Use of the CONSORT Statement and Quality of Reports of Randomized Trials: A Comparative Before-and-After Evaluation

David Moher; Alison Jones; Leah Lepage; et al.

JAMA. 2001;285(15):1992-1995 (doi:10.1001/jama.285.15.1992)

http://jama.ama-assn.org/cgi/content/full/285/15/1992

Use of the CONSORT Statement and Quality of Reports of Randomized Trials

A Comparative Before-and-After Evaluation

David Moher, MSc
Alison Jones, BA
Leah Lepage, PhD
for the CONSORT Group

Context The Consolidated Standards for Reporting of Trials (CONSORT) statement was developed to help improve the quality of reports of randomized controlled trials (RCTs). To date, a paucity of data exists regarding whether it has achieved this goal.

Objective To determine whether use of the CONSORT statement is associated with improvement in the quality of reports of RCTs.

Design and Setting Comparative before-and-after evaluation in which reports of RCTs published in 1994 (pre-CONSORT) were compared with RCT reports from the same journals published in 1998 (post-CONSORT). We included 211 reports from *BMJ*, *JAMA*, and *The Lancet* (journals that adopted CONSORT) as well as *The New England Journal of Medicine* (a journal that did not adopt CONSORT and was used as a comparator).

Main Outcome Measures Number of CONSORT items included in a report, frequency of unclear reporting of allocation concealment, and overall trial quality score based on the Jadad scale, a 5-point quality assessment instrument.

Results Compared with 1994, the number of CONSORT checklist items in reports of RCTs increased in all 4 journals in 1998, and this increase was statistically significant for the 3 adopter journals (pre-CONSORT, 23.4; mean change, 3.7; 95% confidence interval [CI], 2.1-5.3). The frequency of unclear reporting of allocation concealment decreased for each of the 4 journals, and this change was statistically significant for adopters (pre-CONSORT, 61%; mean change, –22%; 95% CI, –38% to –6%). Similarly, 3 of the 4 journals showed an improvement in the quality score for reports of RCTs, and this increase was statistically significant for adopter journals overall (pre-CONSORT, 2.7; mean change, 0.4; 95% CI, 0.1-0.8).

Conclusion Use of the CONSORT statement is associated with improvements in the quality of reports of RCTs.

JAMA. 2001;285:1992-1995 www.jama.com

REPORTS OF RANDOMIZED CONtrolled trials (RCTs) are the "gold standard" by which health care professionals and others make decisions about treatment effectiveness. To assess the strengths and limitations of RCTs, readers need and deserve to know the quality of their methods. Previous studies[1,2] indicate that reports of low-quality RCTs, compared with reports of higher-quality ones, overestimate the effectiveness of interventions by about 30% across a variety of health care conditions.

The Consolidated Standards for Reporting of Trials (CONSORT) Group developed the CONSORT statement,[3] an evidence-based approach to help improve the quality of reports of RCTs. Since its publication in 1996, the CONSORT statement has been widely supported,[4-6] has been translated into several languages, and has an Internet presence (http://www.consort-statement.org) to facilitate awareness and dissemination. However, there is a paucity of data[7] regarding whether the CONSORT statement has improved the quality of reports of RCTs.

See also pp 1987, 1996, 2000, and 2006.

METHODS

We compared a sample of RCTs published in 1994 (pre-CONSORT) to a sample published in 1998 (post-CONSORT). To be included, RCT reports had to be drawn from journals classified by the *Science Citation Index* as "general and/or internal medicine." Four journals were evaluated: *BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine* (*NEJM*). Three of these journals (*BMJ*, *JAMA*, *Lancet*) were early adopters of the CONSORT statement,[4-6] whereas the fourth (*NEJM*) did not formally adopt the CONSORT statement and was used as a comparator.

Author Affiliations: University of Ottawa (Mr Moher) and Thomas C. Chalmers Centre for Systematic Reviews (Mr Moher, Ms Jones, and Dr Lepage), Ottawa, Ontario.
A list of members of the CONSORT Group appears at the end of this article.
Corresponding Author and Reprints: Leah Lepage, PhD, Thomas C. Chalmers Centre for Systematic Reviews, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Ontario, K1H 8L1 Canada (e-mail: llepage@uottawa.ca).

Reports were included only if they involved human subjects and if the study design was identified as an RCT by examining the title and the abstract. Two readers (A.J., L.L.) independently handsearched the 4 journals for RCTs published between January 1 and June 30 of both 1994 and 1998. Hard copies of relevant articles were obtained but were not masked because evidence concerning the effect of masking on assessments of trial quality is inconsistent.[8,9]

Three measures were used to assess the quality of reports of RCTs. First, the CONSORT checklist was modified so that multiple items were listed separately, which resulted in 40 items. Each item was assigned a yes or no response depending on whether the authors had reported it. Second, the reporting of allocation concealment was assessed as adequate, inadequate, or unclear.[2] Third, the Jadad scale,[9] which contains 2 questions for randomization and masking and 1 question evaluating the reporting of withdrawals and dropouts, was used to assess quality. Each question entails a yes or no response option. In total, 5 points can be awarded, with higher scores indicating superior quality. Two reviewers (A.J., L.L.) completed all of these evaluations.

Both reviewers underwent training in evaluating RCTs using the CONSORT checklist. Before training, the definition of each checklist item was discussed. To assess interobserver agreement, 5 items from the checklist were purposefully selected (inclusion criteria, exclusion criteria, point estimate, deviation from protocol, and general interpretation of study findings). A κ statistic was calculated for each item based on a randomly selected set of 10 RCTs, from 1994 and 1998, and these were not included in this study. Discrepancies were resolved by consensus involving a third party (D.M.). A similar approach was used to assess interobserver agreement in assessing reporting of allocation concealment and using the Jadad scale.

Taking time (1994-1998) into consideration and using either unpaired *t* tests or χ^2 tests, we compared the number of checklist criteria included in each report and the mean number of criteria included within each subheading specified in the CONSORT checklist. We also assessed the percentage of studies that reported unclear allocation concealment and the specific item and overall quality score derived from the Jadad scale. The number of CONSORT checklist items reported was treated as the dependent variable in an analysis of variance (ANOVA) with 2 factors: journal and year. The difference between the adopter journals and the comparator journal was derived from the fitted ANOVA using a contrast involving journal (ie, adopters vs comparator) and year (ie, 1998 vs 1994). Least-square estimates and 95% confidence intervals (CIs) also were computed; for significant level, $P = .05$ (2-sided).

RESULTS

Of the 221 RCTs identified, 211 met the inclusion criteria. Six studies were excluded because they were animal stud-

Table 1. CONSORT Checklist Criteria Included in Reports of Randomized Trials for Articles Published in *BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine* (*NEJM*) During the First Half of 1994 and 1998*

	BMJ	*JAMA*	*Lancet*	Total Adopters	*NEJM* (Comparator)
Total No. of items					
1994	14	29	28	**71**	26
1998	20	20	37	**77**	37
Title (n = 1)					
1994, %	21	34	43	35	12
1998, % Change (95% CI)	74 (47 to 100)	31 (2 to 59)†	49 (28 to 70)†	51 (37 to 64)†	–12 (–25 to 2)
Abstract (n = 1)					
1994, %	100	100	0	61	100
1998, % Change (95% CI)	–15 (–32 to 2)	0	100†	36 (23 to 48)†	0
Introduction (n = 3)					
1994 Mean (SD)	1.1 (0.4)	1.3 (0.6)	1.1 (0.7)	1.2 (0.6)	1.3 (0.8)
1998, Change (95% CI)	0.3 (–0.03 to 0.6)	0.1 (–0.3 to 0.4)	0.2 (–0.2 to 0.6)	0.2 (–0.1 to 0.4)	0.02 (–0.4 to 0.4)
Methods (n = 22)					
1994, Mean (SD)	10.4 (2.6)	13.7 (3.6)	11.0 (3.2)	12.0 (3.5)	10.2 (2.2)
1998, Change (95% CI)	3.8 (1.4 to 6.3)†	0.9 (–1.2 to 2.9)	1.7 (–0.1 to 3.4)	1.6 (0.4 to 2.8)†	0.5 (–0.9 to 2.0)
Results (n = 10)					
1994, Mean (SD)	6.4 (1.6)	7.5 (1.2)	6.5 (1.5)	6.9 (1.5)	6.5 (1.1)
1998, Change (95% CI)	1.5 (0.5 to 2.5)†	0.2 (–0.5 to 0.9)	1.6 (0.8 to 2.4)	1.0 (0.6 to 1.5)†	0.5 (–0.03 to 1.1)
Comments (n = 3)					
1994, Mean (SD)	2.0 (0.4)	2.1 (0.6)	2.9 (0.3)	2.4 (0.6)	2.8 (0.4)
1998, Change (95% CI)	0.2 (–0.3 to 0.6)	0.1 (–0.2 to 0.4)	–0.1 (–0.3 to 0.1)	0.1 (–0.1 to 0.3)	–0.2 (–0.4 to 0.1)
Total (n = 40)					
1994, Mean (SD)	21.1 (4.2)	26.0 (4.6)	21.8 (4.8)	23.4 (5.1)	22.0 (3.0)
1998, Change (95% CI)	6.4 (2.9 to 9.9)‡	1.6 (–0.8 to 4.0)	4.9 (2.5 to 7.3)‡	3.7 (2.1 to 5.3)‡	0.8 (–1.1 to 2.7)

*CONSORT indicates Consolidated Standards for Reporting of Trials; CI, confidence interval.
†$P<.05$ (2-sided).
‡$P<.001$ (2-sided).

ies and 4 studies because they were quasi-randomized trials.

Substantial agreement was established for 4 items from the CONSORT checklist and the Jadad instrument (inclusion and exclusion criteria, point estimate and general interpretation [$\kappa = 1.0$], quality assessment, and overall Jadad score [$\kappa = 0.74$]). Moderate agreement was established for allocation concealment ($\kappa = 0.53$) and for 1 item from the CONSORT checklist ($\kappa = 0.54$, deviation from protocol).

There was an increase over time in the number of CONSORT checklist items included in the reports of RCTs in all 4 journals (TABLE 1). This increase was statistically significant for 2 individual journals and overall for adopter journals (pre-CONSORT, 23.4; mean change, 3.7; 95% CI, 2.1-5.3). Over time, the increase in the reporting of CONSORT items was significantly greater for adopter journals when evaluated against the comparator journal (mean difference, 3.8; 95% CI, 1.0-6.5; 2-sided $P = .007$).

The proportion of RCTs with unclear reporting of allocation concealment decreased over time in all 4 journals (TABLE 2) and was statistically significant for adopter journals (pre-CONSORT, 61%; mean change, –22%; 95% CI, –38% to –6%). Over time, 3 of the 4 journals improved the quality of reports of RCTs as assessed by the Jadad scale (Table 2), which was statistically significant for 1 journal (*Lancet*) and across the adopter journals (pre-CONSORT, 2.7; mean change, 0.4; 95% CI, 0.1-0.8).

COMMENT

The quality of reports of RCTs in all 4 journals included in this study improved over time. This improvement appears to be greater for the journals that adopted CONSORT. However, because of our study design, it is only possible to suggest that the improvement may be associated with the implementation of the CONSORT statement.

These results also suggest that aspects of reporting of RCTs still require improvement. For example, the reporting of bias reduction methods, such as masking, is less than optimal. Similarly, our results confirm a concern raised by others[10,11] regarding how the discussion/comment sections of RCTs are reported. It is unclear whether these deficiencies reflect difficulties in using CONSORT experienced by authors, by journals, or by both. To help address these questions, it will be important to obtain data from editors and authors as well as data on the readability of CONSORT reports as a way to gauge their scientific content.

We used 1 journal with a high citation impact factor as the comparator. This approach offers some control over more obvious forms of bias, such as the passage of time, and is considered to be a stronger research design than having no comparator.[12] Ideally, we would have liked to include more comparator journals, but we were unable to identify them. Because of this, we limited our analysis in all cases except 1 to a comparison over time rather than between adopter and nonadopter journals. To strengthen these findings, we recommend that this evaluation be replicated and expanded to include more nonadopter journals. To facilitate such a study, we encourage all journals to indicate, perhaps in their information for authors, whether they support the CONSORT statement. Moreover, to increase generalizability of these results, future evaluations should also include specialty journals.[13]

Another limitation of our study is the time frame in which we completed the evaluation. We chose reports of RCTs published during the first half of 1998, only 12 to 18 months after the endorsement of CONSORT by journals included in this evaluation. It is possible, even likely, that effective dissemination is a slow process and that to estimate the true influence of CONSORT requires more time. In addition, our results pertain to the CONSORT checklist and do not evaluate the use of the CONSORT flow diagram. There are limitations to the present version of the flow diagram[14] in terms of information requested of authors. Additionally, interpretation of the terms used and the sensibility of

Table 2. Quality of Reports of Randomized Trials, Using an Assessment Tool, for Articles Published in *BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine* (*NEJM*) During the First Half of 1994 and 1998*

Journal	Total No. of Items		Randomization		Double-blinding		Dropouts/ Withdrawals		Total		Unclear Allocation Concealment	
	1994	1998	1994, Mean (SD)	1998, Change (95% CI)	1994, Mean (SD)	1998, Change (95% CI)	1994, %	1998, % Change (95% CI)	1994, Mean (SD)	1998, Change (95% CI)	1994, %	1998, % Change (95% CI)
BMJ	14	20	1.1 (0.4)	0.4 (0.04 to 0.8)†	0.2 (0.6)	0.1 (–0.4 to 0.5)	71	–6 (–40 to 28)	2.1 (0.9)	0.4 (–0.3 to 1.2)	79	–29 (–62 to 4)
JAMA	29	20	1.3 (0.6)	0.1 (–0.3 to 0.4)	0.9 (0.8)	0.2 (–0.3 to 0.8)	76	4 (–21 to 29)	3.0 (1.0)	0.4 (–0.3 to 1.0)	59	–14 (–43 to 16)
Lancet	28	37	1.2 (0.4)	0.4 (0.1 to 0.6)†	0.6 (0.8)	0.3 (–0.2 to 0.7)	96	1 (–8 to 10)	2.8 (0.9)	0.7 (0.1 to 1.2)‡	54	–24 (–48 to 1)
Total Adopters	**71**	**77**	1.2 (0.5)	0.3 (0.1 to 0.4)†	0.6 (0.8)	0.2 (–0.1 to 0.4)	83	1 (–11 to 13)	2.7 (1.0)	0.4 (0.1 to 0.8)§	61	–22 (–38 to –6)‖
NEJM comparator	26	37	1.4 (0.5)	0.02 (–0.2 to 0.3)	0.8 (1.0)	0.3 (–0.4 to 0.5)	92	–6 (–21 to 10)	3.1 (1.0)	–0.01 (–0.6 to 0.5)	69	–8 (–33 to 17)

*CI indicates confidence interval.
†$P < .05$ (2-sided).
‡$P = .01$ (2-sided).
§$P = .02$ (2-sided).
‖$P = .008$ (2-sided).

some of the criteria across RCTs has been inconsistent. Egger and colleagues[15] report the results of an evaluation of the CONSORT statement flow diagram in an accompanying article.

In summary, these findings suggest that use of the CONSORT checklist may be associated with improving the quality of reports of RCTs. Higher-quality reports are likely to improve RCT interpretation, minimize biased conclusions, and ultimately facilitate decision making about treatment effectiveness.

Author Contributions: *Study concept and design*: Moher.
Acquisition of data: Moher, Jones, LePage.
Analysis and interpretation of data: Moher, Jones.
Drafting of the manuscript: Moher, Jones.
Critical revision of the manuscript for important intellectual content: Moher, LePage.
Statistical expertise: Moher.
Administrative,technical, or material support: Jones, LePage.
Study supervision: Moher.
David Moher is the guarantor of the article
Contributors to the Article: Members of the CONSORT Group who participated in the Ottawa meeting were Douglas G. Altman, DSc, Centre for Statistics in Medicine, Institute of Health Sciences (Oxford, England); Jesse Berlin, PhD, University of Pennsylvania (Philadelphia, Pa); Frank Davidoff, MD, *Annals of Internal Medicine* (Philadelphia, Pa); Matthias Egger, MD, University of Bristol (Bristol, England); Diana Elbourne, PhD, London School of Hygiene and Tropical Medicine (London, England); Peter Gøtzsche, MD (Nordic Cochrane Centre, Copenhagen, Denmark); Sylvan Green, MD, PhD, School of Medicine, Case Western Reserve University (Cleveland, Ohio); Richard Horton, MB, *The Lancet* (London, England); Terry Klassen, MD, University of Alberta (Edmonton, Alberta); Tom Lang, MA, Tom Lang Communications (Lakewood, Ohio); Curt Meinert, PhD, Johns Hopkins University (Baltimore, Md); David Moher, MSc, Thomas C. Chalmers Centre for Systematic Reviews (Ottawa, Ontario); Drummond Rennie, MD, *JAMA* (Chicago, Ill); Ken Schulz, PhD, Family Health International (Research Triangle Park, NC); and Bruce Squires, MD, World Association of Medical Editors (Ottawa, Ontario).
Funding/Support: Financial support was provided in part by the Jefferson Smurfit Foundation, Dublin, Ireland; National Library of Medicine; Merck; Glaxo-Wellcome; and the Canadian Institutes for Health Research.
Acknowledgment: We thank all members of the following groups: Standards for Reporting Trials (SORT), Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, and the CONSORT Group for their support and encouragement in helping to bring the CONSORT statement to its present form. We also thank Natasha Wiebe, MMath, and Ba' Pham, MMath, for helping with the data analysis, and Howard M. Schachter, PhD, for commenting on early versions of the manuscript.

REFERENCES

1. Moher D, Pham B, Jones A, et al. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1999;352:609-613.
2. 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.
3. Begg CB, Cho MK, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA*. 1996;276:637-639.
4. Rennie D. How to report randomized controlled trials: the CONSORT statement. *JAMA*. 1996;274:649.
5. Altman DG. Better reporting of randomized controlled trials: the CONSORT statement. *BMJ*. 1996;313:570-571.
6. McNamee D, Horton R. Lies, damn lies, and reports of RCTs. *Lancet*. 1996;348:562.
7. Lebeau DL, Steinmann WC, Patel KV. *Has the Randomized Controlled Trial Literature Improved After CONSORT? 50 Years of Clinical Trials: Past, Present and Future*. London, England: BMJ Publishing Group; 1998.
8. Berlin JA, for the University of Pennsylvania Meta-analysis Blinding Study Group. Does blinding of readers affect the results of meta-analyses? *Lancet*. 1997;350:185-186.
9. 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.
10. Clarke MJ, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA*. 1998;280:280-282.
11. Docherty M, Smith R. The case for structuring the discussion of scientific papers. *BMJ*. 1999;318:1224-1225.
12. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Boston, Mass: Houghton Mifflin Co; 1966:13-23.
13. Prusakowski N, Woo C, Hawkins BS. CONSORT and its effect on reports of clinical trials in ophthalmology journals [abstract]. *Control Clin Trials*. 2000;21:65S.
14. Meinert CL. Beyond CONSORT: need for improved reporting standards for clinical trials. *JAMA*. 1998;279:1487-1489.
15. Egger M, Jüni P, Bartlett C, for the CONSORT group. Value of flow diagrams in reports of randomized controlled trials. *JAMA*. 2001;285:1996-1999.