Journal of Clinical Epidemiology 150 (2022) 72−79

Journal of Clinical Epidemiology

Subgroup analyses in randomized controlled trials frequently categorized continuous subgroup information

S. Faye Williamson[a], Michael J. Grayling[a], Adrian P. Mander[b], Nurulamin M. Noor[c], Joshua S. Savage[d], Christina Yap[e], James M.S. Wason[a,*]

[a]*Biostatistics Research Group, Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK*
[b]*Centre for Trials Research, Cardiff University, Cardiff, UK*
[c]*Medical Research Council Clinical Trials Unit at University College London (MRC CTU at UCL), London, UK*
[d]*Cancer Research UK Clinical Trials Unit (CRCTU), Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK*
[e]*Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK*

Accepted 28 June 2022; Published online 2 July 2022

Abstract

Background and Objectives: To investigate how subgroup analyses of published Randomized Controlled Trials (RCTs) are performed when subgroups are created from continuous variables.

Methods: We carried out a review of RCTs published in 2016−2021 that included subgroup analyses. Information was extracted on whether any of the subgroups were based on continuous variables and, if so, how they were analyzed.

Results: Out of 428 reviewed papers, 258 (60.4%) reported RCTs with a subgroup analysis. Of these, 178/258 (69%) had at least one subgroup formed from a continuous variable and 14/258 (5.4%) were unclear. The vast majority (169/178, 94.9%) dichotomized the continuous variable and treated the subgroup as categorical. The most common way of dichotomizing was using a pre-specified cutpoint (129/169, 76.3%), followed by a data-driven cutpoint (26/169, 15.4%), such as the median.

Conclusion: It is common for subgroup analyses to use continuous variables to define subgroups. The vast majority dichotomize the continuous variable and, consequently, may lose substantial amounts of statistical information (equivalent to reducing the sample size by at least a third). More advanced methods that can improve efficiency, through optimally choosing cutpoints or directly using the continuous information, are rarely used.

Keywords: Categorization; Continuous variables; Dichotomization; Moderator analysis; Randomized controlled trials; Subgroup analysis

1. Introduction

Randomized Controlled Trials (RCTs) are the gold standard for evaluating the benefit and harm of an intervention. RCTs usually focus on the difference in pre-specified outcomes between treatment arms in a trial population, which informs whether an intervention is beneficial *overall*. However, some interventions may behave differently in different types (or 'subgroups') of patients. For example: 1) tumour mutations are associated with the effect of targeted oncology treatments [1,2]; 2) ferritin level is associated with the effect of tocilizumab in hospitalized COVID-19 patients [3]; 3) greater baseline visual learning is predictive of the effect of ketamine for major depressive disorder [4].

As understanding of diseases evolves, there is an increasing number of subtypes being identified that may have highly variable outcomes and responses to treatment.

Declaration of interests: All authors declare no competing interests.

Funding: This work was supported by the MRC (MR/S014357/1). JSS was supported by a funding grant from Cancer Research UK (C22436/A25354).

Conflicts of interest: Authors report no conflicts of interest.

Author Contributions: S. Faye Wiliamson: Data curation, formal analysis, investigation, visualization, writing − original draft preparation Michael Grayling: investigation, methodology, writing − reviewing and editing Adrian Mander: investigation, writing − reviewing and editing Nurulamin Noor: investigation, writing − reviewing and editing Joshua Savage: investigation, writing − reviewing and editing Christina Yap: funding acquisition, investigation, writing − reviewing and editing James Wason: Conceptualization, funding acquisition, investigation, writing − original draft preparation.

* Corresponding author. Population Health Sciences Institute, Baddiley-Clark Building, Richardson Road, Newcastle upon Tyne NE2 4AX, UK. Tel.: +44 0191 208 6000.

E-mail address: james.wason@newcastle.ac.uk (J.M.S. Wason).

https://doi.org/10.1016/j.jclinepi.2022.06.017

What is new?

Key findings

- Randomised Controlled Trials (RCTs) often report subgroup analyses in which the subgroup is formed from a continuous variable but analysed as binary.

What this adds to what was known

- To our knowledge, this review is the first to investigate the frequency of continuous subgroup-defining variables and how they are analysed.
- Categorisation of continuous variables is the most adopted approach in practice (94.9%).
- More sophisticated methods that use the continuous information directly are rarely used in practice.

What is the implication and what should change now

- More RCTs should conduct and report analyses that use the continuous information, rather than relying on categorization.

This has led to an emerging area, known as 'precision medicine' (also referred to as individualized, personalized or stratified medicine) [5]. Precision medicine aims to identify patient subgroups that are most likely to benefit from an intervention, and those that will likely receive no benefit (or may be harmed). Even when it is not expected that different subgroups will experience varying levels of benefit, it may be of interest to demonstrate that an intervention has a consistent effect across subgroups [6]. This is typically achieved by performing a subgroup analysis, which investigates the effect of intervention(s) vs. control in different subgroups to detect differential treatment effects. For each relevant subgroup-defining patient variable, the treatment effect can be estimated (which we refer to as 'separate analyses') and tested for significance within the subgroup, although this is generally not recommended in isolation [7]. When assessing heterogeneity of treatment effect between subgroups, an interaction test [8] is appropriate. The most informative approach will implement both separate analyses and an interaction test.

Several publications have investigated subgroup analysis in RCTs. Sun et al. [9] showed that subgroup analyses are common and a high proportion claim significant subgroup effects. However, interpreting and reporting subgroup analyses can be challenging [10] and findings are often misleading due to the increased risk of false positives (owing to unadjusted multiple comparisons) and false negatives (owing to inadequate statistical power) [7,11]. Consequently, there are guidelines available to aid the design, analysis and reporting of subgroup analyses (e.g., [8,12,13]), as well as recommended criteria to evaluate the credibility of a subgroup finding (e.g., [14,15]). Key recommendations include: 1) pre-specifying a small number of subgroups that are well-justified (e.g., based on observed data in previous studies or biological plausibility); 2) appropriately adjusting for multiplicity; 3) reporting all subgroup analyses; 4) using interaction tests instead of just separate analyses. An overview of guidance documents from different regulatory agencies and key organizations is summarized by Wijn et al. [11]. Provided appropriate guidance is followed, several regulatory agencies advocate the use of subgroup analyses in RCTs [16,17].

There are two potential approaches to conduct a subgroup analysis when the subgroup-defining variable is based on continuous information. The first would be to split the continuous variable into two (dichotomization) or more (categorization) discrete subgroups, then use standard categorical subgroup analysis methods. For example, adult participants could be dichotomized into obese or nonobese according to whether their body mass index is above or below a certain cutpoint (e.g., [18]). The treatment effect could be estimated for each of these categories together with an interaction between treatment and category. A second approach would be to assess the interaction between the continuous subgroup-defining variable and treatment effect directly, via a regression model with parameters representing the effect per unit increase in the variable, treatment arm allocation, and interaction between them. The latter approach allows direct estimation of the interaction parameter (representing the change in treatment effect from a one-unit increase in the continuous variable). It is also possible to plot how the estimated treatment effect from the model varies with the level of a variable.

There are benefits and drawbacks to the two approaches. Categorizing a continuous subgroup-defining variable will be required to permit precision medicine (i.e., to use an intervention on some patients but not others). Categorization is also considered to simplify the analysis and, from a clinical perspective, lead to results that are easier to understand and communicate [19,20]. In addition, there may be cases where dichotomizing reduces the impact of measurement error, compared to using continuous information directly [21]. From a statistical perspective, categorizing is not a natural way to analyze continuous data and drawbacks have been discussed extensively in the literature [22]. One main limitation is the loss of information, which results in a lower statistical power (equivalent to discarding at least one-third of the sample [23]). The risk of false positive results may also increase [24]. Furthermore, there is the issue of choosing the cutpoint for categorization, which is generally poorly reported [25]. The cutpoint can be prespecified according to previous studies or well-established clinical criteria. However, this may not lead to optimal statistical properties being used, especially if the cutpoint results in a subgroup with a low proportion of trial

participants [26]. Instead, cutpoints could be chosen in a data-dependent way e.g., by using the median of the observed distribution of the subgroup-defining variable. However, this complicates the comparison of results between different studies [27]. Approaches that avoid using cutpoints may be more powerful [28]. When a cutpoint is not chosen a priori, there is a risk of bias because it may have been chosen to falsely demonstrate that a treatment performs better in a particular subgroup [25]. There is a rich literature on more complex approaches to choosing the cutpoint [29].

The benefits of directly assessing the interaction are that there will be a higher power to detect interactions and it avoids having to specify a cutpoint. Drawbacks include the model assumptions. For example, a regression model may assume a linear relationship between biomarker-level and treatment effect, whereas more complex biological relationships often exist [30]. Applying transformations or other methods for allowing nonlinear relationships, such as fractional polynomials [31], can address this, although may be difficult to interpret.

The primary aim of this paper is to gain a better understanding of current practice by investigating how often subgroup analyses in RCTs use continuous subgroup-defining variables and which methods are used to analyze them. This will contribute to the longer-term aim of promoting wider uptake and development of alternative methods for subgroup analyses in RCTs.

2. Materials and methods

To investigate how often subgroup analyses in RCTs use continuous variables, we searched PubMed on February 5, 2021, using the following search term:

(‘‘subgroup analysis’’ OR ‘‘sub-group analysis’’) AND ffrft [Filter] AND clinical trial [Filter] AND (‘‘2016’’ [Date - Publication]: ‘‘3,000’’ [Date - Publication]) AND English [Language] AND trial [Title/Abstract] AND (randomized [Title/Abstract] OR randomized [Title/Abstract] OR clinical [Title/Abstract])

This restricted results to free-full-text clinical trial articles from 2016 that contained ‘‘subgroup analysis’’ in their text and ‘‘randomized trial’’ in their title/abstract. This returned 410 articles for review. An additional search using the term ‘‘moderator analysis’’ instead of subgroup analysis returned a further 17 articles.

SFW and JW created an extraction spreadsheet with accompanying slides containing further information and illustrative examples, which were presented to the group to discuss ambiguities. Following this, a pilot extraction of 40 papers was conducted, where each reviewer was allocated the same ten papers as another reviewer. Discrepancies were discussed as a group and used to inform further refinements to the eligibility criteria and extraction spreadsheet; the final version is provided as Supplementary File 1. A second pilot stage was conducted, in which 50 further articles were double-reviewed. After this, the remaining papers were single-reviewed. Any papers that were unclear thereafter were flagged for independent assessment by SFW and JW.

For the purposes of this review, we restricted eligibility to papers that reported and/or compared the treatment effect within a subgroup and its *complementary* subgroup(s). For example, a study presenting treatment effects in Japanese participants and non-Japanese participants (such as [32]) would be eligible. However, a study where this was done in Japanese participants only and the overall trial population (such as [33]) would not be eligible. Only subgroups that were defined at baseline and applied to all arms of the trial were considered. Moreover, we only included papers that explicitly reported the methods and corresponding results. If a paper mentioned performing an interaction test but this was not presented in the results or supplementary material, it was excluded.

For each eligible paper, we proceeded with extraction of the relevant information, which included the following key questions:

1. What methods were used for subgroup analysis?
2. Were any subgroups based on continuous information?
3. If 2. was true, were they categorized?
4. If 3. was true, how was the cutpoint determined?

The complete list of extraction questions, with a brief description of each, is summarized in Supplementary Table S1 of the Supplementary Materials.

3. Results

A total of 427 papers were returned from the search; an overview of the main results is presented in Fig. 1. Once all completed reviews were merged and finalized, 147 (34.4%) papers had been independently reviewed by at least two people.

Of the 427 papers reviewed, 258 (60.4%) reported a subgroup analysis according to our definition. The remaining 169 papers were considered noneligible due to: only reporting results in a subgroup (with no analysis of treatment effect); not being a RCT (including six meta-analysis reports); and using post-baseline measurements to form the subgroup.

The characteristics of the 258 eligible RCTs are summarized in Table 1. In most cases (141/258, 54.7%), the phase of the trial was not stated or not applicable (e.g., due to a nondrug intervention) so was recorded as unclear. For the 117/258 (45.3%) papers where the phase was available (in the manuscript or corresponding ClinicalTrials.gov entry), the majority were phase three trials. In 90/258 (34.9%) papers, the subgroup analyses were pre-specified and 55/258 (21.3%) were post-hoc, with 14/258 (5.4%) containing both pre-specified and post-hoc subgroup

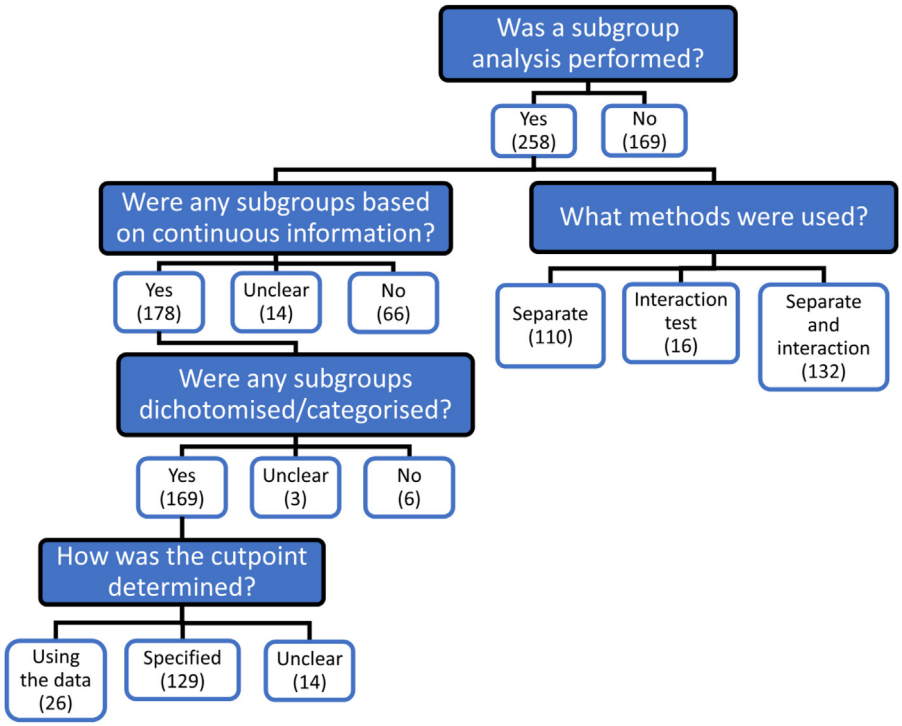

Fig. 1. Overview of review process. The values in parentheses represent the number of papers in the analysis.

analyses. However, many papers (38.4%) did not clearly state whether the subgroup analysis was pre-specified or post-hoc.

Table 1 also reports the type of method used for subgroup analyses. Just over half of the papers reported *both* separate analyses and an interaction test. However, a large proportion of papers (42.6%) reported *only* separate analyses in the subgroups. Interaction tests were exclusively reported in only 16/258 (6.2%) papers.

Table 2 summarizes how often continuous subgroup-defining variables were used and how they were analyzed (see also Fig. 1). Across the 258 eligible RCTs, 178 (69%) included at least one subgroup analysis where the subgroup was formed from a continuous variable (or combinations of continuous variables) and 14 were unclear. For example, in an oncology study [34] where the subgroups of interest were histological subtype and tumour grade, it was not obvious whether the criteria used to make these diagnoses were based on underlying continuous information. Similarly, in a RCT that classified patients with Alzheimer's disease [35] into mild and moderate, these subgroups were not explicitly defined. Although they are likely to be based on some continuous measurement(s), we cannot be sure from the details provided in the paper alone, and hence such cases were reported as 'unclear'. However, if the criteria used to form the subgroup was clearly reported in the text and included at least one continuous measurement, such as the National Heart, Lung and Blood Institute criteria used to classify asthma severity in [36], then the subgroup was recorded as being based on continuous information.

Amongst the 178 papers that had a subgroup based on continuous information, 169/178 (94.9%) dichotomized (or categorized) the variable and analysed using standard binary approaches (e.g., logistic regression). Out of these 169 RCTs, the majority (76.3%) appeared to use specified cutpoints that had been chosen a priori. In some cases, these were clearly stated as being standard for the clinical area. For example, when investigating outcomes in patients with metastatic castration-resistant prostate (e.g., [37]), it is routine to dichotomize at the age of 75. In a subgroup analysis of patients with high and low levels of pain [38], a cutpoint of 20 is used, in line with the pain catastrophising scale [39]. In cases where the choice of cutpoint was not explicitly stated, with no reference to standard practice or the baseline data, we assumed that these had also been specified (e.g., adopted from previous studies). In the 26/174 papers that used the data to determine the cutpoint, most of them (69.2%) used the sample median as the cutpoint. Other quantiles (including tertiles, quartiles, deciles) and the mean were also reported. Meyer et al. [40] used the Youden index [41] to select the cutpoint in plasma concentration that optimized the area under the mortality receiver operating characteristics curve. As well as dichotomizing based on the Youden index, the plasma biomarker concentration was divided into deciles to emulate testing it as a continuous variable and to ''maximize information content''. In some cases, multiple

Table 1. Summary of extracted data. The denominator for computing percentages (given to 1 decimal place) is stated in each heading

Eligible?	***n* (% out of 427)**
Yes	258 (60.4%)
No	169 (39.6%)
Phase of the trial?	*n* (% out of 258)
1/2	1 (0.4%)
2	25 (9.7%)
2b	3 (1.2%)
2/3	4 (1.6%)
3	65 (25.2%)
3/4	1 (0.4%)
4	18 (7%)
Unclear	141 (54.7%)
Was the subgroup analysis pre-specified or post-hoc?	*n* (% out of 258)
Pre-specified	90 (34.9%)
Post-hoc	55 (21.3%)
Both	14 (5.4%)
Unclear	99 (38.4%)
What methods were used for subgroup analysis?	*n* (% out of 258)
Separate	110 (42.6%)
Interaction test	16 (6.2%)
Separate and interaction	132 (51.2%)

cutpoints were used. For example, in Macarulla et al. [42], patients were stratified by age at cutpoints of 65, 70 and 75 to allow ''exploration of outcomes in older patients with greater sensitivity than the pre-specified subgroup analysis with a single cutpoint at age 65''. One RCT [43] used clustering of baseline covariates to form subgroups.

We found only ten examples (5.6%) where a continuous subgroup-defining variable was directly used in regression analysis to explore statistical interactions (six of which were returned from the ''moderator analysis'' search). These are highlighted in the completed extraction spreadsheet (Supplementary File 1). Of these, six used the continuous information directly and did *not* dichotomize; one

Table 2. Summary of how often subgroups were formed from continuous information and how this was accounted for in the analysis

Were any subgroups based on continuous information?	***n* (% out of 258)**
Yes	178 (69%)
No	66 (25.6%)
Unclear	14 (5.4%)
How was continuous information used in the analysis?	*n* (% out of 178)
Dichotomized only	166 (93.3%)
Used directly	6 (3.4%)
Used directly and dichotomized	3 (1.7%)
Used directly and unclear whether also dichotomized	1 (0.6%)
Unclear	2 (1.1%)
How was the cutpoint determined?	*n* (% out of 169)
Specified	129 (76.3%)
Using the data	26 (15.4%)
Unclear	14 (8.3%)

used the continuous information directly and it was *unclear* whether they dichotomized; and three did both.

We highlight the following examples as good practice of conducting subgroup analyses when the subgroup-defining variable is continuous. In Murray et al. [44], pre-specified subgroup analyses for baseline HbA1c, baseline score of the 'Problem Areas in Diabetes' (PAID) scale and duration of diabetes were undertaken, in which each variable was dichotomized for the separate analyses but treated directly as continuous in the interaction. The significance of the interaction between each continuous subgroup and the intervention was assessed using a Wald test. Although interaction *P*-values for the categorised subgroup variables weren't presented, duration of diabetes had a significant interaction with intervention on PAID without strongly inconsistent effect estimates in the dichotomized duration categories. Asche et al. [45] included 14 variables for subgroup analysis, of which body mass index, systolic blood pressure and diastolic blood pressure at baseline were treated as continuous variables in the treatment by subgroup interaction test. These variables were also categorized, with the treatment effect in each category summarized by odds ratios (and corresponding 95% confidence intervals). There was no disagreement in significance between the two approaches, although for diastolic blood pressure, the *P*-value for interaction was 0.01 with the continuous variable and 0.04 with the categorized variable. Krampe et al. [46] investigated how the 'pretreatment willingness to change' score was associated with the estimated intervention effect using a Johnson-Neyman plot. This illustrates when a significant intervention effect becomes non-significant and is an informative alternative to dichotomizing. Lastly, Ferris et al. [47] assessed several immune score variables both as continuous expression levels (with a square root transformation) and dichotomizing at ≤ 5 or > 5 percent positive cells. The results in their supplementary material indicate that no significant interaction effects were found with either method.

Other interesting examples that used Bayesian analysis and incorporated data from previous trials were also highlighted during the search, as has been previously recommended by Song and Bachmann [48]. In a follow-up trial [49], which aimed to confirm a previously reported subgroup analysis, a Bayesian repeated measures linear model was used for the primary endpoint analysis, with 30% of the subgroup information borrowed from the previous trial. Iglesias et al. [50] also used Bayesian methods for the primary and subgroup analyses, with priors constructed using data from each subgroup of patients in a previous trial.

4. Discussion

In this paper, we investigated the frequency of subgroup analyses that use continuous subgroup-defining variables in RCTs over a 5-year period (2016−2021). The results show that, although this is common, the statistical analysis usually ignores the continuous nature through categorizing and applying methods for binary subgroups. This was the case for nearly 95% of the papers reviewed. Consequently, many subgroup analyses in RCTs may be less efficient than they could be. We also found limited use of sophisticated methods for defining a cutpoint, despite the availability of such methods [29], and only a few examples where the rationale for dichotomization, and choice of cutpoint, was justified. Obtaining a better understanding of why dichotomization is the preferred approach in practice and the perceived barriers of using continuous information directly in subgroup analyses would help inform development of alternative methods.

Unlike previous reviews, which typically focus on a specific disease area, this review captures an extensive range of RCTs, covering a variety of disease areas, interventions and countries. The main difficulty we encountered was that the quality of reporting varied substantially. Several items, such as the type of analysis performed and whether the subgroup analysis was pre-specified or post-hoc, were poorly reported (consistent with previous reviews [8]). However, through two pilot phases, regular discussions between reviewers to resolve discrepancies and over a third of the papers being double (or triple) reviewed, we improved the consistency and accuracy of our extractions. By restricting our search to freely available papers that mentioned ''subgroup'' or ''moderator analysis'', our review is not exhaustive. The focus of this review is on the subgroup analysis methods used in RCTs; further work could explore the proportion of these that reported significant subgroup effects.

In conclusion, we have found that it is very common for subgroup analyses in RCTs to categorize continuous subgroup-defining variables. Ideally, we would recommend that at least one reported analysis keeps them continuous, which was only done in 10/178 of the papers reviewed. Further investigation into the rationale for categorization, together with the development and dissemination of better methods, would facilitate more informative subgroup analyses and increased efficiency.

Appendix A

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2022.06.017.

References

[1] Drilon A, Laetsch TW, Kummar S, Dubois SG, Lassen UN, Demetri GD, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. N Engl J Med 2018;378:731−9.

[2] Kahraman S, Yalcin S. Recent advances in systemic treatments for HER-2 positive advanced gastric cancer. Onco Targets Ther 2021; 14:4149−62.

[3] Tom J, Bao M, Tsai L, Qamra A, Summers D, Carrasco-Triguero M, et al. Prognostic and predictive biomarkers in patients with coronavirus disease 2019 treated with tocilizumab in a randomized controlled trial. Crit Care Med 2021;50:398−409.
[4] Matveychuk D, Thomas RK, Swainson J, Khullar A, MacKay M-A, Baker GB, et al. Ketamine as an antidepressant: overview of its mechanisms of action and potential predictive biomarkers. Ther Adv Psychopharmacol 2020;10. 2045125320916657.
[5] Dzau VJ, Ginsburg GS. Realizing the full potential of precision medicine in health and health care. JAMA 2016;316:1659−60.
[6] Tanniou J, van der Tweel I, Teerenstra S, Roes KCB. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. BMC Med Res Methodol 2016;16:20.
[7] Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005;365: 176−86.
[8] Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine — reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189−94.
[9] 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.
[10] Fan J, Song F, Bachmann MO. Justification and reporting of subgroup analyses were lacking or inadequate in randomized controlled trials. J Clin Epidemiol 2019;108:17−25.
[11] Wijn SRW, Rovers MM, Le LH, Belias M, Hoogland J, IntHout J, et al. Guidance from key organisations on exploring, confirming and interpreting subgroup effects of medical treatments: a scoping review. BMJ Open 2019;9:e028751.
[12] Dmitrienko A, Muysers C, Fritsch A, Lipkovich I. General guidance on exploratory and confirmatory subgroup analysis in late-stage clinical trials 2016. J Biopharm Stat 2016;26:71−98.
[13] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials 2010;11:1−8.
[14] Sun X, Ioannidis JPA, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup Analysis: users' guide to the medical literature. JAMA 2014;311:405−11.
[15] Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. BMJ 2015;351:h5651.
[16] Committee for Medicinal Products for Human Use. Guideline on the investigation of subgroups in confirmatory clinical trials. London: Eur Med Agency; 2014.
[17] Temple R. Encouragement of subgroup assessment by the FDA. Pharm Stat 2021;20:923−8.
[18] Egle A, Melchardt T, Obrtlíková P, Smolej L, Kozák T, Steurer M, et al. Rituximab maintenance overcomes the negative prognostic factor of obesity in CLL: subgroup analysis of the international randomized AGMT CLL-8a mabtenance trial. Cancer Med 2019;8:1401−5.
[19] Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. Am J Psychiatry 2001;158: 848−56.
[20] Hunter DJ. Uncertainty in the era of precision medicine. N Engl J Med 2016;375:711−3.
[21] Shentu Y, Xie M. A note on dichotomization of continuous response variable in the presence of contamination and model misspecification. Stat Med 2010;29:2200−14.
[22] MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods 2002; 7:19−40.
[23] Farewell VT, Tom BDM, Royston P. The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. J Am Stat Assoc 2004;99:822−31.
[24] Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. Stat Med 2004;23:1159−78.
[25] 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.
[26] 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.
[27] Buettner P, Garbe C, Guggenmoos-Holzmann I. Problems in defining cutoff points of continuous prognostic factors: example of tumor thickness in primary cutaneous melanoma. J Clin Epidemiol 1997; 50:1201−10.
[28] Weinberg CR. How bad is categorization? Epidemiology 1995;6: 345−7.
[29] Lipkovich I, Dmitrienko A, D'Agostino RB. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med 2017;36:136−96.
[30] Calabrese EJ, Baldwin LA. U-shaped dose-responses in biology, toxicology, and public health. Annu Rev Public Health 2001;22:15−33.
[31] Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Stat Med 2004;23:2509−25.
[32] Oya M, Tomita Y, Fukasawa S, Shinohara N, Habuchi T, Rini BI, et al. Overall survival of first-line axitinib in metastatic renal cell carcinoma: Japanese subgroup analysis from phase II study. Cancer Sci 2017;108:1231−9.
[33] Ohta K, Adachi M, Tohda Y, Kamei T, Kato M, Mark Fitzgerald J, et al. Efficacy and safety of benralizumab in Japanese patients with severe, uncontrolled eosinophilic asthma. Allergol Int 2018;67: 266−72.
[34] Young RJ, Litière S, Lia M, Hogendoorn PCW, Fisher C, Mechtersheimer G, et al. Predictive and prognostic factors associated with soft tissue sarcoma response to chemotherapy: a subgroup analysis of the European Organisation for Research and Treatment of Cancer 62012 study. Acta Oncol 2017;56:1013−20.
[35] Zhao J, Li Z, Cong Y, Zhang J, Tan M, Zhang H, et al. Repetitive transcranial magnetic stimulation improves cognitive function of Alzheimer's disease patients. Oncotarget 2017;8:33864−71.
[36] Goldstein NPN, Frey SM, Fagnano M, Okelo SO, Halterman JS. Identifying which urban children with asthma benefit most from clinician prompting: subgroup analyses from the prompting asthma intervention in rochester−uniting parents and providers (PAIR-UP) trial. Acad Pediatr 2018;18:305−9.
[37] Graff JN, Baciarello G, Armstrong AJ, Higano CS, Iversen P, Flaig TW, et al. Efficacy and safety of enzalutamide in patients 75 years or older with chemotherapy-naive metastatic castration-resistant prostate cancer: results from PREVAIL. Ann Oncol 2016; 27:286−94.
[38] Losina E, Collins JE, Wright J, Daigle ME, Donnell-Fink LA, Strnad D, et al. Postoperative care navigation for total knee arthroplasty patients: a randomized controlled trial: RCT of postoperative care navigation in TKA. Arthritis Care Res 2016;68:1252−9.
[39] Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7:524−32.
[40] Meyer NJ, Reilly JP, Anderson BJ, Palakshappa JA, Jones TK, Dunn TG, et al. Mortality benefit of recombinant human interleukin-1 receptor antagonist for sepsis varies by initial interleukin-1 receptor antagonist plasma concentration. Crit Care Med 2018;46:21−8.
[41] Fluss R, Faraggi D, Reiser B. Estimation of the youden index and its associated cutoff point. Biom J 2005;47:458−72.
[42] Macarulla T, Blanc J-F, Wang-Gillam A, Chen L-T, Siveke JT, Mirakhur B, et al. Liposomal irinotecan and 5-fluorouracil/ leucovorin in older patients with metastatic pancreatic cancer − a subgroup analysis of the pivotal NAPOLI-1 trial. J Geriatr Oncol 2019;10: 427−35.
[43] Penza CW, Horn ME, George SZ, Bishop MD. Comparison of 2 lumbar manual therapies on temporal summation of pain in healthy volunteers. J Pain 2017;18:1397−408.

[44] Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017;7:e016009.

[45] Asche SE, O'Connor PJ, Dehmer SP, Green BB, Bergdall AR, Maciosek MV, et al. Patient characteristics associated with greater blood pressure control in a randomized trial of home blood pressure telemonitoring and pharmacist management. J Am Soc Hypertens 2016;10:873−80.

[46] Krampe H, Salz AL, Kerper LF, Krannich A, Schnell T, Wernecke KD, et al. Readiness to change and therapy outcomes of an innovative psychotherapy program for surgical patients: results from a randomized controlled trial. BMC Psychiatry 2017;17:417.

[47] Ferris RL, Saba NF, Gitlitz BJ, Haddad R, Sukari A, Neupane P, et al. Effect of adding motolimod to standard combination chemotherapy and cetuximab treatment of patients with squamous cell carcinoma of the head and neck: the Active8 randomized clinical trial. JAMA Oncol 2018;4:1583−8.

[48] Song F, Bachmann MO. Cumulative subgroup analysis to reduce waste in clinical research for individualised medicine. BMC Med 2016;14:197.

[49] Abraham WT, Kuck KH, Goldsmith RL, Lindenfeld J, Reddy VY, Carson PE, et al. A randomized controlled trial to evaluate the safety and efficacy of cardiac contractility modulation. JACC Heart Fail 2018;6:874−83.

[50] Iglesias JF, Muller O, Heg D, Roffi M, Kurz DJ, Moarof I, et al. Biodegradable polymer sirolimus-eluting stents versus durable polymer everolimus-eluting stents in patients with ST-segment elevation myocardial infarction (BIOSTEMI): a single-blind, prospective, randomised superiority trial. Lancet 2019;394:1243−53.