Review

CLINICAL TRIALS

Clinical Trials
2021, Vol. 18(3) 351–360

DOI: 10.1177/1740774520984866
journals.sagepub.com/home/ctj
SAGE

A systematic review of subgroup analyses in randomised clinical trials in cardiovascular disease

Korbinian J Brand[1*], Alexander Hapfelmeier[1,2*] and Bernhard Haller[1]

Abstract

Background: Subgroup analyses are frequently used to assess heterogeneity of treatment effects in randomised clinical trials. Inconsistent, improper and incomplete implementation, reporting and interpretation have been identified as ongoing challenges. Further, subgroup analyses were frequently criticised because of unreliable or potentially misleading results. More recently, recommendations and guidelines have been provided to improve the reporting of data in this regard.

Methods: This systematic review was based on a literature search within the digital archives of three selected medical journals, *The New England Journal of Medicine, The Lancet* and *Circulation*. We reviewed articles of randomised clinical trials in the domain of cardiovascular disease which were published in 2015 and 2016. We screened and evaluated the selected articles for the mode of implementation and reporting of subgroup analyses.

Results: We were able to identify a total of 130 eligible publications of randomised clinical trials. In 89/130 (68%) articles, results of at least one subgroup analysis were presented. This was dependent on the considered journal ($p < 0.001$), the number of included patients ($p < 0.001$) and the lack of statistical significance of a trial's primary analysis ($p < 0.001$). The number of reported subgroup analyses ranged from 1 to 101 (median = 13). We were able to comprehend the specification time of reported subgroup analyses for 71/89 (80%) articles, with 55/89 (62%) articles presenting exclusively pre-specified analyses. This information was not always traceable on the basis of provided trial protocols and often did not include the pre-definition of cut-off values for the categorization of subgroups. The use of interaction tests was reported in 84/89 (94%) articles, with 36/89 (40%) articles reporting heterogeneity of the treatment effect for at least one primary or secondary trial outcome. Subgroup analyses were reported more frequently for larger randomised clinical trials, and if primary analyses did not reach statistical significance. Information about the implementation of subgroup analyses was reported most consistently for articles from *The New England Journal of Medicine*, since it was also traceable on the basis of provided trial protocols. We were able to comprehend whether subgroup analyses were pre-specified in a majority of the reviewed publications. Even though results of multiple subgroup analyses were reported for most published trials, a corresponding adjustment for multiple testing was rarely considered.

Conclusion: Compared to previous reviews in this context, we observed improvements in the reporting of subgroup analyses of cardiovascular randomised clinical trials. Nonetheless, critical shortcomings, such as inconsistent reporting of the implementation and insufficient pre-specification, persist.

Keywords

Subgroup analyses, treatment effect heterogeneity, randomized trials, reporting, systematic review

Introduction

Background

Randomised clinical trials (RCTs) in cardiovascular disease often include subgroup analyses.[1–3] These are used to assess heterogeneity of treatment effects, concerning primary, secondary or adverse trial outcomes.[4–6] Corresponding investigations are generally based on the assumption that certain subgroups of patients may benefit more or less from a studied intervention.[5,7] Subgroup

[1]Institute of Medical Informatics, Statistics and Epidemiology, School of Medicine, Technical University of Munich, Munich, Germany
[2]Institute of General Practice and Health Services Research, School of Medicine, Technical University of Munich, Munich, Germany

[*]K.J.B. and A.H. contributed equally

Corresponding author:
Bernhard Haller, Institute of Medical Informatics, Statistics and Epidemiology, School of Medicine, Technical University of Munich, Klinikum rechts der Isar, Ismaninger Str. 22, Munich 81675, Germany.
Email: bernhard.haller@tum.de

analyses are particularly useful when patient characteristics are associated with treatment effects and to define patients with increased risk profiles.[5,8,9] Subgroup analyses represent a valuable source of information, but implementation in (R)CTs may lead to challenges as well, such as the pre-specification of relevant patient characteristics and respective cut-off values for the definition of subgroups.[10–12]

Previous knowledge

Subgroup analyses were often criticised for spurious, meaningless or potentially misleading results.[2,5,13–15] Previous reviews showed that reported information regarding the implementation of subgroup analyses in (R)CTs was not consistent or complete.[1,2,5,14,16–18] Further, performing numerous subgroup analyses may lead to a multiple testing problem and therefore a higher chance for false-positive findings.[10,19–22] Reported (R)CTs did often not include the recommended test of interaction between treatment and the subgroup defining variables.[2,5,18,21] Recommendations and clear guidelines for the implementation and publication of subgroup analyses in the context of RCTs are available, which aim at increasing the comparability, generalizability and error control of results.[23–27] As major outcomes of previous reviews, it was shown that subgroup analyses have been published more frequently for large (R)CTs and in the case of non-significant primary analyses.[1,2,5,16,17] Results of at least one subgroup analysis were reported from 61% to 70% of reviewed (R)CTs,[2,5,14,21] with a median of up to four reported subgroup analyses per trial.[14,15,21] Information about the specification time of published subgroup analyses was available for 32%–41% published (R)CTs,[5,16,21] with 28%–46% of articles reporting results from corresponding interaction tests.[2,5,18,21] In contrast to this current report, not all summarised results of previous reviews did specifically refer to cardiovascular RCTs.

Objective

This systematic review is based on a literature search that was conducted within the digital archives of three high-impact medical journals, covering articles published between 2015 and 2016. We selected relevant articles in the domain of cardiovascular RCTs and compared the implementation and reporting of subgroup analyses between the journals *The New England Journal of Medicine* (*NEJM*), *The Lancet* and *Circulation*. Relying on reported data, we examined the relation between the frequency of reported subgroup analyses and various trial characteristics, such as the statistical significance of the primary analysis, the type of intervention under study and the number of included patients. This was followed by an assessment of the pre-specification of subgroups, the use of interaction tests, the presence of significant results and the number of reported subgroup analyses per article. Based on a comparison to previous reviews,[1,2,5,14,16,18,21] our investigation aimed to detect trends and improvements regarding the implementation and reporting of subgroup analyses within the last two decades which could be attributable to official recommendations and guidelines for subgroup analyses, which have been published in between.

Methods

Information sources

We investigated official guidelines and recommendations for the implementation and reporting of subgroup analyses in (R)CTs. Reference applies to the *European Medicines Agency*, the *International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use* and the *Consolidated Standards of Reporting Trials (CONSORT) Statement*.[25–27] Relevant information about the implementation of subgroup analyses should always be indicated clearly, such as the time of specification or the investigated outcome.[25–27] Ideally, this should also be traceable on the basis of provided trial protocols.[17,25–27] Subgroup analyses should be pre-specified whenever possible, and it should be clearly marked if this was not the case.[25,27] In the case of confirmatory analyses, pre-specification is a mandatory requirement and should, for continuous variables, also include a determination of cut-off values for the definition of subgroups.[25,27] Further, interaction tests are recommended to assess heterogeneous treatment effects.[4,25,27] Potential susceptibilities to false-positive findings because of multiple testing should be considered.[25,27] Even though results were often shown to be questionable, not investigating subgroup analyses might also cause misleading therapeutic recommendations.[25] Because of this, further efforts should be undertaken to improve the implementation and reporting of subgroup analyses within (R)CTs. In general, it can be assumed that working in accordance with the mentioned guidelines also contributes to a qualitative improvement in this regard.[5,23,24]

In addition, author guidelines for selected journals were examined in regard to explicit requirements or instructions for the presentation of subgroup analyses (see journal websites for references). Providing information about the implementation of subgroup analyses in (R)CTs, such as the pre-specification or methods used for hypothesis testing, should be seen as a mandatory requirement for authors according to these guidelines.

In preparation of the present article, we followed the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* statement where appropriate, which provide a summary of requirements and recommendations for systematic reviews.[28] The *Clinical*

Trials (CT) and *EudraCT* registration (www.clinicaltrials.gov, www.clinicaltrialsregister.eu) of the reviewed RCTs were checked to acquire more information about planned subgroup analyses, but these registries did not provide any relevant information in this regard. Finally, we explored results from previous systematic reviews of subgroup analyses in (R)CTs and compared our findings to these.[1,2,5,14,16,18,21]

Search within journal archives

We conducted a systematic literature search within the digital archives of three selected medical journals, *NEJM, The Lancet* and *Circulation* (initial search date: October 2017). The journals' online tools 'Advanced Search' were used to identify 'original research' articles with reference to cardiovascular RCTs, published during the years 2015 and 2016. We searched the whole accessible content and used the term 'random*', to cover the search terms 'randomly', 'randomised', 'randomized' and similar, and to narrow down numbers of indicated results. In addition, we chose the option 'filter by article category' to limit our research to published articles of clinical trials only. Finally, our research was restricted to articles from original journals only, so that no publications from subtitle journals were taken in account.

Eligibility criteria and article selection

We made a thematic reference to issues related to human cardiovascular and circulatory disease according to chapter IX of the *International Classification of Diseases and Related Health Problems (ICD-10)* of the *World Health Organisation*.[29] Correspondingly, all identified full-text articles on clinical trials were searched and checked for eligibility. It was also ensured that no publications from follow-up or post hoc analyses of studies that have been published before were included. If a selected article relied on data from more than one RCT, we still considered this as a single case during the process of our analysis. If several articles referred to the same RCT, we also considered them as one case.

Data collection and synthesis of results

The full texts of all selected articles about RCTs were screened for at least one reported subgroup analysis. Any comparison of treatment groups with regard to defined trial outcomes, such as primary efficacy endpoints, secondary endpoints or safety endpoints with stratification of patients according to baseline characteristics, and the optional use of an interaction test were considered as a subgroup analysis even if not described as such. We also screened the trial protocols, if available, and data supplements for more information.

Based on the collected data, we examined frequencies of reported subgroup analyses, the size of trials, the statistical significance of the trials' analyses of the primary outcome, the type of hypothesis (superiority or non-inferiority) and the kind of evaluated therapeutic intervention.

We also checked whether information about the time of specification was provided. If so, we continued to distinguish between publications that present results from pre-specified subgroup analyses only, contrary to results from subgroup analyses with unclear, inconsistent or post hoc specification. We concluded a pre-specification of the subgroup analyses if they were stated in the respective trial protocols, analysis plans or if described as such in the provided full-text article. The number of reported subgroup analyses per publication was counted as the number of characteristics that were used for subgroup definition or a multiple thereof if subgroup analyses were performed for the evaluation of various outcomes or more than two treatment groups. For all selected cases, we tried to comprehend from the description of the analysis methods or from the presentation if results were adjusted for multiple testing, whether interaction tests were used and whether results achieved statistical significance.

Each article was assessed independently by two raters, and discrepancies were solved by a consensus discussion. Data were collected in tabular form, and statistical analysis was carried out using R (The R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistics, version 24.0 (IBM Corporation, Armonk, New York, USA). Categorical data are presented by absolute and relative frequencies. Continuous data are summarized by mean, median, minimum, maximum and first and third quartile. The relation of the outcome 'reporting of at least one subgroup analysis in an article' to categorical variables was assessed by χ^2 tests. A multivariable logistic regression model was used to explore the effects of multiple variables simultaneously, including the number of subjects, the significance of primary trial results and the featuring journal. To further assess the association between the number of performed subgroup analyses and number of included patients in a trial in articles with at least one reported subgroup analysis, Spearman's rank correlation coefficient was computed. All statistical tests were performed two-sided at a significance level of $\alpha = 5\%$. Performed analyses were conducted in an exploratory manner, and consequently, p values were not adjusted for multiple testing.

Results

Reviewed articles and frequency of subgroup analyses

During the literature search, a total of 1462 records from the journals' digital archives were screened, with 671 records referring to publications of (R)CTs, and 175 eligible articles from cardiovascular RCTs. We excluded 43 articles, as they were based on follow-up or post hoc analyses from RCTs. Three articles published data from the same trial and were considered as one.[30–32] Thus, we explored a total of 130 selected original articles from cardiovascular RCTs (Figure 1, Table 1).

At least one subgroup analysis was reported in 89/130 (68%) articles, with 59/69 (86%) from *NEJM*, 17/28 (61%) from *The Lancet* and 13/33 (39%) from *Circulation* ($p < 0.001$, χ^2 test, Table 1). The likelihood of reporting subgroup analyses was also dependent on the number of patients included in the primary analyses ($p < 0.001$, χ^2 test, Table 1), which ranged from 14 to 24,081 patients with a median of 1136 patients (mean of 2787, Table 2), while larger trials have more often been published by *NEJM* (Table 2). Only 31/89 (31%) included articles presented at least one subgroup analysis with results about a secondary trial outcome.

In total, 80/130 (62%) articles reported significant results from primary trial analyses. These included less often subgroup analyses (46/80 = 58%) than articles from trials with a respective non-significant result (43/50 = 86%, $p < 0.001$, χ^2 test, Table 1).

There was no relevant difference regarding the frequency of reporting subgroup analyses, when comparing selected articles of superiority and non-inferiority trials (Table 1). Also, we found no relevant difference between articles of trials for the evaluation of pharmaceutical, surgical, endovascular or remaining interventions under study (Table 1).

We used multivariable logistic regression for a simultaneous analysis of considered factors and to adjust the analysis for the apparent relation between the journals and the size of the trials. Results showed that the likelihood of reporting subgroup analyses increased with the number of patients (odds ratio (OR) = 1.41 per 500 pts, 95% confidence interval (CI) 1.11–1.77, p = 0.004), with the lack of significance of the primary trial analysis (OR = 4.42, 95% CI 1.55–12.6, p = 0.005) and with the featuring journal (p = 0.020, *NEJM* versus *Circulation*: OR = 4.76, 95% CI 1.57–14.4; *The Lancet* versus *Circulation*: OR = 1.83, 95% CI 0.56–6.01).

Number of reported subgroup analyses

If subgroup analyses were reported for a trial, the number ranged from 1 to 101 with a median of 13 and a mean of 17 (Table 3). This was dependent on the featuring journal, with a mean of 20 for *NEJM*, 16 for *The Lancet* and 8 for *Circulation* (Table 3). Referring to size, more subgroup analyses were reported for larger trials (Spearman correlation: r = 0.41, 95% CI 0.24–0.59, $p < 0.001$). The multiple testing problem was addressed in only two publications.[33,34]

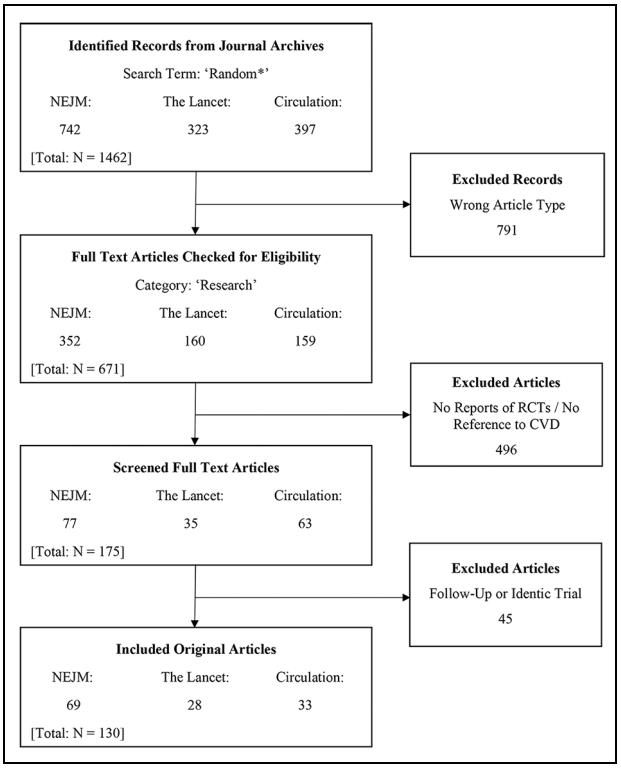

Figure 1. Flow chart – literature search and article selection.
CVD: cardiovascular disease; *NEJM: The New England Journal of Medicine*; RCTs: randomised clinical trials.
Research within journal archives and article selection for the covered date range (January 2015–December 2016). [N] Indicates overall numbers of eligible search results according to defined search criteria. Smaller boxes aside indicate numbers of excluded cases. The number of corresponding records was counted at the date of our literature search (results from October 2017 are shown).

Specification of subgroup analyses

Overall, a total of 55/89 (62%) reviewed articles presented results of exclusively pre-specified subgroup analyses. This included 42/59 (71%) articles from *NEJM*, 8/17 (47%) articles from *The Lancet* and 5/13 (38%) articles from *Circulation*. Further, 14/89 (16%) articles reported results from both, a priori and post hoc defined subgroup analyses. Information about the pre-specification of these subgroup analyses was not traceable in the trials' online registration (*CT* and *EudraCT*). A small amount of 2/89 (2%) articles reported results of only post hoc specified subgroup

Table 1. Articles reporting at least one subgroup analysis, n (%).

	NEJM	*The Lancet*	*Circulation*	Total	p value
Year of publication					
2015	33/42 (78)	7/15 (47)	7/17 (41)	47/74 (64)	0.163
2016	26/27 (96)	10/13 (77)	6/16 (38)	42/56 (75)	
Subjects (n)					
⩽259	4/8 (50)	3/7 (43)	2/17 (12)	9/32 (28)	<0.001
260–1136	13/16 (81)	4/9 (44)	5/8 (63)	22/33 (67)	
1137–2890	19/21 (90)	5/6 (83)	4/6 (67)	28/33 (85)	
⩾2891	23/24 (96)	5/6 (83)	2/2 (100)	30/32 (94)	
Trial design					
Superiority	48/55 (87)	11/20 (55)	12/29 (41)	71/104 (68)	0.925
Non-inferiority	11/14 (78)	6/8 (75)	1/4 (25)	18/26 (69)	
Primary analysis					
Significant	28/37 (76)	11/20 (55)	7/23 (30)	46/80 (58)	<0.001
Not significant	31/32 (97)	6/8 (75)	6/10 (60)	43/50 (86)	
Intervention type					
Pharmaceutical	32/36 (89)	8/14 (57)	4/14 (29)	44/64 (69)	0.675
Surgical	6/7 (86)	0/2 (0)	0/0 (-)	6/9 (67)	
Endovascular	13/16 (81)	7/8 (88)	2/5 (40)	22/29 (76)	
Others	8/10 (80)	2/4 (50)	7/14 (50)	17/28 (61)	
Total	59/69 (86)	17/28 (61)	13/33 (39)	89/130 (68)	

NEJM: The New England Journal of Medicine.

Comparison of trial characteristics and likelihood of reporting subgroup analyses from cardiovascular randomised trials. χ^2 tests were used to test for an association between the trial characteristics and frequency of at least one subgroup analysis. No multiplicity adjustment was considered.

Table 2. Patients included for statistical analyses per trial (n).

Journal	Report of SGA	Articles (N)	Mean	Min.	1st quartile	Median	3rd quartile	Max.
NEJM	*Yes*:	59	4722	110	970	2032	7020	24,081
	No:	10	912	14	93	278	1561	4465
	Total:	69	4170	14	616	1905	5361	24,081
The Lancet	*Yes*:	17	2109	168	454	1215	3116	8404
	No:	11	829	47	109	399	501	4146
	Total:	28	1606	47	261	564	2578	8404
Circulation	*Yes*:	13	1753	60	332	617	1729	7402
	No:	20	340	22	119	203	290	2291
	Total:	33	897	22	151	253	908	7402
Overall	*Yes*:	89	3789	60	622	1905	4265	24,081
	No:	41	611	14	107	246	438	4465
	Total:	130	2787	14	260	1136	2890	24,081

SGA: subgroup analysis; *NEJM: The New England Journal of Medicine*; Min.: minimum, Max.: maximum.

Journal comparison: The number of patients that was included in the primary analysis of cardiovascular randomised trials.

analyses. It was not possible to determine the specification time of reported subgroup analyses for a total of 18/89 (20%) articles.

Considering articles from the journal *NEJM* only, trial protocols of 48/59 (81%) articles contained information regarding the pre-specification of reported subgroup analyses. However, this did not always include a pre-definition of cut-off values used for the categorization of subgroups according to continuously scaled characteristics. Trial protocols were not provided by the other two journals.

A relevant number of the respective articles reported results from subgroup analyses with regard to quantitative variables, such as patients' age (66/89; 74%), body mass index (19/89; 21%) or estimated glomerular filtration rate (17/89; 19%). Based on the body mass index or estimated glomerular filtration rate, subgroups were most often defined by clinically established cut-off values. For the estimated glomerular filtration rate, a cut-off value of 60 mL/min was used in 14/17 (82%) analyses, and in seven studies, multiple cut-off values were considered. Categories defined as <30 kg/m^2 versus >30 kg/m^2 were used for the body mass index in 12/19 (63%) analyses. More heterogeneous cut-off values were used for the categorisation of subgroups according to patients' age, where in 29/66

Table 3. Number of reported subgroup analyses (n).

	Articles (N)	Mean	Min.	1st quartile	Median	3rd quartile	Max.
Journal							
NEJM	59/69	20	3	10	14	24	101
The Lancet	17/28	16	1	5	11	27	46
Circulation	13/33	8	1	2	7	11	21
Subjects (No.)							
⩽269	9/32	10	2	3	8	16	20
260–1136	22/33	11	1	7	10	15	30
1137–2890	28/33	18	1	7	13	24	84
⩾2891	30/32	24	2	11	17	32	101
Primary trial result							
Significant	46/80	17	1	7	13	22	84
Not significant	43/50	18	2	7	13	24	101
Total	89/130	17	1	7	13	22	101

NEJM: The New England Journal of Medicine; Min.: minimum, Max.: maximum.
The number of published subgroup analyses in articles of cardiovascular randomised trials with the report of at least one subgroup analysis.

(44%) trials, an age of 65 years was used, a cut-off value of 75 years in 16/66 (24%), of 60 years in 9/66 (14%) and of 70 years in 8/66 (12%) trials. For 5/66 (8%) trials, a median split for subgroup division was reported. More than two age subgroups were defined in 14/66 (21%) trials. An overview of defined cut-off values for the categorisation of subgroups according to these quantitative variables can be found in Figure S1 (Supplementary Appendix).

Testing for interaction and heterogenous treatment effects

The use of a test for interaction or treatment effect heterogeneity was reported in 84/89 (94%) articles. This included 59/59 (100%) articles from *NEJM*, 15/17 (88%) articles from *The Lancet* and 10/13 (77%) articles from *Circulation*.

Significant heterogeneity of treatment effects for at least one primary, secondary or safety endpoint was described for a total of 36/89 (40%) articles. This refers to a total of 26/59 (44%) articles published by *NEJM*, 6/17 (35%) articles published by *The Lancet* and 4/13 (31%) articles published by *Circulation*.

Discussion

Factors related to subgroup analyses

A majority of the selected original articles of RCTs in cardiovascular disease reported data of subgroup analyses (68%), which corresponds to demands of provided author guidelines for the three journals at the time of analysis. Previous reviews have shown that subgroup analyses were carried out more frequently in clinical trials with larger patient collectives.[2,5,16] We were able to reproduce this finding for the publications of RCTs from the years 2015 and 2016 ($p < 0.001$). Beside this, subgroup analyses seemed to be reported more frequently if the primary analysis did not reach statistical significance ($p < 0.001$), raising the question to what extent significant results were sought within the scope of subgroup analyses. Barraclough and Govindan described a possible 'fishing trip' for positive results in the context of subgroup analyses to value results of overall non-significant trials.[35] Similar findings were described by Sun et al., who conducted a review about subgroup analyses in a total of 469 RCTs published in 2007.[1,21,36] They found that subgroup analyses were reported more frequently if industrial funded RCTs achieved non-significant results for the primary outcome.[1,21] According to Gabler et al.,[18] this also applied to 416 articles of RCTs that were published in the years 2007, 2010 and 2013.[21]

We were also able to confirm these findings with the applied logistic regression model, which was primarily used to assess the independent effects of the examined variables with an apparent relation to the likelihood of reporting subgroup analyses in cardiovascular RCT(s). Further, it was of particular interest to adjust for possible confounding introduced by these variables with regard to a comparison of the investigated journals.

Subgroup analyses and multiple testing

When comparing our findings to previous reviews, a slight increase in the frequency of subgroup analyses may be worth mentioning. For example, Wang et al. were able to identify the use of subgroup analyses in a total of 59/97 (61%) clinical trials, published by *NEJM* from July 2005 to June 2006.[5] In addition, the authors Hernandez et al. found that 39/63 (62%) articles presented results of subgroup analyses, when considering reports of cardiovascular RCTs published by a total of eight selected journals in 2002 and 2004.[2] Assmann et al.[14] compared articles for the publication of clinical

trials from the journals *NEJM, Journal of the American Medical Association* and *The Lancet* from July to September 1997. At least one subgroup analysis was reported in 35/50 (70%) of these articles, which is comparable to our overall finding of 89/130 (68%).

Articles published by the journal *Circulation* included fewer subgroup analyses per case than articles published by *NEJM* and *The Lancet*. This may be due to the inclusion of trials with smaller patient numbers, whereby subgroup analyses are known to require larger numbers of patients for a useful evaluation of treatment effects.[37,38] Consistent therewith, the number of reported subgroup analyses from RCT(s) was dependent on the number of included patients, as publications of larger trials presented results of more subgroup analyses (Spearman correlation: r = 0.41, 95% CI 0.24–0.59).

In comparison with results from previous reviews, the number of reported subgroup analyses per article tended to be much higher. To be specific, Wang et al. showed that only 17/59 (29%) of the included articles reported results of more than eight subgroup analyses,[5] compared to 59/89 (66%) articles considered for the present review. Hernandez et al. found that 26/39 (67%) articles reported data of more than five subgroup analyses,[2] while this was the case in 79/89 (88%) articles considered for this review. Based on results from Assmann et al., the number of reported subgroup analyses from 35 clinical trials in 1997 ranged from 1 to 24 with a median of 4,[14] compared to a range from 1 to 101 and a median of 13 in the present review.

Susceptibility to false-positive findings from subgroup analyses has been pointed out several times in the scientific literature.[5,19,22,35] Whenever a large number of subgroup analyses are performed, the probability of a false-positive finding increases relevantly beyond the nominal level.[5,27,35] The risk of false-positive conclusions about heterogeneous treatment effects increases even more when separate analyses for pairwise comparisons of treatment effects are carried out without considering multiplicity issues. Accordingly, results of subgroup analyses should always be interpreted with caution, especially as adjustment for multiple comparisons is rarely made.[5,10,39] Only two trials were found to correct results of subgroup analyses for the multiple testing problem in the present review.

Specification of subgroup analyses

In summary, we were able to clearly distinguish between the reporting of a priori or post hoc specified subgroup analyses in most reviewed cases (80%), while the correctness of information given in the articles had to be assumed for trials with no published protocol. A majority of the articles presented results of pre-specified subgroup analyses only (62%), or in combination with results from post hoc analyses (16%). However, the specification time of reported subgroup analyses remained unclear in a considerably large number or reviewed publications (20%).

Nonetheless, authors of the reviewed articles generally seemed to put greater emphasis on reporting information about the specification time of presented subgroup analyses, than seen in findings from previous reviews. This may relate to the increasing number of official guidelines and recommendations for the conduct of subgroup analyses in clinical trials.[21,25,27] Wang et al.[5] showed that the specification time of reported subgroup analyses could be reproduced only for 19/59 (32%) reviewed articles of clinical trials published during 2005 and 2006, compared to 71/89 (80%) in the present study. According to Moreira et al.,[16] 7/17 (41%) articles of clinical trials from the year 1998 contained information about whether reported subgroup analyses were planned a priori or post hoc. This was based on a comparison of four scientific journals, *NEJM, Journal of the American Medical Association, American Journal of Public Health* and *The Lancet*.[16] For all mentioned reviews, collected data were solely based on information that was reported within the reviewed articles.

In general, subgroup analyses should be pre-specified in trial protocols.[17,25–27] These were available exclusively for articles published by *NEJM*, with most protocols actually including a pre-definition of published subgroup analyses (81%). Viewed in more detail, this did not always include a pre-definition of cut-off values for the categorisation of subgroups according to continuously scaled characteristics. Based on official recommendations, a priori definition of subgroup analyses should always include examined subgroup characteristics and cut-off values which should be reasoned with clinical evidence or concrete assumptions.[10,17,25,27,40] A post hoc variation of cut-off values may produce biased results and increase the probability of false-positive findings.[12,35]

Although investigated subgroups were often stratified according to the same characteristics in the reviewed clinical trials, comparability of results has been limited by the use of different cut-off values for the definition of subgroups. We observed a more uniform definition of cut-off values to be established for some of the respective variables, such as patients' body mass index or estimated glomerular filtration rate, whereas this was more heterogeneous for the definition of subgroups according to patients' age. This problem could be tackled by choosing statistical methods which do not rely on fixed categorizations of investigated variables.[40–44] In a recently published simulation study, we could show that the probability to detect heterogeneous

treatment effects regarding continuous variables can be increased when data are not split into categories.[45]

Subgroup interaction tests

Most of the reviewed articles reported the use of interaction tests for the comparison of measured treatment effects between subgroups (94%), which almost appeared to be standard in this respect. Especially, this refers to all articles presenting results of subgroup analyses from the journal *NEJM*. In comparison, Wang et al. showed that the use of interaction tests was reported in 27/59 (46%) articles of clinical trials published by *NEJM* in 2005 and 2006.[5] According to findings of the authors Hernandez et al., this was the case in only 11/39 (28%) reviewed cardiovascular RCTs published by a total of eight selected journals in 2002 and 2004.[2] Gabler et al. reviewed a total of 416 articles from RCTs that were published by five journals in the years 2007, 2010 and 2013, *NEJM, The Lancet, Journal of the American Medical Association, British Medical Journal* and *Annals of Internal Medicine*.[18] In this review, 91/270 (34%) of the included trial publications presenting results of at least one subgroup analysis reported the use of interaction tests.

Conclusion

According to our results, subgroup analyses were reported more frequently in reviewed articles of larger RCTs. In comparison with previous reviews, greater emphasis was put on providing information about the specification and conduct of reported subgroup analyses. We were able to comprehend the pre-specification of subgroup analyses in most reviewed cases and subgroup analyses were performed almost as standard in combination with interaction tests. Nonetheless, we also detected some remaining shortcomings. A critical finding is the increased likelihood of reporting subgroup analyses in case of overall non-significant primary analyses, as this might refer to a 'fishing for significance' issue that carries the risk of false-positive results. It is difficult to verify the pre-specification of subgroup analyses since trial protocols have not been provided for publications outside *NEJM*. Therefore, we believe that the publication of a trial protocol should be seen as a prerequisite. In addition, a documentation of planned subgroup analyses should be enabled for the online registration of trials. Besides that, it is worth paying greater attention to a full pre-specification of both subgroup characteristics and cut-off values for the categorization of subgroups according to continuously scaled characteristics. A more uniform stratification of patients could increase the comparability between results of subgroup analyses from different clinical trials. Further reviews could also focus on cut-off values that were used for the definition of subgroups to guide and streamline respective decisions in upcoming clinical trials. However, more powerful statistical analyses do even refrain from using cut-off values. According to previous recommendations, it would be also reasonable to limit the number of subgroup analyses for clinical trials in future, especially as results from subgroup analyses are prone to errors caused by multiplicity issues and rarely are adjusted for multiple testing. To increase reliability, results of such subgroup analyses should be confirmed by further independent clinical trials.

Authors' note

The final version of the article was approved by all authors.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Alexander Hapfelmeier https://orcid.org/0000-0001-6765-6352
Bernhard Haller https://orcid.org/0000-0002-9723-393X

Data accessibility statement

Relevant data extracted from reviewed articles are provided as Supplemental Material.

Supplemental material

Supplemental material for this article is available online.

References

1. 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.
2. Hernandez AV, Boersma E, Murray GD, et al. Subgroup analyses in therapeutic cardiovascular clinical trials: are most of them misleading. *Am Heart J* 2006; 151(2): 257–264.
3. Vidic A, Chibnall JT, Goparaju N, et al. Subgroup analyses of randomized clinical trials in heart failure: facts and numbers. *ESC Heart Fail* 2016; 3(3): 152–157.
4. Ferreira JC and Patino CM. Subgroup analysis and interaction tests: why they are important and how to avoid common mistakes. *J Bras Pneumol* 2017; 43(3): 162.

5. Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine-reporting of subgroup analyses in clinical trials. *N Engl J Med* 2007; 357: 2189–2194.
6. Dahabreh I, Hayward R and Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol* 2016; 45: 2184–2193.
7. Ting N. Statistical interactions in a clinical trial. *Ther Innov Regul Sci* 2018; 52(1): 14–21.
8. Kent DM, Rothwell PM, Ioannidis JP, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010; 11: 85.
9. Rothwell PMM. Treating individuals2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; 365: 176–186.
10. Naggara O, Raymond J, Guilbert F, et al. The problem of subgroup analyses: an example from a trial on ruptured intracranial aneurysms. *AJNR Am J Neuroradiol* 2011; 32(4): 633–636.
11. Lu TP and Chen JJ. Subgroup identification for treatment selection in biomarker adaptive design. *BMC Med Res Methodol* 2015; 15: 105.
12. Fishbane S, Shah HH, Kataria A, et al. Subgroup analyses in nephrology clinical trials. *Clin J Am Soc Nephrol* 2012; 7(11): 1872–1876.
13. Sleight P. Debate: subgroup analyses in clinical trials: fun to look at – but don't believe them!. *Curr Control Trials Cardiovasc Med* 2000; 1: 25–27.
14. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355: 1064–1069.
15. Pocock SJ, Assmann SF, Enos LE, et al. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21: 2917–2930.
16. Moreira ED Jr, Stein Z and Susser E. Reporting on methods of subgroup analysis in clinical trials: a survey of four scientific journals. *Braz J Med Biol Res* 2001; 34: 1441–1446.
17. Fan J, Song F and Bachmann MO. Justification and reporting of subgroup analyses were lacking or inadequate in randomized controlled trials. *J Clin Epidemiol* 2019; 108: 17–25.
18. Gabler NB, Duan N, Raneses E, et al. No improvement in the reporting of clinical trial subgroup effects in high-impact general medical journals. *Trials* 2016; 17: 320.
19. Dmitrienko A, Millen B and Lipkovich I. Multiplicity considerations in subgroup analysis. *Stat Med* 2017; 36: 4446–4454.
20. Wang R and Ware JH. Detecting moderator effects using subgroup analyses. *Prev Sci* 2013; 14(2): 111–120.
21. Tanniou J, van der Tweel I, Teerenstra S, et al. Subgroup analyses in confirmatory clinical trials: time to be specific about their purposes. *BMC Med Res Methodol* 2016; 16: 20.
22. Schulz KF and Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005; 365: 1657–1661.
23. 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(1): 71–98.
24. 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.
25. European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP). *Draft guideline on the investigation of subgroups in confirmatory clinical trials*. London: EMA, CHMP, 2014.
26. International Conference on Harmonisation (ICH). *E9 guideline – adopted by CPMP. Statistical principles for clinical trials*. London: EMA, 1998.
27. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
28. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; 339: b2700.
29. World Health Organisation (WHO). *International statistical classification of diseases and related health problems 10th revision. Diseases of the circulatory system, Chapter IX*. 1990. https://www.cdc.gov/nchs/icd/icd10.htm#:~text=International%20Classification%20of%20Diseases%2CTenth%20Revision%20(ICD% 2D10)&text=The%20International%20Classification%20of%20Diseases, and%20presentation%20of%20mortality%20 statistics.
30. Yusuf S, Lonn EM, Pais P, et al. For the HOPE-3 investigators. Blood-pressure and cholesterol lowering in persons without cardiovascular disease. *N Engl J Med* 2016; 374: 2023–2043.
31. Lonn EM, Bosch J, López-Jaramillo P, et al. Blood-pressure lowering in intermediate-risk persons without cardiovascular disease. *N Engl J Med* 2016; 374: 2009–2020.
32. Yusuf S, Bosch J, Dagenais G, et al. Cholesterol lowering in intermediate-risk persons without cardiovascular disease. *N Engl J Med* 2016; 374: 2021–2031.
33. Wright JT Jr, Williamson JD, Whelton PK, et al. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015; 373: 2013–2016.
34. Kernan WN, Viscoli CM, Karen L, et al. Pioglitazone after ischemic stroke or transient ischemic attack. *N Engl J Med* 2016; 374: 1321–1331.
35. Barraclough H and Govindan R. Biostatistics primer: what a clinician ought to know: subgroup analyses. *J Thorac Oncol* 2010; 5(5): 741–746.
36. Sun X, Briel M, Busse JW, et al. Subgroup analysis of trials is rarely easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. *Trials* 2009; 10: 101.
37. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57(3): 229–236.
38. Burke JF, Sussman JB, Kent DM, et al. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015; 351: h5651.
39. Wallach JD, Sullivan PG, Trepanowski JF, et al. Evaluation of evidence of statistical support and corroboration

of subgroup claims in randomized clinical trials. *JAMA Intern Med* 2017; 177: 554–560.
40. Altman DG and Royston P. The cost of dichotomising continuous variables. *BMJ* 2006; 332: 1080.
41. DeCoster J, Gallucci M and Iselin AMR. Best practices for using median splits, artificial categorization, and their continuous alternatives. *J Exp Psychopathol* 2011; 2: 197–209.
42. Royston P and Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Stat Med* 2004; 23: 2509–2525.
43. Bonetti M and Gelber RD. A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. *Stat Med* 2000; 19: 2595–2609.
44. Naggara O, Raymond J, Guilbert F, et al. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. *AJNR Am J Neuroradiol* 2011; 32: 437–440.
45. Haller B, Ulm K and Hapfelmeier A. A simulation study comparing different statistical approaches for the identification of predictive biomarkers. *Comput Math Methods Med* 2019; 2019: 1–15.