

left-ventricular dysfunction: the CAPRICORN randomized trial. *Lancet*. 2001;357:1385-1390.

5. Eccles M, Steen N, Grimshaw J, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet*. 2001;357:1406-1409.

6. Villar J, Ba'aqeel H, Piaggio G, et al. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. *Lancet*. 2001;357:1551-1564.

7. Carroli G, Villar J, Piaggio G, et al. WHO systematic review of randomised controlled trials of routine antenatal care. *Lancet*. 2001;357:1565-1570.

8. Hill AB. The reasons for writing. *BMJ*. 1965;4:870.

9. Chalmers I. Improving the quality and dissemination of reviews of clinical research. In: Lock S, ed. *The Future of Medical Journals: In Commemoration of 150 Years of the British Medical Journal*. London, England: British Medical Journal Books; 1991:127-146.

10. Chalmers I, Altman DG. How can medical

journals help prevent poor medical research? some opportunities presented by electronic publishing. *Lancet*. 1999;353:490-493.

11. Skelton JR, Edwards SJ. The function of the discussion section in academic medical writing. *BMJ*. 2000;320:1269-1270.

12. Moher D, Schulz KF, Altman DG, Lepage L. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191-1194.

Quality of Reporting of Randomized Trials as a Measure of Methodologic Quality

Karin Huwiler-Müntener, MD

Peter Jüni, MD

Christoph Junker, MD, MSc

Matthias Egger, MD, MSc

THE RANDOMIZED CONTROLLED trial (RCT) is the design of choice for evaluating the effectiveness of health care interventions, but trials are not immune to bias. The validity of their results is threatened by the subversion of randomization, resulting in biased allocation to comparison groups, the unequal provision of care apart from the intervention under evaluation, the biased assessment of outcomes, and the inadequate handling of dropouts and losses to follow-up. Several studies¹ have recently documented these biases. For example, Schulz et al² demonstrated, for trials with binary outcomes included in meta-analyses from the Cochrane Pregnancy and Childbirth Database, that trials in which randomization was inadequately concealed yielded exaggerated estimates of treatment effect in comparison with trials that reported adequate concealment and found a similar (but smaller) overestimation of treatment effects for trials that were not adequately blinded.

Information on trial quality is important for peer review, when considering the results from individual trials and for the conduct of unbiased systematic reviews. The assessment of the methodologic quality of a trial is closely

Context The evaluation of the methodologic quality of randomized controlled trials (RCTs) is central to evidence-based health care. Important methodologic detail may, however, be omitted from published reports, and the quality of reporting is therefore often used as a proxy measure for methodologic quality. We examined the relationship between reporting quality and methodologic quality of published RCTs.

Methods Study of 60 reports of placebo-controlled trials published in English-language journals from 1985 to 1997. Reporting quality was measured using a 25-item scale based on the 1996 issue of the Consolidated Standards of Reporting Trials (CONSORT). Concealment of allocation, appropriate blinding, and analysis according to the intention-to-treat principle were indicators of methodologic quality. Methodologic quality was compared between groups of trials defined by reporting quality scores of low, intermediate, and high. Reporting quality scores were compared between groups defined by high and low methodologic quality.

Results Among 23 trials of low reporting quality (median score, 9 [range, 3.5-10.5]), allocation concealment was unclear for all but 1 trial, but there were 16 trials (70%) with adequate blinding and 9 trials (39%) that had been analyzed according to the intention-to-treat principle. Among 18 trials of high reporting quality (median score, 18 [range 16.5-22.0]), there were 8 trials (44%) with adequate allocation concealment, 16 trials (89%) with adequate blinding, and 13 trials (72%) analyzed according to the intention-to-treat principle. The median reporting score was 15.0 for the 33 trials that were analyzed according to intention-to-treat principle and 14.5 for the 14 trials with on-treatment analyses ($P=.67$).

Conclusions Similar quality of reporting may hide important differences in methodologic quality, and well-conducted trials may be reported badly. A clear distinction should be made between these 2 dimensions of the quality of RCTs.

JAMA. 2002;287:2801-2804

www.jama.com

intertwined with the quality of reporting, that is, the extent to which a report provides information about the design, conduct, and analysis of the trial. Trial reports often omit important methodologic details. A widely used approach to this problem consists in treating reporting quality as a proxy measure for methodologic quality. This could be justified if the assumption were correct that faulty reporting reflects faulty methods.³ The objective of our

study was to examine the relationship between the quality of reporting and the methodologic quality of RCTs.

Author Affiliations: Departments of Social and Preventive Medicine (Drs Huwiler-Müntener, Junker, and Egger) and Rheumatology and Clinical Immunology (Dr Jüni), University of Bern, Bern, Switzerland; and MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Bristol, England (Drs Jüni and Egger).

Corresponding Author and Reprints: Karin Huwiler-Müntener, MD, Department of Social and Preventive Medicine, Finkenhubelweg 11, CH-3012 Bern, Switzerland (e-mail: huwiler@ispm.unibe.ch).

METHODS

Selection of Trial Reports

This analysis is based on a sample of 60 reports of placebo-controlled trials assembled in the context of a study⁴ that examined the interrater reliability of quality scales. We aimed to include a wide range of journals, disciplines, and authors through the use of a database of controlled trials that had been assembled within the framework of a study of language bias.⁵ Described in detail elsewhere,⁶ the database included 89 reports of placebo-controlled trials published from 1985 to 1997 in 29 English-language specialist journals by authors based in German-speaking Europe. We searched MEDLINE to identify 89 additional reports of placebo-controlled trials published in the same journals around the same time by authors from English-speaking countries. Twenty trials were randomly selected from each of these 2 samples (computer-generated random numbers). Finally, we searched MEDLINE to identify another 89 placebo-

controlled trials that were published in 1997 in 1 of the 5 leading general medicine journals (*Annals of Internal Medicine*, *BMJ*, *JAMA*, *New England Journal of Medicine*, and *The Lancet*). We again selected 20 reports at random.

Assessment of Methodologic Quality and Reporting Quality

We used concealment of allocation, blinding, and analysis according to the intention-to-treat principle as indicators of methodologic quality and classified these indicators as adequate, inadequate, or unclear. For concealment, the following approaches to prevent foreknowledge of allocation were considered adequate: sequentially numbered, sealed, opaque envelopes; coded drug packs prepared by an independent pharmacy; and central randomization. Other methods, such as the use of an open random number table, were considered inadequate. For blinding, a trial was classified as adequate if it was described as "double-blind" or the blinding of both patients and outcome

assessors was mentioned. The analysis was considered adequate if all randomized patients were included in the analysis in the group they had been allocated to (intention-to-treat analysis). An on-treatment analysis was considered inadequate.

We developed a 30-item scale based on the 1996 Consolidated Standards of Reporting Trials (CONSORT) statement⁷ to assess reporting quality.⁴ For this analysis, we excluded 5 items that were closely related to our indicators of methodologic quality (TABLE 1). All questions were answered with yes or no, and each yes answer earned 1 point, for a maximum of 25 points.

Two observers (K.H.-M., C.J.) assessed reports. For each article, we calculated the average reporting quality score from the 2 independent assessments. Discrepancies in the assessment of methodologic quality were resolved by consensus. Interobserver agreement was high: the intraclass correlation coefficient for the summary score was 0.81 (95% confidence interval [CI], 0.69-0.88). The κ values for allocation concealment was 0.96 (95% CI, 0.70-1.00); for blinding, 0.94 (95% CI, 0.69-1.00); and for intention-to-treat analysis, 0.93 (95% CI, 0.68-1.00). We compared the methodologic quality among 3 similarly sized groups of trials defined by low, intermediate, and high reporting quality scores and reporting quality scores among groups defined by high and low methodologic quality. We calculated χ^2 tests and Kruskal-Wallis tests using Stata version 7 statistical software (Stata Corp, College Station, Tex).

RESULTS

The average quality of reporting was low (median score, 12.5). The distribution of scores was bimodal (FIGURE), which was explained by better reporting in the 5 general medicine journals compared with the specialist journals (median scores, 10.5 vs 17.0; $P = .0001$ by Kruskal-Wallis test). The median score was 9.0 (range, 3.50-10.5) for the low-reporting-quality group (23 trials), 12.5 (range, 11.5-16.0) for the in-

Table 1. Reporting Quality Scale Based on 1996 CONSORT Statement*

1. Does the title identify the study as a randomized controlled trial?
2. Is the abstract presented in a structured format?
3. Are the objectives stated?
4. Is the hypothesis stated?
5. Is the study population described?
6. Are inclusion and exclusion criteria described?
7. Are the interventions described?
8. Are the outcome measures described?
9. Is a primary outcome specified?
10. Is a minimum important difference for the primary outcome reported?
11. Are power calculations described?
12. Is the rationale for the statistical analyses explained?
13. Are the methods for statistical analyses described?
14. Are stopping rules described?
15. Is the unit of randomization described?
16. Is the method used to generate the allocation schedule described?
17. *Is the method of allocation concealment described?**
18. Is the timing of assignment described?
19. *Is the method to separate those generating the allocation sequence from those assigning participants to groups described?**
20. *Are the mechanisms of blinding described?**
21. Is the number of eligible patients reported?
22. *Is the number of randomized patients reported for each comparison group?**
23. Are prognostic variables by treatment and control group described?
24. Is the number of patients receiving intervention as allocated reported for each comparison group?
25. *Is the number of patients analyzed reported for each comparison group?**
26. Are withdrawals and dropouts described for each comparison group?
27. Are protocol deviations described for each comparison group?
28. Is the estimated effect of the intervention on primary and secondary outcomes stated, including a point estimate and measure of precision (confidence interval)?
29. Are the results stated in absolute numbers?
30. Are summary data and inferential statistics presented in sufficient detail to permit alternative analyses and replication?

*The 5 items of reporting quality in italics were not used in the study because they are related to the indicators of methodologic quality. CONSORT indicates Consolidated Standards of Reporting Trials.

intermediate-reporting-quality group (19 trials), and 18 (range, 16.5-22.0) for the high-reporting-quality group (18 trials). Associations between reporting quality and methodologic quality were evident for all 3 indicators of methodologic quality (TABLE 2), which is not surprising considering that there were many trials with unclear methodologic assessments. This was the case for allocation concealment in 45 trials (75%), for blinding in 12 trials (20%), and for intention-to-treat analysis in 13 trials (22%). The poor quality of reporting meant that for allocation concealment and blinding it was only possible to distinguish between trials with appropriate and unclear methods, whereas trials with appropriate intention-to-treat analyses, inappropriate on-treatment analyses, and unclear analyses were identified.

Although reporting quality is associated with methodologic quality, using reporting quality as a proxy measure for methodologic quality would result in misclassification of a considerable proportion of trials. For example, most trials of low reporting quality were adequately blinded (16 [70%]), and 9 (39%) were analyzed according to the intention-to-treat principle. On the other hand, 5 (28%) of the trials of high reporting quality had presented inappropriate on-treatment analyses rather than intention-to-treat analyses (Table 2). The quality of reporting was similar in trials analyzed appropriately and inappropriately: the median reporting score was 15.0 for the 33 trials that were analyzed according to the intention-to-treat principle and 14.5 for the 14 trials with on-treatment analyses ($P = .67$). Finally, even among trials of high reporting quality, most reports (56%) did not indicate whether allocation had been concealed.

COMMENT

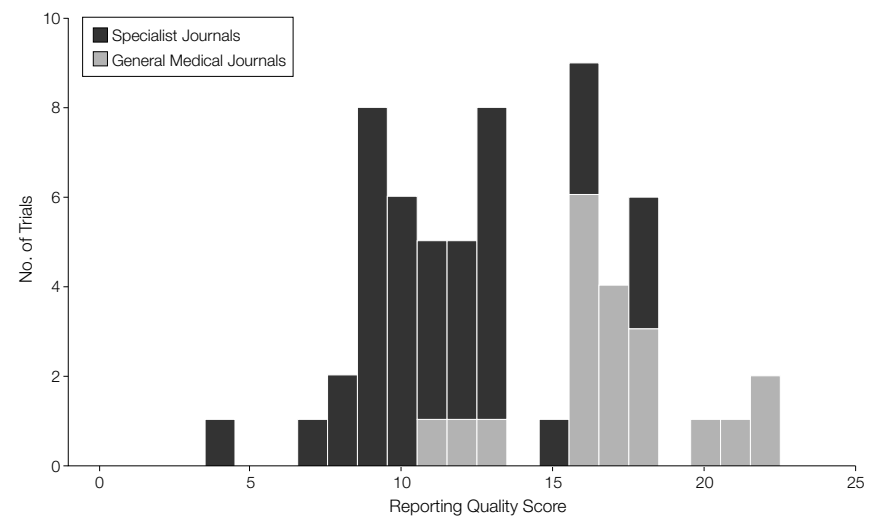
The evaluation of the methodologic quality of RCTs is central to the appraisal of individual trials and the conduct of unbiased systematic reviews.⁸ Assessments of methodologic quality depend on the quality of

reporting, and incomplete reporting is often interpreted as low methodologic quality. We attempted to disentangle the relationship between reporting quality and methodologic quality by using a scale based on the CONSORT statement and by separately assessing 3 central dimensions of methodologic quality. Our results document once again that trial reports frequently omit important methodologic detail: in most trials, it was unclear whether the allocation of participants had been concealed appropriately. The poor

reporting on concealment may be due to the fact that many trials were published before Schulz et al² demonstrated the importance of adequate concealment of allocation for prevention of bias in RCT research. The generally low quality of reporting in specialist journals must be of concern and underscores the importance of the CONSORT statement both for specialist and general medical journals.

As could be expected, the strength of the association between overall quality of reporting and methodologic

Figure. Histogram of Reporting Quality Scores



Specialist journals (include 40 reports from 29 specialist journals), and the general medicine journals include 20 reports from *Annals of Internal Medicine*, *BMJ*, *JAMA*, *New England Journal of Medicine*, and *The Lancet*.

Table 2. Distribution of 60 Placebo-Controlled Trials by Methodologic Quality and Reporting Quality

Methodologic Quality	Reporting Quality, No. (%)*			P Value†
	Low (n = 23)	Intermediate (n = 19)	High (n = 18)	
Allocation concealment				.01
Appropriate	1 (4)	6 (32)	8 (44)	
Unclear	22 (96)	13 (68)	10 (56)	
Blinding				.26
Appropriate	16 (70)	16 (84)	16 (89)	
Unclear	7 (30)	3 (16)	2 (11)	
Analysis				.003
Intention to treat	9 (39)	11 (58)	13 (72)	
On treatment	3 (13)	6 (32)	5 (28)	
Unclear	11 (48)	2 (10)	0 (0)	

*The median score from the reporting quality scale was 9.0 for the low reporting quality group, 12.5 for the intermediate group, and 18 for the high reporting quality group.

†Probabilities from χ^2 tests.

quality depended on the proportion of trials with unclear reporting on the 3 indicators of methodologic quality. The reporting score thus measured what it was supposed to measure; but is it also a good proxy measure for methodologic quality? Our results demonstrate that based on reporting quality alone, the true quality of a substantial proportion of well-conducted trials and of trials of low methodologic quality will be misjudged. Indeed, the quality of reporting of trials that were analyzed according to the intention-to-treat principle was not different from that of trials presenting on-treatment analyses only. The intention-to-treat approach is, of course, essential to maintain treatment groups that are similar apart from random variation, and this crucial feature may be lost if the analysis is not performed on the groups initially produced by the randomization process.⁹

Our study has a number of limitations. The trials analyzed were published several years ago and may no longer reflect current reporting prac-

tices. There is evidence that the quality of reporting has improved in journals that have adopted CONSORT.¹⁰ This would not, however, invalidate our findings on the relationship between reporting quality and methodologic quality. On the contrary, with increasing quality of reporting, scales based on reporting quality will become less accurate measures of methodologic quality. Our reporting scale was based on the 1996 CONSORT statement rather than the revised version published recently.¹¹ It is possible that a scale based on the 2001 version of CONSORT would measure reporting quality more precisely, but it is unlikely that it would be a better measure of methodologic quality. Finally, we assumed that blinding was appropriate if the authors stated that the trial was "double-blind." This term has since been shown to be ambiguous and should no longer be used.^{8,12}

In conclusion, reporting quality is associated with methodologic quality, but similar quality of reporting may hide important differences in methodologic quality and well-conducted tri-

als may be reported badly. A clear distinction should therefore be made between reporting and methodologic quality of trials. Scales that predominantly measure reporting quality, for example, the scales developed by Jadad et al¹³ or Chalmers et al,¹⁴ should not be used to measure methodologic quality. Rather, the important methodologic aspects should be identified a priori and assessed individually. This should generally include the key domains of concealment of treatment allocation, blinding of outcome assessment, and handling of attrition in the analysis. Finally, continued efforts are required to improve the quality of reporting of randomized trials.

Author Contributions: *Study concept and design:* Huwiler-Müntener, Jüni, Junker, Egger. *Acquisition of data:* Huwiler-Müntener, Junker. *Analysis and interpretation of data:* Jüni, Egger. *Drafting of the manuscript:* Egger. *Critical revision of the manuscript for important intellectual content:* Huwiler-Müntener, Jüni, Junker. *Statistical expertise:* Jüni, Egger. *Administrative, technical, or material support:* Huwiler-Müntener, Junker. *Study supervision:* Jüni, Junker, Egger.

REFERENCES

- Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42-46.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408-412.
- Liberati A, Himmel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol*. 1986;4:942-951.
- Huwiler-Müntener K. *Bestimmung der Zwischen-Beurteiler-Verlässlichkeit eines Instrumentes zur Qualitätsbeurteilung der Publikation von randomisierten, kontrollierten Studien [Assessing the Interrater-Reliability of a Quality Assessment Tool for the Reporting of Randomized Controlled Trials]* [MD thesis]. Bern, Switzerland: University of Bern; 2000.
- Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet*. 1997;350:326-329.
- Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA*. 1998;280:247-249.
- Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276:637-639.
- Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663-694.
- Hollis S, Campbell F. What is meant by intention to treat analysis? survey of published randomised controlled trials. *BMJ*. 1999;319:670-674.
- Moher D, Jones A, LePage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285:1992-1995.
- Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *JAMA*. 2001;285:1987-1991.
- Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285:2000-2003.
- Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1-12.
- Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials*. 1981;2:31-49.