manuscript.

Competing interests

All authors: No conflicts related to this study.

Ethical approval

Not required.

References

1. Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet (London, England)* 1995;345(8965):1616-9. doi: 10.1016/s0140-6736(95)90120-5 [published Online First: 1995/06/24]

2. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank quarterly* 2004;82(4):661-87. doi: 10.1111/j.0887-378X.2004.00327.x [published Online First: 2004/12/15]
3. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *Jama* 2007;298(10):1209-12. doi: 10.1001/jama.298.10.1209 [published Online First: 2007/09/13]
4. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet (London, England)* 2005;365(9454):176-86. doi: 10.1016/s0140-6736(05)17709-5 [published Online First: 2005/01/11]
5. Schork NJ. Personalized medicine: Time for one-person trials. *Nature* 2015;520(7549):609-11. doi: 10.1038/520609a [published Online First: 2015/05/01]
6. 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]
7. Wijn SRW, Rovers MM, Le LH, et al. Guidance from key organisations on exploring, confirming and interpreting subgroup effects of medical treatments: a scoping review. *BMJ open* 2019;9(8):e028751. doi: 10.1136/bmjopen-2018-028751 [published Online First: 2019/08/26]
8. European_Medicines_Agency_(EMA). Guideline on the investigation of subgroups in confirmatory clinical trials. London, United Kingdom: European Medicines Agency (EMA), 2019.
9. Food_and_Drug_Administration_(FDA). Integrated Summary of Effectiveness: Guidance for Industry. Maryland, United States: Food and Drug Administration (FDA), 2015.
10. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet (London, England)* 2005;365(9456):341-6. doi: 10.1016/s0140-6736(05)17790-3 [published Online First: 2005/01/25]
11. 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]
12. Riley RD, Debray TPA, Fisher D, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Stat Med* 2020;39(15):2115-37. doi: 10.1002/sim.8516 [published Online First: 2020/05/01]
13. Attwood S, van Sluijs E, Sutton S. Exploring equity in primary-care-based physical activity interventions using PROGRESS-Plus: a systematic review and evidence synthesis. *The international journal of behavioral nutrition and physical activity* 2016;13(1):60. doi: 10.1186/s12966-016-0384-8 [published Online First: 2016/05/22]
14. Frampton GK, Shepherd J, Dorne JL. Demographic data in asthma clinical trials: a systematic review with implications for generalizing trial findings and tackling health disparities. *Social science & medicine (1982)* 2009;69(8):1147-54. doi: 10.1016/j.socscimed.2009.06.012 [published Online First: 2009/07/14]
15. Welch V, Petticrew M, Ueffing E, et al. Does consideration and assessment of effects on health equity affect the conclusions of systematic reviews? A methodology study. *PloS one* 2012;7(3):e31360. doi: 10.1371/journal.pone.0031360 [published Online First: 2012/03/20]
16. 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]

17. Gabler NB, Duan N, Raneses E, et al. No improvement in the reporting of clinical trial subgroup effects in high-impact general medical journals. *Trials* 2016;17(1):320. doi: 10.1186/s13063-016-1447-5 [published Online First: 2016/07/18]
18. Nielsen SM, Storgaard H, Ellingsen T, et al. Population characteristics as important contextual factors in rheumatological trials: an exploratory meta-epidemiological study from an OMERACT Working Group. *Annals of the rheumatic diseases* 2020 doi: 10.1136/annrheumdis-2020-217237 [published Online First: 2020/07/02]
19. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical research ed)* 2010;340:c869. doi: 10.1136/bmj.c869 [published Online First: 2010/03/25]
20. Nielsen SM, Boers M, de Wit M, et al. OMERACT consensus-based operational definition of contextual factors in rheumatology clinical trials: A mixed methods study. *Seminars in arthritis and rheumatism* 2021 doi: 10.1016/j.semarthrit.2021.03.013 [published Online First: 2021/04/21]
21. 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]
22. 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]
23. Bliddal H, Leeds AR, Stigsgaard L, et al. Weight loss as treatment for knee osteoarthritis symptoms in obese patients: 1-year results from a randomised controlled trial. *Annals of the rheumatic diseases* 2011;70(10):1798-803. doi: 10.1136/ard.2010.142018 [published Online First: 2011/08/09]
24. Bliddal H, Christensen R, Højgaard L, et al. Spiritual healing in the treatment of rheumatoid arthritis: an exploratory single centre, parallel-group, double-blind, three-arm, randomised, sham-controlled trial. *Evidence-based complementary and alternative medicine : eCAM* 2014;2014:269431. doi: 10.1155/2014/269431 [published Online First: 2015/01/24]
25. Amris K, Waehrens EE, Christensen R, et al. Interdisciplinary rehabilitation of patients with chronic widespread pain: primary endpoint of the randomized, nonblinded, parallel-group IMPROvE trial. *Pain* 2014;155(7):1356-64. doi: 10.1016/j.pain.2014.04.012 [published Online First: 2014/04/15]
26. Henriksen M, Christensen R, Klokker L, et al. Evaluation of the benefit of corticosteroid injection before exercise therapy in patients with osteoarthritis of the knee: a randomized clinical trial. *JAMA internal medicine* 2015;175(6):923-30. doi: 10.1001/jamainternmed.2015.0461 [published Online First: 2015/03/31]
27. Matthews JN, Altman DG. Interaction 3: How to examine heterogeneity. *BMJ (Clinical research ed)* 1996;313(7061):862. doi: 10.1136/bmj.313.7061.862 [published Online First: 1996/10/05]
28. 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]
29. Christensen R, Bours MJL, Nielsen SM. Effect Modifiers and Statistical Tests for Interaction in Randomized Trials. *Journal of Clinical Epidemiology [in press]* 2021
30. A language and environment for statistical computing [program]. 3.1.2 version. Vienna, Austria: R Foundation for Statistical Computing, 2014.
31. de Wit MP, Berlo SE, Aanerud GJ, et al. European League Against Rheumatism recommendations for the inclusion of patient representatives in scientific projects. *Annals of the rheumatic diseases* 2011;70(5):722-6. doi: 10.1136/ard.2010.135129 [published Online First: 2011/01/25]

32. Cheung PP, de Wit M, Bingham CO, 3rd, et al. Recommendations for the Involvement of Patient Research Partners (PRP) in OMERACT Working Groups. A Report from the OMERACT 2014 Working Group on PRP. *The Journal of rheumatology* 2016;43(1):187-93. doi: 10.3899/jrheum.141011 [published Online First: 2015/04/17]
33. de Wit MP, Kvien TK, Gossec L. Patient participation as an integral part of patient-reported outcomes development ensures the representation of the patient voice: a case study from the field of rheumatology. *RMD open* 2015;1(1):e000129. doi: 10.1136/rmdopen-2015-000129 [published Online First: 2015/10/29]
34. Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21(3):371-87. [published Online First: 2002/01/29]
35. Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2020;192(32):E901-e06. doi: 10.1503/cmaj.200077 [published Online First: 2020/08/12]
36. van Hoorn R, Tummers M, Booth A, et al. The development of CHAMP: a checklist for the appraisal of moderators and predictors. *BMC medical research methodology* 2017;17(1):173. doi: 10.1186/s12874-017-0451-0 [published Online First: 2017/12/23]
37. Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ (Clinical research ed)* 2010;340:c117. doi: 10.1136/bmj.c117 [published Online First: 2010/04/01]
38. Gagnier JJ, Moher D, Boon H, et al. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. *BMC medical research methodology* 2012;12:111. doi: 10.1186/1471-2288-12-111 [published Online First: 2012/08/01]
39. 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]
40. VanderWeele TJ, Knol MJ. A Tutorial on Interaction. *Epidemiologic Methods* 2014;3(1):33. doi: https://doi.org/10.1515/em-2013-0005
41. Lesko CR, Henderson NC, Varadhan R. Considerations when assessing heterogeneity of treatment effect in patient-centered outcomes research. *Journal of clinical epidemiology* 2018;100:22-31. doi: 10.1016/j.jclinepi.2018.04.005 [published Online First: 2018/04/15]
42. Bours MJL. TUTORIAL: A NONTECHNICAL EXPLANATION OF THE COUNTERFACTUAL DEFINITION OF EFFECT MODIFICATION AND INTERACTION. *Journal of clinical epidemiology* 2021 doi: 10.1016/j.jclinepi.2021.01.022 [published Online First: 2021/02/07]
43. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *International journal of epidemiology* 1999;28(5):964-74. doi: 10.1093/ije/28.5.964 [published Online First: 1999/12/22]