

EDITORIAL



When to believe a subgroup analysis: revisiting the 11 criteria

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2022

Eye (2022) 36:2075–2077; <https://doi.org/10.1038/s41433-022-01948-0>

Evidence-based surgery relies on the results of randomised controlled trials (RCTs). The judicious design, analysis, and reporting of RCTs allow surgeons to effectively use the results in routine practice [1–3]. Since the included population in an RCT is not homogenous a priori, treatment effects might vary across different subgroups. Thus, assessing treatment outcome heterogeneity across subgroups and identifying patient characteristics that may modify the effect of the intervention under investigation has become common practice [4]. The subgroup analyses, if true, might have important implications for surgical practice but are often proven to be unreliable and are criticised for their ability to turn negative results into positive ones [5, 6]. Here, we revisit the 11 criteria (Table 1) introduced by Sun et al. [7] and provide literature case examples to illustrate important principles and concepts in the interpretation of a subgroup analysis and to guide researchers in deciding their credibility in the ophthalmological literature.

Subgroup analyses are either planned a priori before randomisation or they emerge after randomisation (post-hoc) [4]. The former is credible if planned based on a prespecified hypothesis, if there is a justified direction of the overall and subgroup effect, and if there is appropriate statistical testing for the underlying hypothesis. For instance, an RCT of 702 patients (Protocol V of the DRCR net) with diabetic macular oedema and good visual acuity (VA) was designed to assess the effect of initial management with aflibercept or laser photocoagulation on vision loss versus observation [8] but failed to find significant changes in VA from either treatment versus observation in the overall population or across the predefined subpopulations. The magnitude and direction of the overall and subgroup effects are likely predictable if they are hypothesised based on a sound biological and clinical plausibility [7, 9]. Post-hoc subgroup analyses, in contrast, are data driven and are considered exploratory or hypothesis generating. Their credibility is compromised by the effect of intervention and lack of statistical power [7, 9].

Simultaneous subgroup analyses create multiplicity, inflating the defined nominal significance level (alpha) [10] which increases the likelihood of spurious and compelling results by chance alone [1]. To combat this, it is recommended to prespecify few highly relevant subgroups, use appropriate statistical tests to examine interactions between treatment effect and subgroup variables, and ensure *p*-values are adjusted for multiple testing [1, 7, 11]. The interaction test determines if treatment effects differ between different subgroups with the assumption that the true effect is the same across each subgroup category [1, 7, 12, 13]. The smaller the *p*-value of the interaction test, the stronger the subgroup effect. For instance, the CATT [14] conducted a non-inferiority trial to compare the efficacy of ranibizumab versus bevacizumab on either a monthly scheduled or an as needed regimen in patients with neovascular age-related macular degeneration and found equivalent gain in VA by treatment and dosing regimen at 1 year.

Each of the monthly scheduled treatment groups were then rerandomized into monthly scheduled or as needed regimen. The Year-2 CATT [15] assessing 2-year effects of the four original groups and the impact of switching from monthly scheduled to an as needed regimen found a similar gain in VA between treatment groups [1.4 letters difference; 95% CI –0.8, 3.7] but greater gain in the monthly scheduled regimen [2.4 letters difference; 95% CI 0.1, 4.8]. The difference did not exceed the non-inferiority margin of 5 letters. To increase power and precision, the treatment and scheduling effects were analysed between treatment and scheduling regimens (interactions *P*-value of ≥ 0.10 for non-inferiority hypothesis) rather than the effects of each drug by scheduling regimen type. The *p*-value for interaction is rarely reported in the ophthalmology literature, making the independence of subgroup effects uncertain [8, 16].

Given differences in the administration of surgical treatments and extent of biological variability, the interaction between treatment effect and various patient variables should be interpreted with caution [1]. The strength with which an inference is made on subgroup effects largely relies on the magnitude of the difference [9, 17]. That is, as the magnitude of treatment effect increases for a subgroup, the likelihood of a real subgroup difference rises. The validity of a subgroup analysis largely depends on reporting all of the conducted subgroup analyses regardless of their statistical significance [1] as well as consistency of the treatment effect across closely related outcomes [9]. A pooled analysis of two RCTs of 107 patients with highly relapsing neuromyelitis optica spectrum disorder [18] illustrates effective adherence to these principles in its design. The study found that the improvements in aggregated proptosis and diplopia responses from teprotumumab intravenous infusions compared to placebo were large and consistent, both in the overall population and across several predefined subgroups.

Arguably, the consistency of the subgroup effects in subsequent well-designed trials provide stronger credibility. Subgroup effects are also more credible if the comparison was made within a study rather than across multiple studies with different methodological qualities [9]. Planning subgroups based on the current understanding of biological mechanisms by anticipating pathophysiological, genetic, or biological heterogeneity [3] is equally important. The accounting for these criteria may be infeasible considering the heterogeneity of intervention, rarity of patient population and poor reporting quality of RCTs in the literature [2, 19]. For instance, a meta-analysis of 17 RCTs examining the effect of omega-3 fatty acid supplementation for the treatment of dry eye disease [20] reported a significant decrease in dry eye symptoms from daily omega-3 fatty acid supplementation with 96% heterogeneity. Post-hoc subgroup analyses by country showed significantly larger treatment effects in trials from India compared to elsewhere. One possible explanation was the predominant vegetarian diet and low intake of omega-3 fatty acids in India. Another explanation might be that five of six trials from India were conducted by the same group of authors on similar setting and population.

Received: 12 January 2022 Revised: 17 January 2022 Accepted: 18 January 2022
Published online: 31 January 2022

Table 1. CRITERIA in assessing the credibility of subgroup analysis.

CRITERIA in assessing the credibility of subgroup analysis
a. Design
1. Was subgroup variable measured before or after randomisation?
2. Was the hypothesis specified a priori? Was the planned statistical test appropriate for the underlying hypotheses?
3. Was the direction of effect specified a priori?
4. Was the subgroup effect one of a small number of tested hypotheses? Was the subgroup effect adjusted for the number of tested hypotheses?
5. Was the subgroup effect suggested by comparisons within rather than between studies?
b. Analysis
6. Did the interaction test suggest a low likelihood of chance explaining the apparent subgroup effect?
7. Was the statistical significance of the subgroup effect independent?
c. Interpretation and context
8. Was there a large subgroup effect? Were all subgroup analyses reported?
9. Was the interaction consistent across closely related outcomes within the study?
10. Was the interaction consistent across studies?
11. Was there indirect evidence supporting the hypothesised interaction (biological plausibility)?

Adapted from Sun et al. [7] with permission.

Well-designed surgical RCTs adequately assess the effectiveness and safety of new surgical treatments in the overall population, but reliable analysis of treatment effects across subpopulations has been slow to adapt [1]. Surgical RCTS should provide a thorough investigation of the benefits and harms of a new treatment in the overall population and key subpopulations. This editorial highlights the 11 criteria as a general guide for clinician readers of evidence regarding its use in clinical settings, but researchers interested in systematic reviews and individual research planning could consider following ICEMAN [21] as a more comprehensive instrument.

Forough Farrokhyar^{1,2}, Philip Skorzewski², Mark R. Phillips¹, Sunir J. Garg³, David Sarraf⁴, Lehana Thabane^{1,5}, Mohit Bhandari^{1,2}, Varun Chaudhary^{1,2} and for the Retina Evidence Trials InterNational Alliance (R.E.T.I.N.A.) Study Group*
¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada. ²Department of Surgery, McMaster University, Hamilton, ON, Canada. ³Mid Atlantic Retina, The Retina Service of Wills Eye Hospital, Philadelphia, PA, USA. ⁴Retinal Disorders and Ophthalmic Genetics, Stein Eye Institute, University of California, Los Angeles, Los Angeles, CA, USA. ⁵Biostatistics Unit, St. Joseph's Healthcare-Hamilton, Hamilton, ON, Canada. *A list of authors and their affiliations appears at the end of the paper. ✉email: vchaudh@mcmaster.ca

REFERENCES

- Dijkman B, Kooistra B, Bhandari M. How to work with a subgroup analysis. *Can J Surg.* 2009;52:515–22.
- Farrokhyar F, Karanicolas PJ, Thoma A, Simunovic M, Bhandari M, Devereaux PJ, et al. Randomized controlled trials of surgical interventions. *Ann Surg.* 2010;251:409–16.
- Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* 2005;365:176–86.
- Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials. *J Biopharm Stat.* 2016;26:71–98.
- Brand KJ, Hapfelmeier A, Haller B. A systematic review of subgroup analyses in randomised clinical trials in cardiovascular disease. *Clin Trials.* 2021;18:351–60.
- Chan AW. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med.* 2008;5:e230.
- 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.
- Baker CW, Glassman AR, Beaulieu WT, Antoszyk AN, Browning DJ, Chalam KV, et al. Effect of Initial Management With Aflibercept vs Laser Photocoagulation vs Observation on Vision Loss Among Patients With Diabetic Macular Edema Involving the Center of the Macula and Good Visual Acuity: A Randomized Clinical Trial. *Jama* 2019;321:1880–94.
- Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med.* 1992;116:78–84.
- Cook DJ, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust.* 2004;180:289–91.
- Ferreira JC, Patino CM. Subgroup analysis and interaction tests: why they are important and how to avoid common mistakes. *J Bras Pneumol.* 2017;43:162.
- Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *Bmj* 2003;326:219.
- Matthews JN, Altman DG. Statistics notes. Interaction 2: Compare effect sizes not P values. *Bmj* 1996;313:808.
- Martin DF, Maguire MG, Ying GS, Grunwald JE, Fine SL, Jaffe GJ. Ranibizumab and bevacizumab for neovascular age-related macular degeneration. *N. Engl J Med.* 2011;364:1897–908.
- Martin DF, Maguire MG, Fine SL, Ying GS, Jaffe GJ, Grunwald JE, et al. Ranibizumab and Bevacizumab for Treatment of Neovascular Age-related Macular Degeneration: Two-Year Results. *Ophthalmology* 2020;127:S135–s45.
- Zhang C, Zhang M, Qiu W, Ma H, Zhang X, Zhu Z, et al. Safety and efficacy of tocilizumab versus azathioprine in highly relapsing neuromyelitis optica spectrum disorder (TANGO): an open-label, multicentre, randomised, phase 2 trial. *Lancet Neurol.* 2020;19:391–401.
- Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *Bmj.* 2012;344:e1553.
- Kahaly GJ, Douglas RS, Holt RJ, Sile S, Smith TJ. Teprotumumab for patients with active thyroid eye disease: a pooled data analysis, subgroup analyses, and off-treatment follow-up results from two randomised, double-masked, placebo-controlled, multicentre trials. *Lancet Diabetes Endocrinol.* 2021;9:360–72.
- Lai TY, Wong VW, Lam RF, Cheng AC, Lam DS, Leung GM. Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. *Ophthalmic Epidemiol.* 2007;14:390–8.
- Giannaccare G, Pellegrini M, Sebastiani S, Bernabei F, Roda M, Taroni L, et al. Efficacy of Omega-3 Fatty Acid Supplementation for Treatment of Dry Eye Disease: A Meta-Analysis of Randomized Clinical Trials. *Cornea.* 2019;38:565–73.
- Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, 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:E901–e6.

AUTHOR CONTRIBUTIONS

FF was responsible for writing, critical review and feedback on manuscript. PS was responsible for writing and critical review on manuscript. MRP was responsible for conception of idea, critical review and feedback on manuscript. SJG was responsible

for critical review and feedback on manuscript. DS was responsible for critical review and feedback on manuscript. LT was responsible for critical review and feedback on manuscript. MB was responsible for conception of idea, critical review and feedback on manuscript. VC was responsible for conception of idea, critical review and feedback on manuscript.

COMPETING INTERESTS

FF: Nothing to disclose. PS: Nothing to disclose. MRP: Nothing to disclose. SJG: Consultant: Allergan, Apellis, Bausch and Lomb, Boehringer Ingelheim, Johnson and Johnson, Kanaph; Research funds: American Academy of Ophthalmology, Apellis, Boehringer Ingelheim, NGM Bio, Regeneron—unrelated to this study. DS: Consultant: Amgen, Bayer, Genentech, Novartis, Optovue; Research funds: Amgen, Genentech, Heidelberg, Optovue, Regeneron, Topcon—unrelated to this study. LT: Nothing to

disclose. MB: Research funds: Pendopharm, Bioventus, Acumed—unrelated to this study. VC: Advisory board member: Alcon, Roche, Bayer, Novartis; Grants: Bayer, Novartis—unrelated to this study.




ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Varun Chaudhary.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

FOR THE RETINA EVIDENCE TRIALS INTERNATIONAL ALLIANCE (R.E.T.I.N.A.) STUDY GROUP

Varun Chaudhary ^{1,2,✉}, Mohit Bhandari ^{1,2}, Charles C. Wykoff^{6,7}, Sobha Sivaprasad⁸, Lehana Thabane ^{1,5}, Peter Kaiser⁹, David Sarraf⁴, Sophie J. Bakri¹⁰, Sunir J. Garg³, Rishi P. Singh^{11,12}, Frank G. Holz¹³, Tien Y. Wong^{14,15} and Robyn H. Guymer^{16,17}

⁶Retina Consultants of Texas (Retina Consultants of America), Houston, TX, USA. ⁷Blanton Eye Institute, Houston Methodist Hospital, Houston, TX, USA. ⁸NIHR Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, UK. ⁹Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ¹⁰Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA. ¹¹Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. ¹²Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. ¹³Department of Ophthalmology, University of Bonn, Bonn, Germany. ¹⁴Singapore Eye Research Institute, Singapore, Singapore. ¹⁵Singapore National Eye Centre, Duke-NUS Medical School, Singapore, Singapore. ¹⁶Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, Australia. ¹⁷Department of Surgery, (Ophthalmology), The University of Melbourne, Melbourne, VIC, Australia.