

# Reporting on methods of subgroup analysis in clinical trials: a survey of four scientific journals

E.D. Moreira Jr.<sup>1,2</sup>,  
Z. Stein<sup>3</sup> and  
E. Susser<sup>4</sup>

<sup>1</sup>Núcleo de Epidemiologia e Estatística, Centro de Pesquisas Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, BA, Brasil

<sup>2</sup>Núcleo de Apoio à Pesquisa, Associação Obras Sociais Irmã Dulce, Salvador, BA, Brasil

<sup>3</sup>Department of Psychiatry, Columbia University, New York, NY, USA

<sup>4</sup>Division of Epidemiology, School of Public Health, Columbia University, New York, NY, USA

## Abstract

Results of subgroup analysis (SA) reported in randomized clinical trials (RCT) cannot be adequately interpreted without information about the methods used in the study design and the data analysis. Our aim was to show how often inaccurate or incomplete reports occur. First, we selected eight methodological aspects of SA on the basis of their importance to a reader in determining the confidence that should be placed in the author's conclusions regarding such analysis. Then, we reviewed the current practice of reporting these methodological aspects of SA in clinical trials in four leading journals, i.e., the New England Journal of Medicine, the Journal of the American Medical Association, the Lancet, and the American Journal of Public Health. Eight consecutive reports from each journal published after July 1, 1998 were included. Of the 32 trials surveyed, 17 (53%) had at least one SA. Overall, the proportion of RCT reporting a particular methodological aspect ranged from 23 to 94%. Information on whether the SA preceded/followed the analysis was reported in only 7 (41%) of the studies. Of the total possible number of items to be reported, NEJM, JAMA, Lancet and AJPH clearly mentioned 59, 67, 58 and 72%, respectively. We conclude that current reporting of SA in RCT is incomplete and inaccurate. The results of such SA may have harmful effects on treatment recommendations if accepted without judicious scrutiny. We recommend that editors improve the reporting of SA in RCT by giving authors a list of the important items to be reported.

## Key words

- Subgroup analyses
- Randomized clinical trials
- Research design
- Epidemiological methods
- Effect modification

## Correspondence

E.D. Moreira Jr.  
Rua Waldemar Falcão, 121  
40295-001 Salvador, BA  
Brasil  
Fax: +55-71-356-2155  
E-mail: edson@cpqgm.fiocruz.br

This study was partially funded by a Fogarty grant from the U.S. Public Health Service (NIH D43 TW 00018).

Received January 4, 2001  
Accepted September 4, 2001

## Introduction

The randomized clinical trial (RCT) has been increasingly accepted as a main source of evidence for making clinical decisions and basing new policies. RCT now have more precise definitions of patients' eligibility, treatment schedules, and outcome criteria,

appropriate blinding and objectivity in assessments of patients, and better data collection and processing. In addition, methods of statistical analysis such as significance testing and estimation techniques have become essential features in the reporting of trial findings (1-3). Despite the robustness of this strategy, results of an RCT cannot be

adequately interpreted without information about the methods used in the study design and the data analysis. Readers need these specific data in order to make informed judgments regarding the internal and external validity of published reports of clinical trials. Such reports, however, frequently omit important features of design and analysis (4,5). DerSimonian and colleagues (6) found that investigators reported only 56% of eleven methodological items in 67 trials published in four general medical journals.

Subgroup analyses are frequently encountered in reports of clinical trials. In a survey of three leading medical journals, Pocock and co-workers (7) found that 23 out of 45 trials (51%) had at least one subgroup analysis that compared the response to treatment in different subsets of patients. While there have been some reports on guidelines for assessing the strength of inferences based on subgroup analyses, and assisting clinicians in making decisions regarding whether to base a treatment decision on the results of such analyses (8-13), neither the Consolidated Standards of Reporting Trials (CONSORT) statement (a checklist of 21 items and a flow diagram), which identifies key pieces of information necessary to evaluate the validity of a trial report (14), nor the Instructions for Authors in leading medical journals include specific information on reporting subgroup analyses in RCT.

In this paper we review the current practice of reporting methodological aspects of subgroup analysis in clinical trials in four leading journals. Our aim here was to show how often inaccurate and incomplete reports are published. We also suggest a list of important items to be reported in subgroup analysis of RCT.

## Material and Methods

We examined 32 reports of randomized controlled clinical trials published in the New England Journal of Medicine (NEJM),

the Lancet, the Journal of the American Medical Association (JAMA), and the American Journal of Public Health (AJPH). The choice of journals was based on the size of their audience and their wide international acceptance. Therefore, two leading American medical journals (NEJM and JAMA), and one leading British medical journal (Lancet) were selected. The fourth journal chosen is a leading journal in the field of Epidemiology, since we wanted to have a sample representative of both medically oriented journals as well as an epidemiologically oriented journal. Eight consecutive reports from each journal published after July 1, 1998 were included. The reports were published within two months in the NEJM and Lancet, three months in JAMA, and 6 months in the AJPH. The trials involved many diseases (including 11 infectious, 3 cardiovascular, 2 respiratory, and 4 neurological disorders) and were considered to be representative of clinical trials reported in major medical journals.

We then surveyed the selected trials to determine how frequently certain aspects of design and analysis of subgroup analysis were reported. We recorded whether a particular methodological aspect of a clinical trial was mentioned and not whether a particular method was used. If the author said that a specific test for interaction was not performed, that was coded positively as a report on the item, as was a statement that a test for interaction had been carried out. We selected eight items on the basis of their importance to a reader in determining the confidence that should be placed in the author's conclusions regarding the subgroup analysis, their applicability across a variety of medical specialty areas, and their ability to be detected by the scientifically literate general medical reader. We chose the following items: 1) *a priori/post hoc* subgroup analysis (information on whether the subgroup analysis preceded or followed the analysis); 2) number of subgroups examined (in-

formation about how many subgroups were examined); 3) justification for subset definition (information explaining how the subgroups were stratified); 4) when subgroups were delineated (information on whether the subgroups were defined after or before randomization); 5) statistical methods (the names of specific tests or techniques used for statistical analyses); 6) power (information describing the determination of sample size or the size of detectable differences); 7) clinical significance (information on the clinical significance of the interaction), and 8) overall treatment comparison (information about the prominence of the comparison between the main treatment groups) (Table 1).

For each trial, one of us (E.D.M.) completed a standardized evaluation of each item to determine whether the item was reported, not reported, or unclear (when the information on the item was found to be incomplete or ambiguous). If an item was clearly not applicable to a particular study, it was regarded as reported.

## Results

Of the 32 trials surveyed, 17 (53%) had at least one subgroup analysis. Since our inter-

est was centered on subgroup analysis, henceforth the results presented will refer to this subset of articles. The frequency of reporting each of the eight items included in our list is shown in Table 2. The proportion of clinical trials reporting a particular methodological aspect ranged from 23 to 94%. Only 7 (41%) of the studies presenting results of subgroup analysis reported whether the analysis was planned *a priori* or not, and overall the problem of subgroup analysis not planned *a priori* was restricted to one quarter of the 32 papers evaluated. Among the 14 (82%)

Table 1. List of important methodological items when reporting subgroup analysis in randomized clinical trials.

- A priori/post hoc subgroup analysis (information on whether the subgroup analysis preceded or followed the analysis)
- Number of subgroups examined (information about how many subgroups were examined)
- Justification for subset definition (information explaining how the subgroups were stratified)
- When subgroups were delineated (information on whether the subgroups were defined after or before randomization)
- Statistical methods (the names of specific tests or techniques used for statistical analyses)
- Power (information describing the determination of sample size or the size of detectable differences)
- Clinical significance (information on the clinical significance of the interaction)
- Overall treatment comparison (information about the prominence of the comparison between the main treatment groups)

Table 2. Frequency of reporting eight important aspects for planning and interpreting the results of subgroup analysis in randomized clinical trials in four scientific journals.

Methodological aspect	NEJM (N = 4)			JAMA (N = 3)			Lancet (N = 6)			AJPH (N = 4)			Total (N = 17)		
	R	U	O	R	U	O	R	U	O	R	U	O	R (%)	U (%)	O (%)
A priori/post hoc subgroup analysis	2	-	2	2	-	1	1	1	4	2	-	2	7 (41)	1 (6)	9 (53)
Number of subgroups examined	1	1	2	1	1	1	2	-	4	2	-	2	6 (35)	2 (12)	9 (53)
Justification for subset definition	2	1	1	2	-	1	3	1	2	4	-	-	11 (65)	2 (12)	4 (23)
Subgroup delineated after/before randomization	3	-	1	2	1	-	5	-	1	4	-	-	14 (82)	1 (6)	2 (12)
Statistical methods	4	-	-	3	-	-	4	-	2	4	-	-	15 (88)	-	2 (12)
Power estimation	-	-	4	1	-	2	2	-	4	1	-	3	4 (23)	-	13 (77)
Clinical significance	3	1	-	2	1	-	4	1	1	3	-	1	12 (71)	3 (18)	2 (12)
Overall treatment comparison	4	-	-	3	-	-	6	-	-	3	-	1	16 (94)	-	1 (6)

NEJM = The New England Journal of Medicine; JAMA = The Journal of the American Medical Association; AJPH = American Journal of Public Health. R = reported; U = unclear; O = omitted.

trials reporting when the subgroups were delineated, 4 (29%) defined the subgroups after randomization had been conducted. Although 15 (88%) articles stated the statistical methods used, only 6 (35%) carried out specific tests or techniques for assessing the presence of interaction. The remaining 9 (53%) inappropriately employed P values for subgroups for assessing whether the treatment effect varied between subgroups. Of the 16 (94%) reports including the results of the main comparison of treatment groups, 10 (63%) yielded either nonsignificant ( $N = 7$ ) or significantly worse ( $N = 3$ ) results for the treatment under investigation. Of the total possible number of items to be reported, NEJM, JAMA, Lancet and AJPH clearly mentioned 59, 67, 58 and 72%, respectively.

## Discussion

The frequency of reporting a particular item varied widely from 23 to 94% in all four journals combined. Although all items were relevant, they were not deemed equally important. Information on whether the subgroup analysis was intended in the trial protocol and the number of subgroups examined were thought to be essential. Approximately two thirds of the trials in this survey either omitted or ambiguously mentioned these items. This affects other methodological issues, as it may increase the risk of a type I error, and can make the interpretation of the statistical tests used extremely difficult (15, 16). One should be particularly cautious about analysis of a large number of subgroups, even if investigators have clearly specified their hypothesis in advance 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. Unfortunately, as our data indicate, the reader is quite often not sure about the number of possible interactions that were tested.

Because subgroup analysis always in-

cludes fewer patients than does the overall analysis, they carry a greater risk of making a type II error, falsely concluding that there is no difference. Thus, just as it is possible to observe spurious interactions, chance is likely to lead to some studies in which even a real difference in treatment effect is not apparent. Statistical power estimations were omitted in 13 (76%) of the articles. Therefore, we are commonly left with a limited knowledge of the probability of missing a real difference in the trials reported. Moreover, this finding might indicate that a high proportion of the subgroup analyses were not intended when the trial protocol was planned. We surveyed whether a particular methodological aspect of a subgroup analysis was mentioned, not whether a particular method was appropriately used (i.e., whether the method was valid and did not violate the specific assumptions required). Hence, our finding that the vast majority of the trials reported the statistical methods employed in their analysis does not mean that these methods were adequate. In fact, only one third of the articles (6 of 17) utilized appropriate statistical techniques for assessing whether the treatment difference varied between subgroups. Instead, more than half of the trials (9 of 17) inadequately focused on subgroup P values.

Differences in the effect of treatment are likely to occur due to biological variability. However, it is only when these differences are practically important (that is, when they are large enough that they would lead to different clinical decisions for different subsets) that there is any point in considering them further. Twelve (71%) of the reports went beyond the appraisal of the statistical significance of the treatment differences observed, to address the clinical significance of the findings.

Our findings suggest that there is a wide chasm between what a subgroup analysis of a trial should report, and what is actually published in the literature. In response to increasing evidence that reporting of RCT is

imperfect, there has been a concerted effort to set standards for reporting RCT (1,6,7,14-16). Among these different proposals, none specifically addressed the problems of reporting subgroup analysis. In this regard, the CONSORT statement (14) included in their list of recommendations just one of our suggestions, precisely, reporting whether or not the subgroup analysis was included in the trial protocol.

Although the items included in our checklist have been long recognized as important and necessary information for interpreting evidence based on subgroup analysis, none of the journals surveyed included in their respective instructions for authors specific recommendations for reporting subgroup analysis. We intended to perform a formal compilation of the most relevant methodological aspects involved in this type of analysis and then review the current practice for reporting them in a sample of major scientific journals. We hope that such checklist may be used to improve the reporting of subgroup analysis of RCT. In this way, already existing guidelines for assisting clinicians in making decisions regarding whether to base a treatment decision on the results of a subgroup analysis could be better used (8-11,17-20).

We conclude that the current reporting of subgroup analysis in RCT is incomplete and inaccurate. In view of their prodigality in reported trials, the results of such subgroup analysis may have harmful effects on treat-

ment recommendations if accepted without judicious scrutiny. Ideally, the report of such an evaluation needs to convey to the reader relevant information regarding the design, conduct and analysis of the trial's subgroup analysis. This should permit the reader to make informed judgments concerning the validity of the subgroup analysis of the trial. Accurate and complete reporting would also be of benefit to editors and reviewers in their deliberations regarding submitted manuscripts.

We recommend that editors improve the reporting of subgroup analysis in RCT by giving authors a list of the important items to be reported. This list will ultimately lead to more comprehensive and complete reporting of RCT. We recognize that this list of recommendations will need revision as it is disseminated and made available to wider audiences. We invite all interested editors and readers to join us in using and perfecting this checklist.

### Acknowledgments

We are very grateful to Dr. Richard Garfield for reviewing the manuscript and for his comments, to Dr. Warren Johnson Jr., Dr. Mary Latka and Mrs. Liliane Zaretsky for their advice, to Isabella Rebouças Arapiraca for secretarial assistance, and to Dr. Beatriz Helena Carvalho Tess for her wise suggestions that ultimately made it possible to accomplish this work.

### References

1. Moher D, Dulberg CS & Wells GA (1994). Statistical power, sample size, and their reporting in randomized controlled trials. *Journal of the American Medical Association*, 272: 122-124.
2. Rennie D (1995). Reporting randomized controlled trials: an experiment and a call for responses from readers. *Journal of the American Medical Association*, 273: 1054-1055.
3. Schulz KF, Chalmers I, Grimes DA & Altman DG (1994). Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *Journal of the American Medical Association*, 272: 125-128.
4. Mosteller F (1979). Problems of omission in communications. *Clinical Pharmacology and Therapeutics*, 25: 761-764.
5. Reiffenstein RJ, Schillroth AJ & Todd DM (1968). Current standards in reported drug trials. *Canadian Medical Association Journal*, 99: 1134-1135.
6. DerSimonian R, Charette LJ, McPeck B & Mosteller F (1982). Reporting on methods in clinical trials. *New England Journal of Medicine*, 306: 1332-1337.
7. Pocock SJ, Huges MD & Lee RJ (1987). Statistical problems in the reporting of clinical trials. *New England Journal of Medicine*, 317: 426-432.
8. Oxman AD & Guyatt GH (1992). A consumer's guide to subgroup analyses. *Annals of Internal Medicine*, 117: 71-77.

- nals of Internal Medicine, 116: 78-84.
9. Stallones RA (1987). The use and abuse of subgroup analysis in epidemiological research. *Preventive Medicine*, 16: 183-194.
  10. Buyse ME (1989). Analysis of clinical trial outcomes: some comments on subgroup analyses. *Controlled Clinical Trials*, 10: 187S-194S.
  11. Bulpitt CJ (1988). Subgroup analysis. *Lancet*, 2: 31-34.
  12. Byar D (1985). Assessing apparent treatment-covariate interactions in randomized clinical trials. *Statistics in Medicine*, 4: 255-263.
  13. Shuster J & van Eys J (1983). Interaction between prognostic factors and treatment. *Controlled Clinical Trials*, 4: 209-214.
  14. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D & Stroup DF (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, 276: 637-639.
  15. The Standards of Reporting Trials Group (1994). A proposal for structured reporting of randomized controlled trials. *Journal of the American Medical Association*, 272: 1926-1931.
  16. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature (1994). Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature: a position paper. *Annals of Internal Medicine*, 121: 894-895.
  17. Simon R (1982). Patient subsets and variation in therapeutic efficacy. *British Journal of Clinical Pharmacology*, 14: 473-482.
  18. Armitage P (1981). Importance of prognostic factors in the analysis of data from clinical trials. *Controlled Clinical Trials*, 1: 347-353.
  19. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J & Smith PG (1976). Design and analysis of randomized clinical trials requiring prolonged observation on each patient. I. Introduction and design. *British Journal of Cancer*, 34: 585-612.
  20. Liberati A, Himel HN & Chalmers TC (1986). A quality assessment of randomized control trials of primary treatment of breast cancer. *Journal of Clinical Oncology*, 4: 942-951.