



Editorial

Same Old Challenges in Subgroup Analysis—Should We Do More About Methods Implementation?

Stefan Schandelmaier, MD, PhD; Gordon Guyatt, MD, MSc

Sherry et al¹ considered the methodological quality of subgroup analyses reported in 379 oncology trials. The authors explored a number of fundamental problems with these analyses: in forest plots, use of linear rather than logarithmic scales and failure to present overall pooled estimates; failure to conduct and highlight tests of interaction; and readiness to make unwarranted inferences regarding the credibility of postulated subgroup effects. They reported a distressingly high frequency of each of these problems.¹

The study thus adds to an already large body of subgroup literature finding that authors' presentations of forest plots are often suboptimal. It also adds indirect evidence for the often-suspected abuse of subgroup analyses for post hoc data dredging in search of interesting findings²: trials that failed to show a significant effect reported more subgroup analyses.¹ The primary message of the study is, therefore, that the methodological limitations of subgroup analyses and misleading inferences in oncology trials remain the same as those that the methods community has been discussing for more than 3 decades.^{3,4} We will briefly discuss these key limitations in subgroup analyses.

+ Related article

Author affiliations and article information are listed at the end of this article.

Failure to Design and Interpret Subgroup Analyses in the Light of Prior Knowledge

In only 10% of the subgroup claims did the primary study authors consider prior evidence or prespecify hypotheses.¹ In failing to do so, they fail to inform their audience whether the claim is consistent with prior knowledge. This information is critical for evaluating the credibility of the authors' subgroup claim.

Not Reporting the Essential Statistic

The test of interaction is the single most crucial statistic for subgroup analysis: it tells us the extent to which chance can explain the apparent difference in effect sizes across subgroups. Sherry et al¹ found that only 17% of the trials reported a *P* value or estimate of interaction, leaving evidence users in the dark regarding the extent to which subgroup differences were compatible with random error (and they mostly are⁵).

Ignoring Multiplicity Issues

The more hypotheses one tests, the more likely one will capitalize on a chance finding and then claim a spurious subgroup effect. With an average number of 9 subgroup analyses per oncology trial,¹ these subgroup analyses run a high risk of being misled—and misleading their audience—by the play of chance.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Poor Statistical Practice of Dichotomizing Continuous Effect Modifiers

For continuous variables such as age, trial authors set thresholds and reported effects on patients above and below their chosen threshold. A superior alternative would be to examine whether effects differ across the range of the continuous variable. The choice of a single threshold results in a high risk of further capitalizing on the play of chance, especially when choosing a threshold that maximizes apparent difference between groups, and weakening the analysis through discarding the extra information that the continuum provides.

And finally, as Sherry et al¹ confirmed, none of the included trials applied credibility criteria for subgroup effects that have been available since the 1990s.^{3,4,6} We recently refined these criteria in the first formal rigorously developed instrument for judging the credibility of subgroup effects.⁷

What can we do to improve the methodological quality of subgroup analyses? Learning from the past, publishing more commentaries, meta-studies, simulation studies, and guidance papers about the challenges and solutions in subgroup analysis, even if they are largely repetitive, seems unlikely to help.

One strategy, a systematic greater focus on methods implementation, has thus far failed to attract the attention it might deserve. A small number of pioneering studies have started to identify and better understand barriers to methods implementation⁸⁻¹⁰ identify and test strategies for better methods implementation,¹¹ and raise the issue of whether the principles of implementation science could work in the methods context.¹²

One simple application of implementation science would be to make it easier for investigators to identify methods guidance relevant to their studies.¹³ Another would be to improve methods guidance by encouraging those providing such guidance to involve their target audience in ensuring the accessibility of the guidance they provide.¹⁴ Reporting guidelines represent another approach that has demonstrated appreciable—although still perhaps somewhat disappointing—improvement in study design and methods implementation.

We might utilize the full potential of reporting guidelines by including more methodological details (eg, items addressing the typical limitations of subgroup analyses we have highlighted) in the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guideline and corresponding items in the Consolidated Standards of Reporting Trials (CONSORT) guideline. Some investigators might, however, find the complexity associated with these requirements daunting. Such concerns highlight the need for more systematic research to identify and understand the main barriers (lack of awareness, training, opportunity) to guide the development—and subsequent testing—of new strategies for methods implementation.

It remains speculative whether shifting some of the notoriously scarce resources for methodological research toward implementation science would eventually result in better research quality. The substantial potential for improvement, as Sherry et al¹ highlighted for subgroup analyses in oncology trials, and the obvious failure of our current approaches, suggest that now may be a good time to give it a try.

ARTICLE INFORMATION

Published: March 28, 2024. doi:[10.1001/jamanetworkopen.2024.3339](https://doi.org/10.1001/jamanetworkopen.2024.3339)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2024 Schandelmaier S et al. *JAMA Network Open*.

Corresponding Author: Stefan Schandelmaier, MD, PhD, Division of Clinical Epidemiology, Department of Clinical Research, University Hospital and University of Basel, Totengässlein 3, 4051 Basel, Switzerland (s.schandelmaier@gmail.com).

Author Affiliations: Division of Clinical Epidemiology, Department of Clinical Research, University Hospital and University of Basel, Basel, Switzerland (Schandelmaier); MTA-PTE Lendület Momentum Evidence in Medicine Research Group, Medical School, University of Pécs, Pécs, Hungary (Schandelmaier); Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada (Guyatt); Department

of Medicine, McMaster University, Hamilton, Ontario, Canada (Guyatt).

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Sherry AD, Hahn AW, McCaw ZR, et al. Differential treatment effects of subgroup analyses in phase 3 oncology trials from 2004 to 2020. *JAMA Netw Open*. 2024;7(3):e243379. doi:10.1001/jamanetworkopen.2024.3379
2. 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:d1569. doi:10.1136/bmj.d1569
3. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266(1):93-98. doi:10.1001/jama.1991.03470010097038
4. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116(1):78-84. doi:10.7326/0003-4819-116-1-78
5. Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern Med*. 2017;177(4):554-560. doi:10.1001/jamainternmed.2016.9125
6. Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? updating criteria to evaluate the credibility of subgroup analyses. *BMJ*. 2010;340:c117. doi:10.1136/bmj.c117
7. 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*. 2020;192(32):E901-E906. doi:10.1503/cmaj.200077
8. Pullenayegum EM, Platt RW, Barwick M, Feldman BM, Offringa M, Thabane L. Knowledge translation in biostatistics: a survey of current practices, preferences, and barriers to the dissemination and uptake of new statistical methods. *Stat Med*. 2016;35(6):805-818. doi:10.1002/sim.6633
9. Hughes KL, Clarke M, Williamson PR. A systematic review finds Core Outcome Set uptake varies widely across different areas of health. *J Clin Epidemiol*. 2021;129:114-123. doi:10.1016/j.jclinepi.2020.09.029
10. Coffey T, Williamson PR, Gillies K; Trials Methodology Research Partnership Working Groups. Understanding implementation of findings from trial method research: a mixed methods study applying implementation frameworks and behaviour change models. *Trials*. 2024;25(1):139. doi:10.1186/s13063-024-07968-3
11. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open*. 2019;9(5):e026589. doi:10.1136/bmjopen-2018-026589
12. King KM, Pullmann MD, Lyon AR, Dorsey S, Lewis CC. Using implementation science to close the gap between the optimal and typical practice of quantitative methods in clinical science. *J Abnorm Psychol*. 2019;128(6):547-562. doi:10.1037/abn0000417
13. Hirt J, Schönenberger CM, Ewald H, et al. Introducing the Library of Guidance for Health Scientists (LIGHTS): a living database for methods guidance. *JAMA Netw Open*. 2023;6(2):e2253198. doi:10.1001/jamanetworkopen.2022.53198
14. Hirt J, Ewald H, Lawson DO, Hemkens LG, Briel M, Schandelmaier S. A systematic survey of methods guidance suggests areas for improvement regarding access, development, and transparency. *J Clin Epidemiol*. 2022;149:217-226. doi:10.1016/j.jclinepi.2022.05.005