

BMJ Open Guidance from key organisations on exploring, confirming and interpreting subgroup effects of medical treatments: a scoping review

Stan R W Wijn,¹ Maroeska M Rovers,¹ Ly H Le,² Michail Belias,³ Jeroen Hoogland,² Joanna Int'Hout,³ Thomas Debray,² Johannes B Reitsma²

To cite: 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**:e028751. doi:10.1136/bmjopen-2018-028751

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-028751>).

Received 07 January 2019
Revised 03 July 2019
Accepted 09 July 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Operating Rooms, Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, The Netherlands

²Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

³Department for Health Evidence, Radboud Institute for Health Sciences, Radboudumc, Nijmegen, The Netherlands

Correspondence to

Stan R W Wijn;
stan.wijn@radboudumc.nl

ABSTRACT

Objectives With the increasing interest in personalised medicine, the use of subgroup analyses is likely to increase. Subgroup analyses are challenging and often misused, possibly leading to false interpretations of the effect. It remains unclear to what extent key organisations warn for such pitfalls and translate current methodological research to detect these effects into research guidelines. The aim of this scoping review is to determine and evaluate the current guidance used by organisations for exploring, confirming and interpreting subgroup effects.

Design Scoping review.

Eligibility criteria We identified four types of key stakeholder organisations: industry, health technology assessment organisations (HTA), academic/non-profit research organisations and regulatory bodies. After literature search and expert consultation, we identified international and national organisations of each type. For each organisation that was identified, we searched for official research guidance documents and contacted the organisation for additional guidance.

Results Twenty-seven (45%) of the 60 organisations that we included had relevant research guidance documents. We observed large differences between organisation types: 18% (n=2) of the industry organisations, 64% (n=9) of the HTA organisations, 38% (n=8) of academic/non-profit research organisations and 57% (n=8) of regulatory bodies provided guidance documents. The majority of the documents (n=33, 63%) mentioned one or more challenges in subgroup analyses, such as false positive findings or ecological bias with variations across the organisation types. Statistical recommendations were less common (n=19, 37%) and often limited to a formal test of interaction.

Conclusions Almost half of the organisations included in this scoping review provided guidance on subgroup effect research in their guidelines. However, there were large differences between organisations in the amount and level of detail of their guidance. Effort is required to translate and integrate research findings on subgroup analysis to practical guidelines for decision making and to reduce the differences between organisations and organisation types.

Strengths and limitations of this study

- It is the first overview of research guidance on exploring, confirming and interpreting subgroup effects by key organisations.
- All organisations included in the review were contacted in an attempt to receive the most up-to-date guidance documents.
- Although organisations were included from different continents, the search was limited to English guidelines or translated documents.
- Due to the wide variety of organisations, a systematic review of all organisations and documents was not feasible.

INTRODUCTION

Individuals may respond to medical interventions in different ways due to genetic, environmental, demographic, disease, behavioural, or comorbidity variations that strengthen or weaken the outcome of a treatment. Patients with similar variations can be grouped in so-called subgroups of patients that respond better or worse to medical interventions. In order to tailor medical treatment to an individual, personalised medicine aims to provide patients with the most relevant and optimal treatment, possibly resulting in better patient outcomes and/or lower medical costs.^{1 2} Interest in subgroup effects is likely to grow as more information becomes available from biomarkers, companion diagnostics, high-resolution imaging, function tests, and genetic information. However, investigating subgroup effects can be complex and challenging as there is a natural tension between the urge to detect clinically relevant subgroup effects and avoiding incorrect claims.^{3 4}

Given both the complexity and relevance of investigating subgroup effects, many papers have been published in the scientific literature on this topic, both methodological

and statistical, with new methods to adequately detect subgroups.^{5–7} Although reaching a consensus regarding the optimal method can be challenging,⁸ several published reviews tried to provide guidance by focusing either on the current practice of subgroup analyses, giving an overview of (competing) statistical approaches or providing direction on best practices.^{3 9 10} However, it remains unclear to what extent key organisations (stakeholders with respect to the issue of subgroup analyses) translate recommendations from methodological research into practical guidance. These key organisations may be interested in subgroup effects as they either have a regulatory function, are part of an academic organisation, produce medicines and/or other treatments or generate evidence on the (cost-)effectiveness of treatments used in decision making required to determine whether treatments may be reimbursed or introduced on the market. We assume that these organisations provide guidance on subgroup research for their members, to assist them in this kind of analyses or research.

To identify and evaluate the nature of the guidance on exploring, confirming and interpreting subgroup effects during these steps, we have performed a scoping review. A scoping review is a useful instrument that can be used to generate an overview on a broad topic to highlight the nature of what is being discussed, the overlap in information and what is not discussed.¹¹ The aim of this scoping review is to discover how key organisations use findings and recommendations from methodological research in their research guidance on exploring, confirming and interpreting subgroup effects. We focus specifically on their definition of subgroup effects, level of detail in the guidance, the inclusion of factors affecting the credibility of subgroup effects and the promoting or discouraging of specific statistical approaches.

METHODS

We investigated how subgroup effect research has been included in research guidelines written by a wide range of key organisations worldwide. Since numerous organisations exist worldwide, it was not feasible to analyse all of them. Therefore, we restricted the inclusion to the major types of stakeholders in this field (see criteria below) and adopted a scoping review design.^{2 11} We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist in reporting our results.¹²

Organisations eligibility criteria

We identified four main types of organisations that represent major stakeholders with a clear interest in subgroup effects of medical treatment: (1) industry (developing the medical treatment), (2) health technology assessment (HTA) organisations (evaluating the cost-effectiveness), (3) academic/non-profit research organisations (evaluating effectiveness) and (4) regulatory bodies (health market authorisation and reimbursement). Our goal was

to capture all organisations that are involved throughout the lifecycle of a medical treatment from concept to implementation.

Information sources

We selected a limited number of organisations focusing on the following regions: global (worldwide organisations), North America (USA and Canada), Europe (UK, Germany and The Netherlands) and Asia-Pacific (Japan and Australia). Other regions were not considered because of their limited number of HTA organisations/regulatory bodies, due to language restrictions and to limit the scope of this review.¹³

Patient and public involvement

Patients and/or public were not involved.

Search strategy

Initially, we adopted a stakeholder list of Makady *et al* describing 23 key organisations.¹⁴ We then gradually expanded that list of potential organisations by contacting methodology experts in the field of subgroup analysis to identify specific organisations within these organisation types and completed the list by means of a literature search. The methodological experts were identified via the Cochrane collaboration methods groups, key opinion papers on this topic and HTA organisations. The key opinion papers are listed in online supplementary S1. Next, we searched for similar organisations across regions (eg, Food and Drug Administration (FDA) in USA vs European Medicines Agency (EMA) in Europe). We excluded organisations if: (1) they work in specific medical areas (eg, food safety, surgery, cardiology, oncology or infectious diseases); (2) their mission was not related to the purpose of this review, (3) they only collect guidelines from other organisations, rather than produce their own or (4) they only had non-English documents. For each organisation, we used the following strategy to determine whether they had official documents in which they provide research guidance: (1) using the search engine of their official website, (2) Google, (3) PubMed and (4) citations. Documents were considered eligible if they contained methodological research guidance and included 'subgroup', or terms analogous to the definition of subgroup research: investigate treatment effect according to baseline characteristics, covariate treatment interaction, heterogeneity, effect modifying factors, split the population into subgroup or explicitly mention particular characteristics should be analysed or presented separately (eg, gender). The research strategy is included in online supplementary S2.

Additionally, all organisations were contacted by email or contact form, asking whether they have any research guidance on subgroup effects or an updated version of the ones we already found. If there was no response after one reminder email, we discontinued the search for that specific organisation.

A guidance document was considered to be official when it clearly named the organisation, displayed the logo prominently or included statements about publication rights. We excluded books, talks or lectures written by individual author(s) even when they worked or conducted projects for the organisation. When an organisation published multiple versions of the same document, we only included the most recent version.

Data extraction

Out of the included documents, we extracted data on the following topics: general information about the organisation, purpose and context of the guidance documents, definitions and concepts of subgroup analysis challenges and recommendations for and against specific methods. Data extraction was performed in two stages: first, all documents were scanned to detect whether they mention the predefined topics. Second, we analysed the documents in more detail. Any information on subgroup research was extracted. Subsequently, the 10 credibility criteria described by Sun *et al* to investigate the credibility of authors' claims of subgroup effects were used to determine if and how subgroup analysis credibility criteria were reported.⁵ These criteria were used because they were up to date, widely accepted, well defined, and the multiple criteria enabled us to compare the documents. Although these criteria were developed to assess the credibility of subgroup effects in randomised controlled trials, we used these to determine whether these topics were mentioned in any document, including documents aimed at non-randomised trials or systematic reviews. The criteria were based on previous published subgroup credibility criteria that were focused on subgroup analyses in general and are thus also applicable to other types of data.^{15 16} The data extraction method, including the definitions of general and detailed information, are described in online supplementary S3.

Analysis

Using descriptive statistics, similarities and differences between the types of organisations and regions were analysed. The frequency of similar recommendations was analysed and compared. We analysed the frequency at which the credibility criteria for subgroup analysis were mentioned and calculated the average number of criteria mentioned in the documents for the four organisation types.⁵

RESULTS

From the list of organisations of Makady *et al* and initially received suggestions, 40 organisations were included. Thirty-four organisations were added after expert consultation and inclusion of counterpart organisations. In total, after removing organisations outside the scope of our search strategy, we found 60 organisations operating globally or in one of the prespecified regions or countries that matched our inclusion criteria (figures 1 and 2).

In the end, we included 11 industry organisations, 14 HTA organisations, 21 academic/non-profit research organisations and 14 regulatory bodies divided over global organisations (23%), Europe (32%), North America (27%) and Asia-Pacific (18%). Out of the 60 organisations, 27 (45%) had relevant research guidance documents and 26 (43%) responded to our inquiry. Five organisations responded but stated that they had no documents and six organisations of whom we received no response had documents online available. The availability of documents differed among organisation types, with 18%, 64%, 38% and 57% for industry, HTA, academic/non-profit and regulatory organisations, respectively.

Looking at regional differences, Europe and North America had relevant research guidance documents in half of the organisations we included, with 53% and 50% of the organisations, respectively, while Asia-Pacific had research guidance in 27% of the organisations. In total, we retrieved two documents from industry organisations,^{17 18} 19 documents from HTA organisations,¹⁹⁻³⁷ 12 documents from academic organisations³⁸⁻⁴⁹ and 19 documents from regulatory organisations.⁵⁰⁻⁶⁸ Eleven documents were excluded because they were not in line with our research objective, including presentations, advertisement guidelines or political documents. The list of the included organisations and documents can be found in online supplementary S4.

Document details

The general characteristics of the included documents are described in table 1. The mean and median year of publication of the documents were 2011 and 2014 (range: 1988–2018), respectively. HTA and regulatory organisations had the most documents (n=19) and industry organisations the least (n=2). A separate section on subgroup research was present in 0%, 42%, 17% and 42% of the industry, HTA, academic and regulatory documents, respectively. Both, a definition of subgroups and heterogeneity, appeared more often when a separate subgroup section was available. Regulatory bodies included the most clinical trial guidance of the organisation types (n=19, 37%) but had no guidance on non-randomised studies. In total, out of the 52 documents, five were aimed at non-randomised studies (10%), three by HTA and 2 by academic organisations. HTA organisations were the only type of organisation that had documents for cost-effectiveness studies (n=10, 19%). Both included documents from industry organisations were aimed at clinical trials.

Types of subgroup effects

The term subgroup effects can be related to a range of different concepts. Therefore, organisations typically distinguish between different types of subgroup effects, such as subgroup effects from exploratory analysis, subgroup effects from confirmatory analysis, qualitative effects and quantitative effects.^{10 69} These concepts are elucidated in box 1.

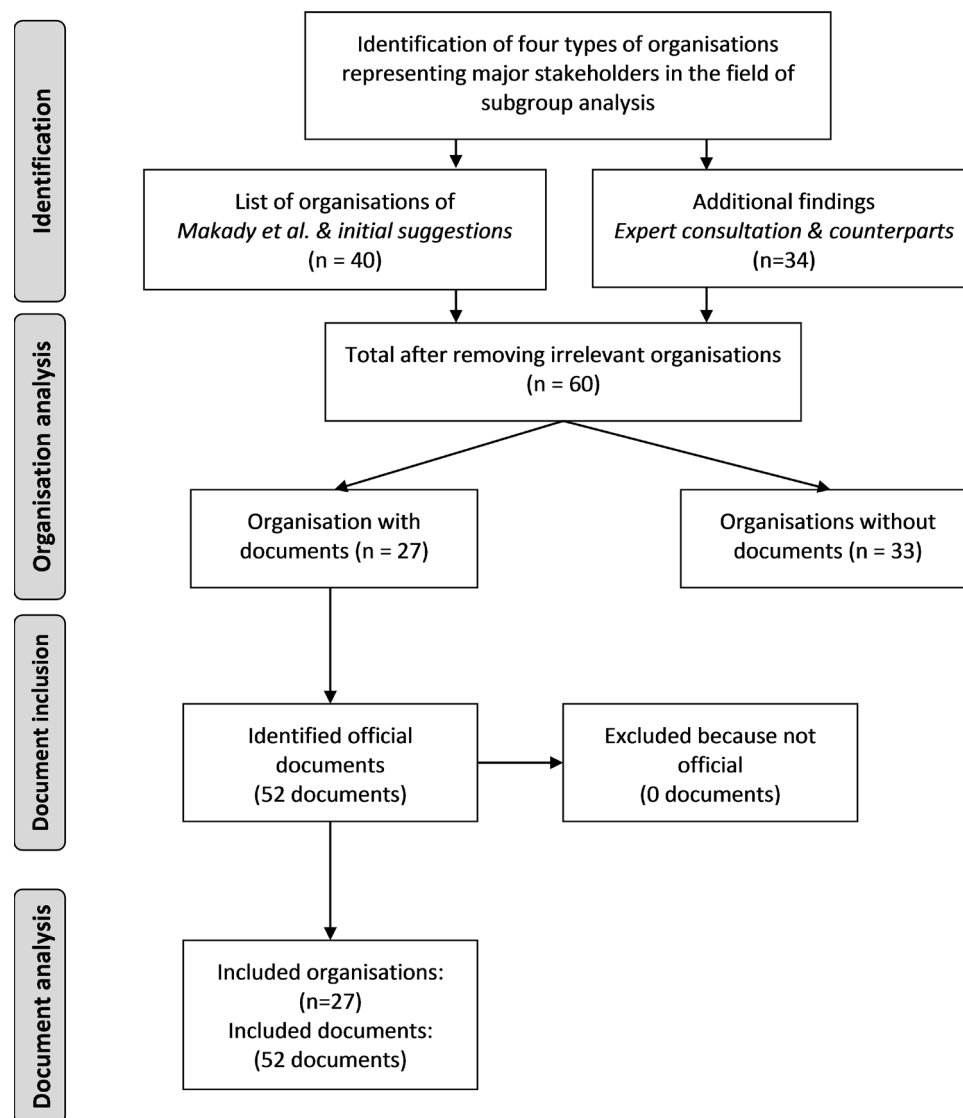


Figure 1 PRISMA flow chart of organisation and document inclusion and analysis. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Subgroup effects from exploratory analyses were mentioned most often by the organisations, in 23 (44%) of all documents, divided in 0%, 58%, 25% and 47% for industry, HTA, academic and regulatory organisations, respectively. Guidance for exploratory subgroup analysis mostly advised to compare different treatment effects in gender or age.

Subgroup effects from confirmatory subgroup analyses were described in 13 documents (25%), divided in 0%, 42%, 17% and 16% for industry, HTA, academic and regulatory organisations. One document suggested to report the 95% prediction interval in confirmatory subgroup analysis next to the regular 95% CI in order to reflect the variation of the treatment/subgroup effect in a different setting.

Qualitative subgroup effects and quantitative subgroup effects were rarely mentioned. Only HTA organisations mentioned both concepts in three different documents (14% of all HTA guidelines, 6% of all documents).

Regarding the actual clinical relevance of the subgroup effect, no document in our review highlighted the importance of reporting absolute differences between subgroups. That is, a significant test result (to detect a subgroup) may not lead to a change in treatment decisions, whereas in the absence of a significant result a clinically relevant subgroup may still exist.

Challenges in subgroup analysis

Performing adequate subgroup analysis can be difficult due to several statistical and methodological challenges. Informing researchers and other stakeholders about the challenges in subgroup analysis can be a way to improve the outcome of the analyses. The most common challenges are: type I errors, multiplicity problems, lack of power and ecological bias (explained in [box 1](#)).

Typically, these challenges were not regularly mentioned or discussed. Overall, the risk of a false positive (type I error) and multiple testing problem (multiplicity) were

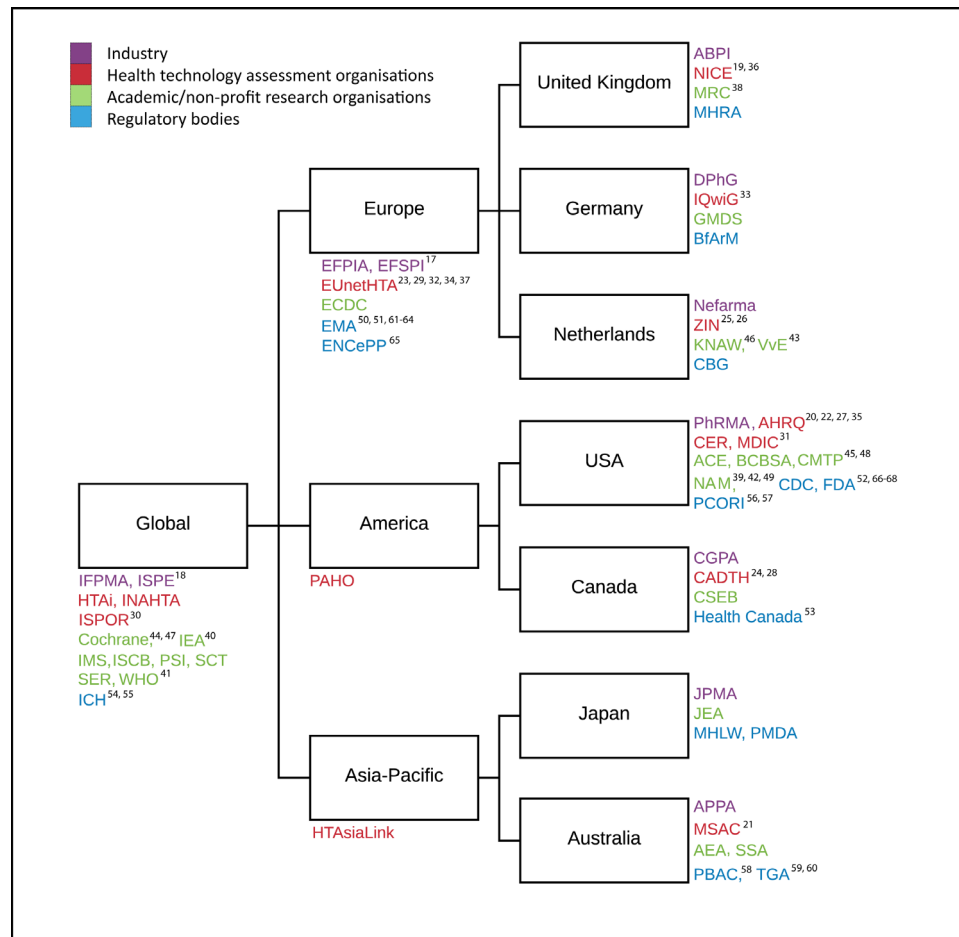


Figure 2 Overview of organisations and their corresponding type and geographic.

most frequently mentioned, by 12 (23%) and 13 (25%) documents, respectively. Some documents suggest that limited testing or multiplicity adjustment is required. Next to the multiplicity and false positive issue, five documents (10%) mentioned the lack of power to detect real subgroup effects due to small sample sizes, increasing the risk of false negatives. Ecological bias was mentioned in three documents (6%). Lack of power and ecological bias was only mentioned by HTA and academic organisations.

Credibility criteria

In 1992, years before the publication of most documents included in our review, Oxman and Guyatt suggested seven criteria to assess subgroup effects.¹⁶ Sun *et al* updated the list resulting in 10 criteria for randomised trials to evaluate the credibility of subgroup analyses.⁵ These include: (1) hypothesis specified a priori, (2) subgroup variable a baseline characteristic, (3) subgroup stratification factor at randomisation, (4) independent interactions, (5) consistent with previous studies, (6) consistent across outcomes, (7) small number of tests, (8) significant p value of interaction test, (9) indirect evidence to support apparent effect and (10) direction of subgroup effect prespecified. We assessed how the documents reflect the credibility criteria. In total, 34 documents (62%) mentioned one or more credibility criteria.

On average, HTA documents included 4.1 criteria, while the averages for industry, academic and regulatory organisation documents were lower with 1.5, 2.3 and 1.8 criteria, respectively (as shown in figure 3).

Of the 10 criteria, 'Hypothesis specified a priori' and 'Subgroup variable a baseline characteristic' were addressed most frequently, in 33 (63%) and 29 (56%) of the documents, respectively. 'Direction of subgroup effect prespecified' was mentioned the least (six times, 12%).

Statistical recommendations

Statistical recommendations were often lacking, and when present, not up-to-date with current standards. The recommendations were limited to a formal test of interaction using a threshold of $p < 0.05$ (16 documents (31%)) and/or stratified analyses (nine documents (17%)). One document described the option to use Bayesian models in a subgroup analysis.³⁰ In total, 37% of the documents provided statistical recommendations. There were no recommendations against certain statistical tests or methods.

Publication checklists

There are several checklists/guidelines for reporting trials or reviews created to strengthen the reporting of medical studies. These checklists are often combined

Table 1 General characteristics of the included documents

Type	Industry (N=11)	HTA (N=14)	Academic (N=20)	Regulatory (N=15)	Total (N=60)
Total documents (n)	2	19	12	19	52
Documents per organisation ratio	0.18	1.36	0.60	1.27	0.87
Number of pages (median, first/third quartile)	9 (9/9)	113 (75/219)	71 (39/162)	39 (20/81)	74 (36/162)
'Subgroup' mentioned (median, first/third quartile)	1 (1/1)	24 (11/41)	5 (2/17)	12 (8/21)	13 (5/27)
'Effect modification' mentioned at least once (n, %)	0 (0)	8 (42)	2 (17)	3 (16)	13 (25)
'Treatment covariate interaction' mentioned at least once (n, %)	0 (0)	2 (11)	0 (0)	2 (11)	4 (8)
Specific subgroup paragraph (n, %)	0 (0)	8 (42)	2 (17)	8 (42)	18 (35)
Mention heterogeneity (n, %)	0 (0)	11 (58)	5 (42)	8 (42)	24 (46)
Include a subgroup definition (n, %)	0 (0)	9 (47)	3 (25)	8 (42)	20 (38)
Research purpose:* (n, %)					
Cost-effectiveness study (n, %)	0 (0)	10 (53)	0 (0)	0 (0)	10 (19)
Clinical trial (n, %)	2 (100)	9 (47)	7 (58)	19 (100)	37 (71)
Non-randomised study (n, %)	0 (0)	3 (16)	2 (17)	0 (0)	5 (10)
Systematic review (n, %)	0 (0)	9 (47)	5 (42)	7 (37)	21 (40)

Note: Presented as frequency.

*Some documents had multiple research purposes.

HTA, health technology assessment; N, number of organisations; n, number of documents.

with a statement that includes best practices for research publication, including optimal methods for subgroup analysis reporting.

In our search, we found that 13 documents (25%) recommended the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) checklist for systematic reviews, 12 documents (23%) recommended the Consolidated Standards of Reporting Trials (CONSORT) checklist for randomised trials, 10 documents (19%) recommended the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for observational studies and the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement for economic evaluations was mentioned in two documents (4%). Documents that described a checklist often described more than one; 29 documents (56%) did not recommend any checklist or guidance for publication.

Individual patient data

Like we mentioned before, detecting patient subgroups can be difficult due to the limited number of patients that have been included in studies. To overcome this problem, researchers may perform a review to identify and analyse the individual patient data from multiple trials. This approach is also known as individual patient data meta-analysis (IPDMA).⁷⁰ Of the 52 documents, five (9.6%) describe the IPDMA as a method to detect

subgroups, and only one document provided methodological guidance on IPDMA.

DISCUSSION

This scoping review provides insight into the adoption of methodological subgroup effects research in guidelines, both by international and national key organisations. Investigating subgroup effects can be challenging and resulted in the publication of several non-reproducible subgroup effects.^{71–73} With the growing discussion on adequate subgroup analysis,^{4 10} proper guidance could reduce the number of non-reproducible findings and improve individual patient care. However, only 45% of the 60 organisations included in our review had some guidance documents on subgroup analysis research, with large differences between organisation types. Overall, HTA organisations provided most documents with detailed and up-to-date methodological information, while industry organisations provided hardly any guidance documents, possibly because these organisations use research guidance from other organisations (eg, regulatory organisations). Organisations that included subgroup analysis in their guidelines often only mention subgroup research as a tool to evaluate heterogeneity between (sub)groups.

Box 1. Summary of types of subgroup effects and challenges in subgroup analyses

Types of subgroup effects:

- Confirmatory: subgroup effects from confirmatory analyses are subgroup effects that have been detected in previous trials and require confirmation of the effect in another population that is sufficiently powered to detect the subgroup, increasing credibility of the subgroup effect.
- Exploratory: subgroup effects from exploratory analyses are subgroup effects that are detected during post hoc analysis in a trial or review, without prior knowledge of the effect or predefining the direction of the subgroup effect.
- Qualitative subgroup effects: treatment effects that are beneficial to one subgroup but harmful in another.
- Quantitative subgroup effects: difference in size of treatment effect between subgroups but not the direction. For example, a treatment can be beneficial across different age ranges but to a larger extent for the elderly (as compared to adolescents).

Challenges in subgroup analyses:

- Type I error (false positive): statistical tests for detecting subgroup effects typically involve a null hypothesis that assumes absence of treatment–covariate interaction. The significance level is the probability of rejecting the null hypothesis when it is true. For instance, when the significance level is set to 0.05 (5%), then, if there is indeed no treatment–covariate interaction, there is a 5% chance of incorrectly rejecting the null hypothesis. This probability to make a type I error can be decreased by lowering the significance level, which also decreases the power to detect a genuine treatment–covariate interaction.
- Type II error (false negative): a type II error occurs when the null hypothesis is false (ie, treatment–covariate interaction is present) but is not rejected. The probability to make a type II error is related to the power and can be reduced by increasing the significance level or the sample size.
- Multiplicity: when multiple (related) hypothesis tests are performed in a single dataset, the probability of at least one type I error increases. This is a common problem when assessing effect modification by more than 1 covariate. If the study is not properly powered to detect treatment–covariate interaction, it can also increase the probability of making at least one type II error (failure to reject a false null hypothesis: a false negative). Multiplicity problems can be resolved by statistical multiplicity correction (eg, based on family-wise error rates) or reducing the number of (exploratory) subgroup analyses (which need to be prespecified).
- Power: the power is the probability that an effect is found assuming there is an effect. When a study is not powered for subgroup analyses, splitting the study population into different subgroups could result in a lack of power. This means that the risk of a type II error increases (false negative finding).
- Ecological bias: ecological bias may occur in systematic reviews when the presence of treatment–covariate interaction is assessed across studies (rather than within studies). For instance, meta-regression can be used to assess whether published treatment effect estimates are associated with a particular population characteristic (eg, the proportion of males in each trial). Although the presence of such trial-level association is commonly treated as evidence of treatment–covariate interaction, this is often misleading. For instance, even if gender does not have any effect on treatment efficacy, a trial-level association between treatment effect and gender

Continued

Box 1 Continued

may still appear when trials that included relatively more males adopted a higher treatment dosage for the active comparator.

Our assumption was that most organisations would include some guidance on how to perform credible subgroup analysis, both for exploratory and confirmatory reasons. However, most documents lacked detail, especially regarding statistical and methodological recommendations. Given that research in personalised medicine is growing, we hope that key organisations will update and refine their guidance and at least include the credibility criteria of Sun *et al.*⁵

Multiple sources for research guidance can be useful, however, it can also increase confusion if the suggestions for subgroup research differ between documents and/or key organisations and do not include the latest recommendations. More consistency in recommendations would help to reduce reproducibility issues and likely improve reporting standards. Perhaps we should strive to a more coherent set of guidance distinguishing between different main purposes for performing subgroup analyses.

In our search for official guidance documents, some organisations had papers that stated that the findings and conclusions in these documents were those of the authors and that the findings and conclusions did not necessarily represent the views of the key organisation. Still, we included those documents in our review because they were located on the official website and often were presented with an official logo of the organisation/company.

Although most documents were written in the last 10 years, a delay can exist between methodological publications and the implementation of this research into guidelines. However, methodological and statistical difficulties

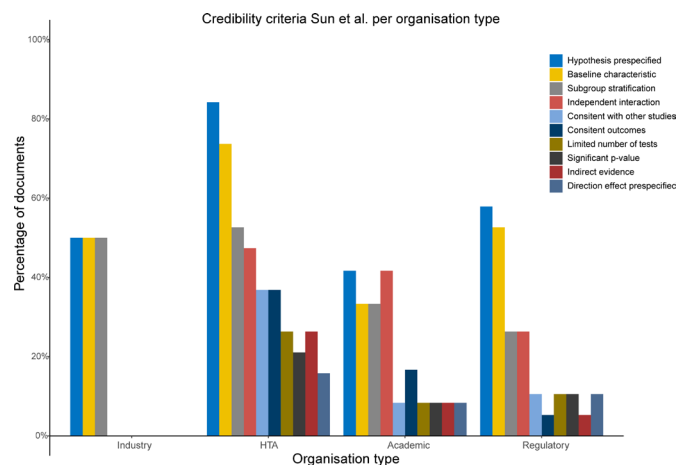


Figure 3 Total number of credibility criteria per organisation type. Hypothesis specified a priori was mentioned most often. HTA organisations included the most credibility criteria, while industry organisations rarely included these criteria. HTA, health technology assessment.

in subgroup analyses have been reported over the last three decades.^{16 73 74}

Strengths and limitations

This is the first scoping review targeted at guidance on exploring, confirming and interpreting subgroup effects by key organisations. The scoping review is a useful instrument that can be used to generate an overview of available guidance documents by different organisations in different regions of the world. As recommended by Levac *et al*, we kept the research question broad and simultaneously considered the different concepts of interest.⁷⁵ To balance between feasibility and comprehensiveness, we arranged regular meetings with methodological experts to discuss and justify decisions and, if needed, adapt the search strategy.

Some potential limitations should also be discussed. First, our review was targeted at research guidance documents. To search for these documents, we primarily used the official website of the concerning organisations. This might have limited the number of documents in our review to documents that were publicly available on their website. To overcome this, we tried to contact all organisations in an attempt to receive the most up-to-date documents. We limited the number of reminder emails to two to ensure the practicality of the scoping review. Second, out of the 26 organisations that responded to our inquiry, 21 (81%) had documents. This might indicate that organisations without documents might be less likely to respond to our email or that they had documents but not openly available and were not willing to respond. This could lead to bias as we marked these latter organisations as 'no documents', which would be an underestimation of the results.

Third, although we included organisations from different continents, we limited our search to English guidelines or translated documents. This might have had a negative impact on articles from Asia-Pacific, possibly explaining why we found fewer organisations and guidelines from Asia-Pacific, compared with North America and Europe. Fourth, data extraction and primary analysis were performed by one reviewer of our research team. It might have improved the accuracy of the results if it was done by multiple researchers. However, this was not feasible for our review. To ensure the quality of the review, we discussed all discrepancies and difficulties with the research team during all steps of the reviewing process.

In conclusion, less than half of the organisations included in our scoping review did provide some kind of guidance on exploring, confirming and interpreting subgroup effects in their guidelines. Furthermore, information was often minimal, and methodological and statistical guidance of best current practices was missing. To further improve and speed-up personalised medicine, more effort is required to uniformly disseminate and integrate research findings on subgroup analysis into practical guidelines for decision making.

Acknowledgements This work was in part supported by the "The University of Sydney – Utrecht University Partnership Collaboration Awards".

Contributors MMR, JH, TD and JBR have contributed to the conception and design of the study. All authors have contributed to the final design and focus of the scoping review. Reviewing has been performed by LHL and SRWW. Scoring and data analysis has been performed by SRWW. SRWW drafted the manuscript. All authors have made contributions to the drafting and revising of the article. All authors have read, reviewed and approved the final version of the manuscript before submission.

Funding This study was funded by a TOP grant by the Netherlands Organisation for Health Research and Development (ZonMw) Number: 91215058.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Search strategies and data extraction documents are available on request to the corresponding author.

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

REFERENCES

- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793–5.
- Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Heal Inf Libr J* 2009;26:91–108.
- Sun X, Briel M, Busse JW, *et al*. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. *BMJ* 2011;342.
- Senn S. Statistical pitfalls of personalized medicine. *Nature* 2018;563:619–21.
- Sun X, Briel M, Busse JW, *et al*. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012;344:e1553.
- Burke JF, Sussman JB, Kent DM, *et al*. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015;351.
- Wang S-J, Hung HMJ. A regulatory perspective on essential considerations in design and analysis of subgroups when correctly classified. *J Biopharm Stat* 2014;24:19–41.
- European Medicines Agency (EMA). European Medicines Agency workshop on the investigation of subgroups in confirmatory clinical trials [Internet], 2015. Available: <https://www.youtube.com/playlist?list=PL7K5dNgKwZld20zYEhLs2ByEP3Pm9BU> [Accessed cited 21 Nov 2018].
- Sun X, Ioannidis JPA, Agoritsas T, *et al*. How to use a subgroup analysis. *JAMA* 2014;311.
- Dmitrienko A, Muysers C, Fritsch A, *et al*. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J Biopharm Stat* 2016;26:71–98.
- Rumrill PD, Fitzgerald SM, Merchant WR. Using scoping literature reviews as a means of understanding and interpreting existing literature. *Work* 2010;35:399–404.
- Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 2018;169:467.
- Hill S, Velazquez A, Tay-Teo K, *et al*. Global survey on health technology assessment by National authorities. *WHO* 2015;2015:1–40.
- Makady A, de Boer A, Hillege H, *et al*. What is real-world data? A review of definitions based on literature and Stakeholder interviews. *Value Health* 2017;20:858–65.
- Sun X, Briel M, Walter SD, *et al*. Is a subgroup effect believable? updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010;340:c117–4.
- Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78–84.
- Fletcher C, Driessen S, Burger HU, *et al*. European Federation of Statisticians in the pharmaceutical industry's position on access to clinical trial data. *Pharm Stat* 2013;12:333–6.
- Public Policy Committee, International Society of Pharmacoeconomics. Guidelines for good pharmacoepidemiology practice (Gpp). *Pharmacoepidemiol Drug Saf* 2016;25:2–10.

19. Dias S, Sutton AJ, et al. *NICE DSU Technical Support Document 3 : Heterogeneity : Subgroups, Meta-Regression, Bias and Bias-Adjustment*. London, United Kingdom: National Institute for Health and Care Excellence (NICE), 2011.
20. AHRQ Publication No. 10(14)-EHC063-EF. Methods guide for effectiveness and comparative effectiveness reviews. In: . Rockville, Maryland, United States: Agency for Healthcare Research and Quality (AHRQ), 2014.
21. Medical Services Advisory Committee (MSAC). Technical guidelines for preparing assessment reports for the medical services Advisory Committee – service type: investigative. Canberra, Australia Medical Services Advisory Committee (MSAC); 2017.
22. AHRQ Publication. Methods guide for medical test reviews. Rockville, Maryland, United States Agency for Healthcare Research and Quality (AHRQ); 2012.
23. Swedish Council on Health Technology Assessment (SBU) / Sweden. Methods for health economic evaluations - A guideline based on current practices in Europe. Diemen, Netherlands European network for health technology assessment (EUNETHTA); 2015.
24. Wells GA, Sultan SA. Indirect evidence: indirect treatment comparisons in meta-analysis. Ottawa, Canada Canadian Agency for Drugs and Technologies in Health (CADTH); 2009.
25. Staal P, Heymans J, et al. *Beoordeling stand van de wetenschap en praktijk*. Diemen, Netherlands: Zorginstituut Nederland (ZIN), 2015.
26. Knies S, Dragt NM, Enzing JJ, Hermen LAH, Vijgen SMC. Guideline for economic evaluations in healthcare. Diemen, Netherlands Zorginstituut Nederland (ZIN); 2016.
27. Dreyer NA, Nourjah P, Velentgas P. Developing a protocol for observational comparative effectiveness research: a user's guide. Rockville, Maryland, United States Agency for Healthcare Research and Quality (AHRQ); 2013.
28. Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. Ottawa, Canada Canadian Agency for Drugs and Technologies in Health (CADTH); 2017.
29. Institute for Quality and Efficiency in Health Care (IQWiG) / Germany. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness. Diemen, Netherlands European network for health technology assessment (EUNETHTA); 2017.
30. Ramsey SD, Wilke RJ, Glick H, et al. Cost-Effectiveness analysis alongside clinical trials II-An ISPOR good research practices Task force report. *Value Health* 2015;18:161–72.
31. MDIC PCBR Project steering committee. Medical device innovation Consortium patient centered benefit-risk project report: a framework for incorporating information on patient preferences regarding benefit and risk into regulatory assessments of new medical technology. Arlington, Virginia, United States Medical Device Innovation Consortium (MDIC); 2015.
32. Kleijnen S, Goettsch W. Relative effectiveness assessment (REA) of pharmaceuticals background review. Diemen, Netherlands European network for health technology assessment (EUNETHTA); 2011.
33. Institute for Quality and Efficiency in Health Care. General Methods - Version 5.0. Cologne, Germany Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG); 2017.
34. UMIT (University for Health Sciences, Medical Informatics and Technology). *Therapeutic medical devices*. Diemen, Netherlands: European network for health technology assessment (EUNETHTA), 2015.
35. Dreyer N, Leavy M, Gliklich R. Registries for Evaluating Patient Outcomes : A User's Guide. Rockville, Maryland, United States Agency for Healthcare Research and Quality (AHRQ); 2014.
36. Earnshaw J, Lewis G. Nice guide to the methods of technology appraisal. *Pharmacoeconomics* 2008;26:725–7.
37. Garrido MV, Kristensen FB. Health technology assessment and health policy-making in Europe. current status, challenges and potential. Diemen, Netherlands European network for health technology assessment (EUNETHTA); 2008.
38. Medical Research Council (MRC). Developing and evaluating complex interventions: new guidance. London, United Kingdom Medical Research Council (MRC); 2008.
39. National Academy of Medicine (IOM / NAM). *Sharing clinical trial data: maximizing benefits, minimizing risk*. Washington, D.C., United States: The national academies press, 2015.
40. International Epidemiological Association (IEA). IEA Council Handbook. Chicago, Illinois, United States International Epidemiological Association (IEA); 2015.
41. World Health Organization (WHO). *Handbook for Guideline development*. 2nd ed. World Health Organization (WHO), 2014.
42. National Academy of Medicine (IOM / NAM). *Finding what works in health care: standards for systematic reviews*. Washington, D.C., United States: The national academies press, 2011.
43. The Netherlands Epidemiological Society (VvE). *Responsible epidemiologic research practice*. Leiden, Netherlands: The Netherlands Epidemiological Society (VvE), 2017.
44. Higgins J, Lasserson T, Chandler J, Tovey D, Churchill R. Methodological expectations of Cochrane intervention reviews. London, United Kingdom Cochrane; 2018: 69.
45. Green Park Collaborative. *Best practices for the design, implementation, analysis, and reporting of oncology trials with high rates of treatment switching: a guidance document from the green Park collaborative*. Baltimore, Maryland, United States: Center for Medical Technology Policy (CMTP), 2016.
46. Royal Netherlands Academy of Arts and Sciences (KNAW). *Evaluation of new technology in health care. in need of guidance for relevant evidence*. Amsterdam, Netherlands: Royal Netherlands Academy of Arts and Sciences (KNAW), 2014.
47. O'Connor D, Green S, Higgins JP. *Cochrane Handbook for systematic reviews of interventions: cochrane book series*. West Sussex, England, United Kingdom: John Wiley & Sons Ltd, 2008.
48. Velentgas P, Mohr P, Mohr P, Messner DA. Making informed decisions: assessing the strengths and weaknesses of study designs and analytic methods for comparative effectiveness research. Baltimore, Maryland, United States Center for Medical Technology Policy (CMTP); 2012.
49. National Academy of Medicine (IOM / NAM). Discussion framework for clinical trial data sharing: guiding principles, elements, and activities. Washington, D.C., United States The national academies press; 2014.
50. European Medicines Agency (EMA). Guideline on adjustment for baseline covariates in clinical trials. London, United Kingdom European Medicines Agency (EMA); 2015.
51. European Medicines Agency (EMA). Guideline on the investigation of subgroups in confirmatory clinical trials. London, United Kingdom European Medicines Agency (EMA); 2014.
52. Food and Drug Administration (FDA). *Good review practice: clinical review of investigational new drug applications*. Silver Spring, Maryland, United States: Food and Drug Administration (FDA), 2013.
53. Health Canada. Considerations for inclusion of women in clinical trials and analysis of data by sex. Ottawa, Canada Health Canada; 2012.
54. ICH Expert Working Group. ICH Harmonised Tripartite Guideline - Statistical Principles for Clinical Trials E9. Geneva, Switzerland International Conference on Harmonisation (ICH); 1998.
55. ICH Expert Working Group. *ICH Harmonised Tripartite Guideline - Structure and Content of Clinical Study Reports*. Step 4. Geneva, Switzerland: International Conference on Harmonisation (ICH), 1995.
56. Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. The PCORI methodology report. Washington, D.C., United States Patient-Centered Outcomes Research Institute (PCORI); 2017.
57. Patient-Centered Outcomes Research Institute (PCORI). *The PCORI methodology standards academic curriculum*. Washington, D.C., United States: Patient-Centered Outcomes Research Institute (PCORI), 2016.
58. Pharmaceutical Benefits Advisory Committee (PBAC). Guidelines for preparing a submission to the pharmaceutical benefits Advisory Committee. Canberra, Australia Pharmaceutical Benefits Advisory Committee (PBAC); 2016.
59. Therapeutic Goods Administration (TGA). Note for guidance on good clinical practice. Canberra, Australia Therapeutic Goods Administration (TGA); 2000.
60. Therapeutic Goods Administration (TGA). Note for guidance on clinical safety data management: definitions and standards for Expedited reporting Therapeutic Goods Administration (TGA); 2000.
61. European Medicines Agency (EMA). ICH Topic E 3 - Structure and Content of Clinical Study Reports. London, United Kingdom European Medicines Agency (EMA); 2009.
62. European Medicines Agency (EMA). Points to consider on application with 1. meta-analyses; 2. one pivotal study. London, United Kingdom European Medicines Agency (EMA); 2001.
63. European Medicines Agency (EMA). Guideline on multiplicity issues in clinical trials (draft). London, United Kingdom European Medicines Agency (EMA); 2016.
64. European Medicines Agency (EMA). ICH Step 5 - Gender Considerations in the Conduct of Clinical Trials. London, United Kingdom European Medicines Agency (EMA); 2005.
65. European Network of Centres for Pharmacovigilance and Pharmacovigilance (ENCePP). Guide on methodological standards in pharmacoepidemiology (revision 5). London, United Kingdom

- The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP); 2015.
66. Food and Drug Administration (FDA). Integrated summary of effectiveness - Guidance for industry. Rockville, Maryland, United States U.S. Department of Health and Human Services Food and Drug Administration; 2015.
 67. Food and Drug Administration (FDA). *Guideline for the format and content of the clinical and statistical sections of an application*. Rockville, Maryland, United States: Food and Drug Administration (FDA), 1988.
 68. Food and Drug Administration (FDA). Evaluation of sex-specific data in medical device clinical studies - Guidance for industry and food and drug administration staff. Rockville, Maryland, United States Food and Drug Administration (FDA); 2014.
 69. Yusuf S, Wittes J, Probstfield J, *et al*. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93–8.
 70. Tierney JF, Vale C, Riley R, *et al*. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med* 2015;12:e1001855.
 71. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 2018;45:k4245.
 72. Hernández AV, Boersma E, Murray GD, *et al*. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading? *Am Heart J* 2006;151:257–64.
 73. Assmann SF, Pocock SJ, Enos LE, *et al*. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *The Lancet* 2000;355:1064–9.
 74. Cui L, Hung HMJ, Wang SJ, *et al*. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002;12:347–58.
 75. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.