Subgroup analysis in Haematologic Malignancies Phase III Clinical Trials: A systematic review

Nerea Báez Gutiérrez[1], Héctor Rodríguez-Ramallo[1], Laila Abdel-kader Martín[1], and Sandra Flores-Moreno[1]

[1]Virgen del Rocio University Hospital

June 4, 2020

Abstract

Aims: To assess the appropriateness of the use and interpretation of subgroup analysis in haematology randomized clinical trials (RCT). Method: A systematic review of Medline including Haematology phase III RCT published between January 2013 and October 2019 was carried to identify subgroup analysis reported. Information related to trials characteristics, subgroup analysis reported and claims of subgroup difference were collected. Results: A total of 98 studies reporting subgroup analyses were identified. Of those, 24 RCT reported 46 claims of subgroup difference. Among them, 44 were claims for the primary outcome, of which 25 were considered strong claims and 17 were considered suggestions of a possible effect. Authors included subgroup variables for the primary outcome measured at baseline for 38 claims (n = 86.36%), used subgroup variable as stratification factor at randomization for 15 (34.09%), clearly prespecify their hypothesis for 11 (25%), the subgroup effect was one of a small number of hypothesised effects tested for 17 (38.36%), carried out a test of interaction that provide statistically significant for 18 (40.91%), documented replication of a subgroup effect with previously related studies for 11 (25%), identify consistency of a subgroup effect across related outcome for 10 (22.72%), and provided a biological rationale for the effect for 8 (18.18%). Of the 44 claims for the primary outcome, 34 (77.27%) met 4 or fewer of the 10 credibility criteria. Conclusion: Credibility of subgroup claims reported in haematology RCT lack of credibility, even when claims are strong. Information about subgroup difference should be interpreted ca

Introduction

Subgroup analysis are important elements in the report of results of randomized clinical trials (RCTs)[1 -2]. Clinical practice guidelines (GPC) recommendations in haematology are guided by phase III RCTs results. Usually only average results are reported in RCTs and trial participants are frequently recruited from a heterogeneous population. Subgroup analysis are born with the aim to detect subgroup effects, in other words they have the aim of predicting which patients will benefit more from therapies[2-4].

Interpretation of subgroup analysis is potentially important for treatment decisions in medical practice. Subgroup analysis can provide clinicians with a better perspective on the individualized treatment of patients, which is particularly interesting in the field of haematology due to the lower therapeutic index and higher toxicity of used-drugs[5-6]. However, subgroup analysis can introduce analytical challenges leading to misleading and exaggerated results, which may result in denial of a beneficial treatment or even receiving a potentially harmful or ineffective treatment[7-9.]

Subgroups analysis have the potential to generate hypotheses for further prospective investigation[10], their exploratory nature requires results to be confirmed in a new study to ensure their findings with statistical reliability. However, confirmatory studies are generally never carried out and decisions in clinical practice are made with this lack of information. On the other hand, the option of completely discarding subgroups

analysis finding is also a decision that has its consequences, and is especially controversial in situations with very high risks or costs that are difficult to assume, which are not uncommon in clinical practice[11].

Concerns about the correct interpretation of subgroup analysis has recently grown. With the intention of reducing the problems related to subgroup analysis misinterpretation, several tools have been developed to assess the credibility of the effects of subgroups reported in RCTs[12-17].

With the results of this study we will be able to determine if subgroup analysis claims of phase III RCTs in haematology malignancies are carried out correctly.

The main objective of this study is to assess the appropriateness of the use and interpretation of subgroup analysis in recently published haematologic malignancies RCTs. To achieve our objective the following aspects will be evaluated:

1- To describe subgroups analysis and claims of subgroup effects.

2- To assess study characteristics of subgroup analyses.

3- To examine the analysis and interpretation of subgroup effects for primary outcomes and to assess the credibility of subgroup claims using "the 10 criteria for assessing the credibility of a subgroup claim" by Sun et al 2012[17].

Methods:

Literature Search

This Systematic review was designed to summarize the available data addressing the following research question, framed in the Population-Intervention-Comparator- Outcome-Study design (PICOS) framework: (Population) Patients with haematological malignancies; (Intervention) subgroup analysis; (Comparison) studies with comparator will be considered; (Outcomes), subgroup analysis; (Study design), phase III randomized clinical trials.

A systematic search was performed following Preferred Reporting Items for a Systematic Review and Meta-analysis (PRISMA) guidelines[18]. The search was performed using Mesh terms-controlled vocabulary and keywords in MEDLINE database (OVID interfaz including In-process and Epub ahead of print) between January 2013 and October 2019, to identify publications of phase III RCT assessing systemic therapies for haematological malignancies.

The search was performed on October 2019. The full literature search strategy is available at supplemental material (Appendix A).

The following criteria were used for trial selection

Eligibility criteria:

We considered eligible all published Phase III randomized clinical trials for haematological malignancies with subgroup analysis reported. Not language restriction was applied.

Exclusion criteria:

1. Paediatric patients (<18 years of age).
2. Pooled data from two or more trials.
3. Studies exploring devices, behavioural or supportive care interventions.
4. The report does not include the entire population enrolled in the original article (i.e. the report focuses on a subset of the original study population).

In cases were multiple publications from the same trial were identified, the initial publication was used for the analysis if it was published during the studied period.

Study Screening and Selection

Two investigators independently examined the titles and abstracts of the search results using the predefined inclusion criteria. For all titles that appear to meet the inclusion criteria or those where there was some uncertainty, full text was accessed. The two reviewers assessed whether the articles met the selection criteria. Any disagreements were resolved by discussion or arbitration from a third reviewer. Reasons for excluding studies were recorded and is available at supplemental material.

Data extraction

For data extraction additional sources referenced in the included study (i.e., trial register, published protocol and online supplements) were used. Data were extracted and entered in a structured Microsoft Excel (Redmond, WA, USA) database.

Eligible RCTs were evaluated to determine whether a subgroup analysis was reported. A subgroup analysis was defined as a statistical analysis that explores whether effects of the intervention differ according to status of a subgroup variable. A subgroup effect was defined as a difference in the magnitude of a treatment effect across a group of a study population[16]. For each RCT reporting subgroup analysis and subgroup claims the following information was collected:

1. *Trial characteristics:* information on funding source, year and journal of publication, journal impact factor (<10 o >10), haematological malignancy type, disease status (naive/untreated or refractory/relapse), type of intervention (chemotherapy, immunotherapy or haematopoietic transplant),centre (multicentric or unicentric), trial design (parallel, cross-over or factorial), trial type (superiority, non-inferiority or equivalence), allocation concealment, blinding of patients, number of patients recruited and randomized for the trial and number of treatment arms. The primary endpoint was categorized according to whether results were statistically significant and the type of outcome variable (time-to-event, binary, continuous or count).
2. *Reporting of subgroup analysis* : number of subgroup factors, type of subgroup factors (clinical factors or biomarkers), number of subgroup analysis and outcomes for subgroup analysis reported, forest plots used, prespecify or post hoc subgroup, statistical method used to assess heterogeneity of the treatment effect (descriptive only, subgroup P values and confidence interval or interaction test).

A subgroup factor was defined as each of the subgroup analysed in the RCT (i.e. sex, age, presence of a mutation).

Claims of subgroup effects: Subgroup claims mode of presentation (abstract or text only), number of subgroup claims, subgroup variable (primary or secondary outcome) and number of outcomes for subgroup claims were recorded. A subgroup effect was considered claimed when the authors states in the abstract or discussion that the effect of intervention differs between the categories of the subgroup variable. Claims of subgroup effect were classified according to the strength of the claim into 3 categories: Strong claim, claim of a likely effect or suggestion of a possible effect based on Sun et al 2009 clasification[16](Appendix B). To evaluate the credibility of subgroup claims for primary outcomes "the 10 criteria for assessing the credibility of a subgroup claim" by Sun et al 2012[17] were applied (Appendix C). These criteria have been widely used[13-15,17] and are recommended for assessing how much confidence to place in subgroup analyses[19]. If the subgroup claim met less than half of criteria, the credibility of this claim was considered low.

Assessment of risk of bias

Risk of bias was assessed using the Cochrane Collaboration´s tool for assessing risk of bias in randomized trials[20]. This tool is composed by 5 domains: bias arising from the randomization process; bias due to deviations from intended interventions; bias due to missing outcome data; bias in measurement of the outcome; and bias in selection of the reported result. For each domain, the tool comprises:

a series of 'signalling questions'; a judgement about risk of bias for the domain, which is facilitated by an algorithm that maps responses to the signalling questions to a proposed judgement; free text boxes to justify responses to the signalling questions and risk-of-bias judgements; and an option to predict (and explain) the likely direction of bias. Risk of bias was assessed by two independent reviewers. Possible disagreements

between reviewers were resolved by discussion or arbitration by a third reviewer when consensus could not be reached.

Data analysis

A descriptive analysis was developed. Continuous and categorical variables were presented as mean (range) and n (%), respectively.

For those RCTs that stated a subgroup effect without providing an interaction test, p interaction was calculated using the Joaquin Primo calculator[21], to verify that there was indeed statistical significance.

Results

The literature search identified 1622 studies. After a first review by title or abstract and removing duplicates, 321 articles were selected for a full text review. Finally, 98 articles were included. (Figure 1). Articles excluded and the reason of exclusion are available at supplemental material (Appendix D).

Characteristics of trials included in the analysis

The characteristics of the trials included in this study are listed in table 1. These 98 publications reported data on 48,245 randomized patients (Median: 402; range: 82-1623).

A 77.25% (n = 76) studies were funded by industry. Most of the trials were published during 2015 (18.36%; n = 18) and 2016 (20.41%; n = 20). The New England Journal of Medicine (26.53%; n = 26) and Lancet Oncology (20.41%; n = 20) were the most selected journals for publication of these trials. An 85.7% (n = 84) of the studies were published in high impact journals (impact factor >10).

The most common malignancies explored were Non-Hodgkin lymphoma (25.51%; n = 25), multiple myeloma (20.41% n= 20), acute myeloid leukaemia (20.41%; n = 20), and chronic lymphocytic leukaemia (20.41%; n = 20). The most common intervention was chemotherapy (50%; n = 49). Stated primary endpoint was statistically significant in 65.31% (n = 64) of trials.

Subgroup analysis

Characteristics of reported subgroup analysis are listed in table 2. Subgroup analysis were mentioned in the method section for 46.94% (n = 46) trials, 89.90% (n = 88) in the results sections, 56.12% (n=55) in the discussion section and 26.53% (n = 26) in supplemental appendix.

At least 6 subgroup factors were reported in 63.26 % (n = 62) of trials. Related the type of subgroup factors 30.61% (n = 30) were clinical factor and 66.33% (n = 65) were clinical factor plus biomarkers. More than 6 subgroup analysis were reported in 71.43% (n = 70) of the trials. More than one outcome was reported in 25.51% (n = 25) of trials (mean:1; range:1-3). To show the results of subgroup analysis forest plots were used in 77.55% (n = 76) of the trials.

For 11.22% (n = 11) of trials, it was unclear whether subgroup analysis was prespecify or post hoc, in 50% (n = 49) of trials were prespecify and 31.63% (n = 31) were post hoc.

Only 18.37% (n = 18) use an interaction test to assess heterogeneity of the treatment effect; a 17.35% (n =17) reported subgroup analysis without any statistical analysis.

Claims of subgroup effects

Characteristics of subgroup claims are listed in table 3. In 24 RCTs authors claim heterogeneity of treatment effect of at least one subject subgroup, 13 made a claim for a primary outcome, 2 for secondary outcomes and 9 for both primary and secondary outcomes. Six (25.00%) of these RCTS presented subgroup claims in the articles abstract and five (20.83%) were based on significant interaction tests, whereas the claims were based only on within-subgroup comparisons for most of trials (54.17%; n = 13). More than one subgroup claim was made in 54.17% (n = 13) of trials.

A total of 46 subgroup difference were claimed in these 24 trials (44 for primary outcomes and 2 for secondary outcomes). These claims were classified as 26 (59.10%) strong claims, two (4.54%) as claims of a likely effect and 18 (40.91%) as suggestion of a possible effect.

Respect to the 10 criteria to assess credibility of subgroups claims (table 4): Authors included subgroup variables for the primary outcome measured at baseline for 38 claims (86.36%), used subgroup variable as stratification factor at randomization for 14 (34.09%) claims, clearly prespecify their hypothesis for 11 (25.00%) claims, correctly prespecify direction for 5 (11.36%) claims, tested a small number of hypothesis for 17 (38.63%) claims, carried out a test of interaction that provide statistically significant for 18 (40.91%) claims, documented replication of a subgroup effect with previously related studies for 11 (25.00%) claims, identify consistency of a subgroup effect across related outcome for 10 (22.72%) claims, and provided a biological rationale for the effect for 8 (18.18%) claims. Of the 44 claims for the primary outcomes, 34 (77.27%) met 4 or fewer of the 10 criteria. For strong claims, 15 (60.00 %) met three or less criteria and only 6 (24.00%) met more than 5 criteria.

Risk of Bias Graphs Within Studies and across studies is available at supplemental material (Appendix D).

Discussion

Limitations of reporting subgroup analysis in RCT have been widely reported on the literature. Inflated false positives due to multiple testing, high false negatives due to inadequate statistical power and inappropriate a priori specification are well-known limitations of subgroup analysis[2,7-8,22-24]. A prespecified subgroup analysis is one that is planned and documented before any examination of the data. They are more reliable than those no prespecified because their hypotheses are based on biological rationale or data obtained on previous studies. In this review only half of trials conducted prespecified subgroup analysis. When analysis of a large number of subgroups are made, even if a hypothesis has been clearly specified, their results should be considered cautiously, since the strength of inference associated with the apparent confirmation of any single hypothesis will decrease if it is one of a large number that have been tested[25]. In this systematic review, multiple subgroup analyses were performed, around three quarters of trials reported at least 6 subgroups. Statistical analysis of interaction establishes the difference in benefit between subgroups by calculating interaction probability (p), which suggests that chance is an unlikely explanation for apparent differences, therefore the interaction test is the appropriate method to analyse subgroups. In this review only a few trials (18.37%) used an interaction test to assess heterogeneity of the treatment effect.

Due to important methodological problems bias, subgroup interpretation can lead to erroneous conclusions, producing wrongful clinical decision making. Several tools have been developed to assess the credibility of the effects of subgroups reported in clinical trials[12-17]. In our study we have based ourselves on the "10 criteria to assess credibility of subgroup claims" by Sun et al 2012[17]. The credibility of subgroup claims in phase III haematology RCT was low. Of the 44 claims of a subgroup effect for the primary outcome identified, 26 were strong claims and only 24% (n = 6) of these claims were able to satisfy at least half of the credibility criteria and none satisfied all criteria. Multiple significant interactions were the only criteria satisfied by more than 50% of the claims. All 24 assessed studies failed to prespecify the correct direction of the subgroup hypotheses, and the hypothesis was prespecified for only 11 (25%) claims.

Sun et al 2012[17] considered three out of their 10 criteria as critical: the use of subgroup variables measured at baseline, prespecification of subgroup hypothesis and statistical significance of interaction test. In our study the first of these criteria was met for most of trials (86.36%), however the other two criteria were only met by 25.2% and 40.91% respectively. As stated before, interaction test is the appropriate method to analyse subgroups, but only a 40% of strong claims of this review were made base on this test. This finding indicates that most authors are unaware of how to interpret a subgroup analysis correctly and make statements based on intragroup comparisons, instead of intergroup comparisons. The latter determines evidence of differences in the results for different subgroups, this comparison is made by the interaction test. The lack of compliance of previously cited criteria in the claims of the haematology RCTS demonstrates their limited credibility.

Similar results have been reported in other studies areas. Zhang et al 2015[26], reported low credibility of

subgroup claims in phase III RCT solid tumours using The CONSORT statements to evaluate subgroup claims[27]. They found as most common problems for reporting subgroup analysis the great number of subgroups reported, although frequently not prespecified and the underused of interaction test. Sun et al. 2012[17] reported low credibility of subgroup claims in pharmacological RCT published in 2007. Most of these trials failed to prespecify the hypotheses or present significant interaction tests. Two recent reviews investigated subgroup analysis quality in low back pain management trials[28-29] and reported the failure to specify the subgroup hypotheses a prior as a common problem in trials, which is also consistent with our findings. Vidic et al 2016[10] reviewed phase III cardiovascular RCTs with subgroup analysis, concluding that subgroup analysis were reported with several shortcomings, including lack of prespecification and testing of a large number of subgroups without the use of the statistically appropriate test for interaction. All these studies reported the failure to specify the subgroup hypotheses, many subgroup analyses conducted and underuse of interaction test as common problems in trials, which is consistent with our findings.

By contrast in other studies the number of claims of subgroup effect in this review was low. Zhang et al 2015[26], Sun et al 2012[17], Saragiotto et al[29] and Vidic et al 2016[10] reported that a 54.26%, 40.10%, 57.57%, 53.84% of trials assessed made claims of subgroup effect, respectively. The number of subgroup claims identify in haematological trials was half of those reported in other areas.

This study had several strengths: It is the first systematic review of the credibility of subgroup analysis reported on haematological malignancies RCTs. A rigorous systematic review method was employed, and standardized criteria were used for assessing credibility of subgroup claims[17].

This study had several limitations: This study is based on authors' reported trial information in published articles, which may be vulnerable to selective reporting or underreporting. Our study was limited to phase III RCT, although Sun et al 2012[17]criteria could be applied to all phase clinical trials. The low number of subgroup claims identified is also a limitation of this study.

Conclusions

In summary, subgroup analysis in phase III haematology malignancies RCTs are of poor quality, identifying flaws already described in other areas of study, such as the great number of subgroups reported, inappropriate a priori specification and the underused of interaction test.

Although not as frequent as in other areas, subgroup claims credibility was low. Most claims do not meet critical criteria; therefore, clinicians should interpret these results with caution. Subgroup analysis should be carried out due to the potential information they can provide, however researchers should be more cautious before claiming the existence of a subgroup effect.

Bibliography

1. Guyatt G, Rennie D, Meade MO, Cook, D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Chapter 1. Third Edition. 2015.
2. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. Lancet 2005; 365 (9454): 176-86.
3. VanderWeele TJ, Knol MJ. Interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. Ann Intern Med 2011;154:680e3.
4. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G.. How to use a subgroup analysis: users' guide to the medical literature. JAMA 2014;311:405e11.
5. Thanarajasingam G, Minasian LM, Baron F, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies [published correction appears in Lancet Haematol. 2019 Mar;6(3):e121]. Lancet Haematol. 2018;5(11):e563-e598.
6. Bennett CL, Tigue CC, Angelotta C, McKoy JM, Edwards BJ Adverse effects of drugs used to treat hematologic malignancies: surveillance efforts from the research on adverse drug events and reports project. Semin Thromb Hemost. 2007;33(4):365-372.

7. Wittes J. On looking at subgroups. *Circulation* . 2009;119(7):912-915.
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(21):2189-2194.
9. Koch A, Framke T. Reliably basing conclusions on subgroups of randomized clinical trials. *J Biopharm Stat* . 2014;24(1):42-57.
10. Vidic A, Chibnall JT, Goparaju N, et al. Subgroup analyses of randomized clinical trials in heart failure: facts and numbers. ESC Heart Fail. 2016 Sep; 3(3): 152–157.
11. 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(8):e028751.
12. Gil-Sierra MD, Fenix-Caballero S, Abdel-Kader Martin L, et al. Checklist for Clinical Applicability of Subgroup Analysis. J Clin Pharm Ther. 2020 Jun;45(3):530-538
13. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;116:78-84.
14. Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010;340:c117.
15. Sun X, Ioannidis JPA, Agoritsas T, et al. How to use a subgroup analysis. JAMA 2014;311:405
16. 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.
17. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012;344:e1553.
18. Moher D., Liberati A., Tetzlaff J, et al. Group P. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int. J. Surg. 2010;8:336–341.
19. Oxman AD. Subgroup analyses. BMJ 2012;344:e2022.
20. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
21. Primo J, Escrig J. MetaSurv: Excel calculator for survival meta-analyzes. 2008. Avaialble in: http://www.redcaspe.org/herramientas/descargas/MetaSurv.xls
22. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004; 57: 229–236.
23. Assmann SF, Pocock SJ, Enos LE, Kasten LE.. Subgroup analysis and other (mis) uses of baseline data in clinical trials. Lancet 2000; 355: 1064–1069.
24. Feinstein AR. The problem of cogent subgroups: a clinicostatistical tragedy. J Clin Epidemiol 1998; 51: 297–299.
25. Moreira ED Jr, Stein Z, Susser E. Reporting on methods of subgroup analysis in clinical trials: a survey of four scientific journals. Braz J Med Biol Res. 2001;34(11):1441-1446.
26. Zhang S, Liang F, Li W, Hu X. Subgroup Analyses in Reporting of Phase III Clinical Trials in Solid Tumors. *J Clin Oncol* . 2015;33(15):1697-1702.
27. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Int J Surg. 2011;9(8):672-677.
28. Mistry D, Patel S, Hee SW, Stallard N, Underwood M. Evaluating the quality of subgroup analyses in randomized controlled trials of therapist-delivered interventions for nonspecific low back pain: a systematic review. Spine (Phila Pa 1976) 2014;39:618e29.
29. Saragiotto BT, Maher CG, Moseley AM, et al. A systematic review reveals that the credibility of subgroup claims in low back pain trials was low. J Clin Epidemiol. 2016;79:3-9.

Hosted file

`Table 1.docx` available at https://authorea.com/users/329737/articles/456662-subgroup-analysis-in-haematologic-malignancies-phase-iii-clinical-trials-a-systematic-review

Hosted file

Table 2.docx available at https://authorea.com/users/329737/articles/456662-subgroup-analysis-in-haematologic-malignancies-phase-iii-clinical-trials-a-systematic-review

Hosted file

Table 3.docx available at https://authorea.com/users/329737/articles/456662-subgroup-analysis-in-haematologic-malignancies-phase-iii-clinical-trials-a-systematic-review

Hosted file

Table 4.docx available at https://authorea.com/users/329737/articles/456662-subgroup-analysis-in-haematologic-malignancies-phase-iii-clinical-trials-a-systematic-review

Hosted file

Figure 1. Flow chart.pptx available at https://authorea.com/users/329737/articles/456662-subgroup-analysis-in-haematologic-malignancies-phase-iii-clinical-trials-a-systematic-review