

Standards of Reporting of Randomized Controlled Trials in General Surgery

Can We Do Better?

Sabapathy P. Balasubramanian, FRCS Ed,* Martin Wiener, MB, ChB,* Zeiad Alshameeri,* Ravindranath Tiruvoipati, FRCS Ed,† Diana Elbourne, PhD,‡ and Malcolm W. Reed, MD*

Objective: To evaluate the quality of reporting of surgical randomized controlled trials published in surgical and general medical journals using Jadad score, allocation concealment, and adherence to CONSORT guidelines and to identify factors associated with good quality.

Summary Background Data: Randomized controlled trials (RCTs) provide the best evidence about the relative effectiveness of different interventions. Improper methodology and reporting of RCTs can lead to erroneous conclusions about treatment effects, which may mislead decision-making in health care at all levels.

Methods: Information was obtained on RCTs published in 6 general surgical and 4 general medical journals in the year 2003. The quality of reporting of RCTs was assessed under masked conditions using allocation concealment, Jadad score, and a CONSORT checklist devised for the purpose.

Results: Of the 69 RCTs analyzed, only 37.7% had a Jadad score of ≥ 3 , and only 13% of the trials clearly explained allocation concealment. The modified CONSORT score of surgical trials reported in medical journals was significantly higher than those reported in surgical journals (Mann-Whitney U test, $P < 0.001$). Overall, the modified CONSORT score was higher in studies with higher author numbers ($P = 0.03$), multicenter studies ($P = 0.002$), and studies with a declared funding source ($P = 0.022$).

Conclusion: The overall quality of reporting of surgical RCTs was suboptimal. There is a need for improving awareness of the CONSORT statement among authors, reviewers, and editors of surgical journals and better quality control measures for trial reporting and methodology.

(*Ann Surg* 2006;244: 663–667)

Well-designed and properly executed randomized controlled trials (RCTs) provide the best evidence about the relative effectiveness of different interventions.¹ The

results of these trials can have a profound and immediate impact on patient care. Trials conducted with inadequate methodological approaches are associated with exaggerated treatment effects.^{2–4} Inadequate methodology may be due to several reasons: subversion of randomization resulting in biased allocation to comparison groups; unequal provision of care apart from the intervention under evaluation; biased assessment of outcomes and inadequate handling of dropouts and losses to follow-up.⁵ Results of such biased studies can mislead decision making in health care at all levels, from treatment decisions for an individual patient to formulation of national public health policies. There is strong evidence to indicate that the quality of reporting of RCTs in the medical literature is less than optimal.^{6,7} To overcome such problems, the Consolidated Standards for reporting of Trials (CONSORT) Group developed the CONSORT statement⁸ in 1996, which was followed by a revised version in 2001.⁷

At least 25 scales have been described for the evaluation of trial quality,⁹ among which the Jadad score is commonly used and has been validated using established methodologic procedures.¹⁰ Although several studies have evaluated the quality of RCTs published in medical journals,^{10–12} to date there have been none directed at general surgical literature published in general surgery journals.

The aim of this study was to assess the quality of reporting of published general surgical RCTs using a checklist based on the CONSORT statement (providing a modified CONSORT score), the Jadad score, and the criteria of allocation concealment alone. We aimed to evaluate differences in reporting quality between studies published in surgical and general medical journals. We also determined whether specific study characteristics such as the number of authors, single or multicenter, statistician/epidemiologist involvement, and funding source were associated with reporting quality.

METHODS

RCTs published during the year 2003 in 6 of the top 10 impact factor surgery journals [*Annals of Surgery* (AS), *British Journal of Surgery* (BJS), *World Journal of Surgery* (WJS), *Journal of Surgery* (JS), *Journal of American College of Surgeons* (JACS), and *American Journal of Surgery* (AJS)] and the 4 major general medical journals [*British Medical Journal* (BMJ), *Lancet*, *Journal of American Medical Association*

From the *Academic Unit of Surgical Oncology, University of Sheffield; †Department of ECMO, Glenfield Hospital, Leicester; and ‡London School of Hygiene and Tropical Medicine, University of London, London, UK.

Reprints: Sabapathy P. Balasubramanian, FRCS Ed, Academic Unit of Surgical Oncology, K Floor, Royal Hallamshire Hospital, University of Sheffield, S10 2JF, UK. E-mail: s.p.balasubramanian@sheffield.ac.uk.

Copyright © 2006 by Lippincott Williams & Wilkins

ISSN: 0003-4932/06/24405-0663

DOI: 10.1097/01.sla.0000217640.11224.05

(JAMA), and *New England Journal of Medicine* (NEJM)] were retrieved for assessment of the quality of reporting. The search strategy involved screening the titles and abstracts of all the manuscripts published in the above journals in 2003 for the key words “randomized controlled trial.” Studies were only included if they were truly randomized, involved human subjects, reported predominantly on clinical outcomes and were on “surgical” topics. Full reports were obtained for those studies that appeared to meet the criteria or where there was insufficient information in the title and abstract.

The citation impact factor for 2002 obtained from journal citation reports was used to help in choosing the above journals for the study. From the top 10 journals, we chose 6 for which the full text was available at the University of Sheffield library. The “instructions to authors” section of all the journals was accessed to find out whether the journals endorsed the CONSORT statement.

Specific descriptive characteristics for each article such as the number of authors, number of centers involved, involvement of a statistician/epidemiologist, source of funding (if any), and country of study were noted by one observer (Z.A.), who then masked the manuscripts for details such as journal, authors, and institution. Part of the methodology used to assess the quality of the RCT reports was similar to that described by Moher et al.¹¹ The reporting of allocation concealment was assessed as adequate, inadequate, or unclear as previously described.² The Jadad scale was used, which contains 2 questions each for randomization and masking and 1 question evaluating the reporting of withdrawals and dropouts. Each question entailed a yes or no response option. In total, 5 points were awarded, with a score of 3 or more indicating superior quality.¹⁰ Allocation concealment and Jadad score were assessed by one observer (S.P.B.). A checklist of 30 items concerning reporting and/or methodology was then prepared from the revised CONSORT guidelines¹ published in 2001. The score for each item ranged from 1 (corresponding to no descrip-

tion of the item) to 3 (corresponding to adequate description and methodology). The CONSORT guidelines were studied and the definitions of each checklist item were discussed by the reviewers in detail. Each article was then assessed for every item on the checklist and scored independently by 2 observers (S.P.B. and R.T.), who also later arrived at a consensus score. The scores for the 30 items were added and a percentage score was calculated for each trial (as some items were non applicable and did not merit any score).

The data were collected on an excel spreadsheet and exported to SPSS (version 12.0 for Windows) for analysis. Initial presentations were descriptive, which included the median (interquartile range) Jadad and modified CONSORT scores of the individual journals and the specific characteristics of the studies. Comparisons were then made between the quality of the reports in medical and surgical journals, and the association between study characteristics (number of authors, single or multicenter, statistician/epidemiologist involvement, and funding source) and the modified CONSORT score (measured as greater or lesser than the median) was studied. Non-parametric tests were used for all analyses.

RESULTS

Of the 87 manuscripts that were initially retrieved from the 10 different journals (6 surgical and 4 general medical), 69 were eligible for inclusion in the study. The reasons for exclusion included previous reporting of study methodology/results (6), nonsurgical topics (5), not truly randomized (3), cluster randomized trials (1), trials with predominantly nonclinical outcomes (1), and studies reporting trial design only (1).

Table 1 shows the different journals along with their impact factor for 2002, number of articles analyzed, endorsement of the CONSORT statement in their “instructions to authors,” median Jadad score, and median modified CONSORT score calculated as percentage. The agreement between the pair of ob-

TABLE 1. Characteristics of RCTs Meeting Study Inclusion Criteria and Published in the Different Medical and Surgical Journals

Journal	Impact Factor in 2002	Articles	CONSORT Endorsement	Jadad Score Median (IQR)	Median Modified CONSORT Score (IQR)
Surgical Journals					
<i>American Journal of Surgery</i>	1.758	4	No	2 (2–3.5)	63.13 (61.95–69.38)
<i>Annals of Surgery</i>	6.073	20	No*	2 (2–3)	70.24 (66.98–74.35)
<i>British Journal of Surgery</i>	3.444	20	No	2 (1–2.75)	66.67 (61.59–73.81)
<i>Journal of American College of Surgeons</i>	2.369	7	Yes	2 (2–4)	72.41 (70.24–78.16)
<i>Surgery</i>	2.631	4	No	2 (2–2.75)	66.01 (62.35–75.88)
<i>World Journal of Surgery</i>	1.777	6	No	2.5 (1.75–4)	66.67 (63.64–67.82)
Total		61		2 (2–3)	68.97 (62.89–73.11)
Medical Journals					
<i>British Medical Journal</i>	7.585	0	Yes	—	—
<i>Journal of American Medical Association</i>	16.8	0	Yes	—	—
<i>Lancet</i>	15.397	4	Yes	3 (2.25–4.5)	85.81 (81.25–88.14)
<i>New England Journal of Medicine</i>	31.736	4	No*	3.5 (2.25–4.75)	85.12 (73.94–85.95)
Total		8		3 (2.25–4.75)	85.45 (81.09–86.13)

*Endorsed after 2003.

IQR indicates inter quartile range.

TABLE 2. Characteristics of the 69 Surgical Randomized Controlled Trials Meeting the Inclusion Criteria

Characteristic Assessed	Subgroup	No. (%)
Authors	<6	19 (27.5)
	6–10	39 (56.5)
	>10	11 (15.9)
Statistician/epidemiologist involvement	No mention	45 (65.2)
	Involvement acknowledged	9 (13)
	Involved as coauthor	15 (21.7)
No. of centers	Single center	42 (60.9)
	Multicenter	27 (39.1)
Funding source	No mention	35 (50.7)
	Commercial	3 (4.3)
	Public sector	21 (30.4)
	Mixed	10 (14.5)
Country	United States	10 (14.5)
	Europe	40 (58)
	Others	14 (20.3)
	Spanning continents	5 (7.2)

servers who independently assessed the RCTs using the CONSORT checklist was good (ICC = 0.85; 95% CI = 0.77–0.91; $P < 0.001$). Where there were initial disagreements, these were resolved and the consensus score was used for all further analyses. The 3 quality assessment tools used in this study (Jadad, allocation concealment, and the modified CONSORT score) correlated moderately with each other (Jadad and AC: Spearman's rho = 0.63; $P < 0.001$; Jadad and modified CONSORT: Spearman's rho = 0.66; $P < 0.001$; and AC and modified CONSORT: Spearman's rho = 0.68; $P < 0.001$).

There was a good correlation between impact factor of the journals studied and median modified CONSORT score of the articles published in each of the journals (Spearman's rho = 0.75; $P = 0.031$). Table 2 shows the study characteristics (including the number of authors, involvement of statistician/epidemiologist, single or multicenter study, and funding source) of all the RCTs included in the analysis. Table 3 shows the quality as assessed by Jadad score, allocation concealment, and the CONSORT checklist for the RCTs published in the surgical and medical journals separately. Table 4 shows the distribution of scores on all the items on the CONSORT checklist.

Univariate analysis was carried out to determine the association of reporting quality (measured as greater or lesser than the median modified CONSORT score of 70) with author numbers, statistician/epidemiologist involvement, number of centers, and funding source. It was found that studies with higher author numbers (Mann-Whitney U test; $z = -2.164$; $P = 0.03$), multicenter studies (χ^2 test with Yates correction; $\chi^2 = 10.029$; $P = 0.002$) and studies with a declared funding source (χ^2 test with Yates correction; $\chi^2 = 5.267$; $P = 0.022$) were of significantly better quality. Involvement of a statistician/epidemiologist in the study also tended to be associated with better quality (χ^2 test with Yates correction; $\chi^2 = 2.271$; $P = 0.132$).

DISCUSSION

“Surgical research or comic opera?” queried a *Lancet* editorial in 1996,¹³ stimulating a heated debate and serving to highlight the lack of RCTs in surgery and the limitations and difficulties of conducting one. Reports of RCTs should ideally convey relevant information to enable the reader to make an informed and a justified judgment concerning the validity of the trial and the effectiveness of the treatment.⁸ Furthermore, assessing the validity of the primary studies has been defined as one of the most important steps of the peer-review process.¹⁴

The need to improve the quality of reporting of RCTs has been highlighted in several specialties across health-care.^{6,11,15,16} We performed an exploratory survey of the quality of reporting of surgical RCTs in the surgical literature and assessed similar studies published in general medical literature, which served as a comparator. We found that the quality of reporting of general surgical RCTs leaves considerable room for improvement. The quality of reporting in surgical journals was clearly inferior to the quality of reporting of surgical trials in medical journals as assessed by allocation concealment, Jadad score, and the modified CONSORT score (Table 3). We acknowledge that the impact factors of the medical journals evaluated are greater than the surgical ones (Table 2) and would therefore attract good quality trials. We have not hypothesized that the quality of trials in these 2 groups of journals would be the same, but have only used the trials in medical journals as a standard with which to compare the surgical ones. Of trials reported in medical and surgical journals, patient blinding was not feasible in 50% and 27.9%, respectively, and assessor blinding was not feasible in 12.5% and 9.8%, respectively. That

TABLE 3. Quality Assessment of Reporting of Surgical RCTs Published in the General Surgical and General Medical Journals

Assessment Criterion	Subgroup	Medical Journals (n = 8)	Surgical Journals (n = 61)	P
Jadad score	Low (<3)	2 (25%)	41 (67%)	0.046*†
	High (≥3)	6 (75%)	20 (33%)	
Allocation concealment as assessed by Schulz et al ²	Unclear/inadequate	4 (50%)	50 (82%)	0.06†
	Adequate	4 (50%)	11 (18%)	
Modified CONSORT Score	Median (interquartile range)	85.45 (81.09–86.13)	68.97 (62.89–73.11)	<0.001**

*Significant ($P < 0.05$).

†Fisher exact test.

**Mann-Whitney U test.

TABLE 4. Distribution of Consensus Scores for All the Items in the Modified CONSORT Checklist

Item	n*	No Description (%)	Inadequate (%)	Adequate (%)
Justification for the trial	69	0	29	71
Explicit definition of eligibility criteria	69	0	15.9	84.1
Detailed description of settings/location of recruitment and data collection	69	2.9	68.1	29.0
Details of intervention studied	69	0	44.9	55.1
Clear statement of hypothesis or objectives	69	0	2.9	97.1
Identification and definition of outcome measures	69	0	31.9	68.1
Description of pre-study sample size calculation	69	43.5	26.1	30.4
Method of generation of random sequence	69	58.0	11.6	30.4
Details of any restriction used in randomization	68	54.4	10.3	35.3
Description of allocation concealment	69	68.1	18.8	13.0
Details of personnel involved in sequence allocation, enrollment, and assignment	69	66.7	27.5	5.8
Details of blinding of subjects	48	66.7	4.2	29.2
Details of blinding of treatment providers	24	41.7	12.5	45.8
Details of blinding of assessors	62	59.7	9.7	30.6
Details of blinding of analysts	69	100	0	0
Details of measurement of success of blinding	69	98.6	0	1.4
Description of statistical methods	69	1.4 (inadequate)	63.8 (minor errors)	34.8 (no obvious errors)
Flow chart describing patient numbers at different stages	69	75.4	4.3	20.3
Clear description of protocol deviations	48	6.3	12.5	81.3
Description of dates of recruitment	69	18.8	4.3	76.8
Details of follow-up	45	2.2	17.8	80
Description of baseline characteristics	69	1.4	20.3	78.3
Use of intention-to-treat principle	69	20.3	10.1	69.6
Complete reporting of results	69	0	10.1	89.9
Use of confidence intervals	69	63.8	21.7	14.5
Multiple testing and corrections	48	95.8	0	4.2
Description of side effects/adverse events	60	3.3	16.7	80
Trial limitations and weaknesses	69	31.9	52.2	15.9
External validity of trial results	69	4.3	65.2	30.4
Literature review	69	0	73.9	26.1

*The number is less than 69 for certain items, as there were trials in which certain items were not applicable or not relevant.

blinding was not feasible in a greater percentage of the trials in medical journals indicates that difficulty in blinding cannot explain the low scores in the trials in surgical journals. Overall, high author numbers, multicenter studies, and declaration of funding source were factors found to be significantly associated with better reporting quality. This finding of significant associations of some study characteristics with better quality may well be spurious as unknown confounding factors (such as size of any available grants) could be responsible. There is evidence that inadequate reporting and/or methodology of key aspects affect the quality and usefulness of such studies. The majority of studies (58%) did not state the method of randomization (Table 4) and allocation concealment was adequately explained in only 13% of the studies. Proper randomization eliminates selection bias and is a crucial component of high-quality RCTs.¹⁷ Successful randomization hinges on 2 steps: generation of an unpredictable allocation sequence and concealment of this sequence from the investigator enrolling the participant (allocation concealment).⁹ The latter helps to prevent selection bias, protects the

randomization sequence before and until the interventions are given to study participants, and can always be implemented. Trials that have not reported adequate allocation concealment have been found to be associated with exaggerated treatment effects.² Potential difficulties associated with the application of RCTs to surgical problems include the difficulty in successfully blinding patients, investigators, and assessors, the variability of surgical techniques and operator skills, and the "learning curve," which influences the efficacy of many interventions under study. Blinding of patients and assessors was not feasible in 30.4% and 10.1%, respectively, of all trials evaluated in this study. Of the trials in which blinding was possible, patient and assessor blinding were clearly reported only in 29.2% and 30.6%, respectively. Studies have shown that nonrandomized trials and RCTs that do not incorporate blinding are more likely to show advantages of a new intervention over the standard treatment.¹⁸ Trials that cannot be double blinded could still score 3 points on the Jadad scale if they included randomization (1 point), appropriate generation of randomization sequence (1 point), and a

detailed account of withdrawals and dropouts (1 point). Furthermore, blinding can be achieved at levels other than participants and assessors, such as treatment providers, analysts, and reporters. Of all the analyzed trials, 43.5% had no description of prestudy sample size calculations and a further 26.1% described the parameters involved in the calculation of the sample size inadequately. Prestudy sample size calculations based on a clearly defined outcome are considered essential for both scientific and ethical reasons. Studies with small sample sizes are often inadequately powered to detect small but clinically significant differences between interventions and are therefore not a valid justification to negate the usefulness of the new treatments.¹⁹ “Intention-to-treat” analyses are usually favored over “per-protocol” analyses as they avoid bias associated with nonrandom loss of participants.^{20–22} Although such analyses may underestimate the real benefit of treatments for which noncompliance may be an issue, they address the effectiveness question that is the more pragmatic approach to the application of the interventions in clinical practice. Only 69.6% of the trials in our analyses clearly reported on intention-to-treat analyses and 81.3% described protocol deviations adequately. There is some evidence to suggest that trials that reported an intention-to-treat analysis and those that reported exclusions are associated with other aspects of good study design when compared with those that did not report on these aspects.^{23,24}

The limitations of our study include the well-known difficulty of separately assessing methodology and reporting. Although many criteria, such as blinding, allocation concealment, recruitment, outcomes assessment, and statistical analysis, clearly fall within the domain of “methodology,” the design of our study necessitates the assessment of “study methodology” through the window of “reporting.” We acknowledge that space would have been a limiting factor for the inclusion of all information that we have used to assess reporting quality, but critical information with a bearing on scientific validity should always be prioritized. Although failure to report critical elements does not imply lack of implementation, we think that adequate reporting is vital for the credibility of a trial’s conclusions. The attempted blinding of the reviewers to the journal, author, and institution names may not have been completely successful as the reviewers (who could be considered domain experts) had previously encountered some of the articles assessed and could have been prejudiced by preexisting knowledge on the subject. In addition, the CONSORT checklist contained items whose scoring was subjective and dependent on reviewers’ perceptions and domain knowledge. The assessment of the item of “allocation concealment” as part of the modified CONSORT checklist was slightly in variance with the assessment described by Schulz et al in 1995,² which explains the difference in the scoring as shown in Tables 3 and 4.

Our findings suggest that the reporting quality of RCTs in general surgery falls well below the optimum and cannot be explained by potential limitations in the scientific assessment of operative interventions. We have recently demonstrated similar results in cardiothoracic literature and shown that authors of RCTs lack awareness of guidelines such as CONSORT.²⁵ As

results of RCTs have a significant impact on clinical decision-making, vigilance is required when relying on the results of these trials for implementing new treatments. The responsibility for improvement should primarily lie with the investigators, but reviewers and editors of surgical journals could facilitate the process by endorsing guidelines such as the CONSORT statement.

REFERENCES

- 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.
- Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408–412.
- Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA.* 1998;279:1489–1491.
- Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ.* 2001;323:42–46.
- Huwiler-Muntener K, Juni P, Junker C, et al. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA.* 2002;287:2801–2804.
- Schulz KF, Chalmers I, Grimes DA, et al. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA.* 1994;272:125–128.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet.* 2001;357:1191–1194.
- 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.
- Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials.* 1995;16:62–73.
- 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.
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA.* 2001;285:1992–1995.
- Junker CA. Adherence to published standards of reporting: a comparison of placebo-controlled trials published in English or German. *JAMA.* 1998;280:247–249.
- Horton R. Surgical research or comic opera: questions, but few answers. *Lancet.* 1996;347:984–985.
- Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA.* 1994;272:96–97.
- Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke.* 1998;29:2203–2210.
- Mills E, Loke YK, Wu P, et al. Determining the reporting quality of RCTs in clinical pharmacology. *Br J Clin Pharmacol.* 2004;58:61–65.
- Jadad AR. *Randomized Controlled Trials.* London: BMJ, 1998.
- Grimes DA. Randomized controlled trials: ‘it ain’t necessarily so.’ *Obstet Gynecol.* 1991;78:703–704.
- Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ.* 1995;311:485.
- Lee YJ, Ellenberg JH, Hirtz DG, et al. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med.* 1991;10:1595–1605.
- Lachin JL. Statistical considerations in the intent-to-treat principle. *Control Clin Trials.* 2000;21:167–189.
- Lewis JA, Machin D. Intention to treat—who should use ITT? *Br J Cancer.* 1993;68:647–650.
- Ruiz-Canela M, Martinez-Gonzalez MA, de Irala-Esteviz J. Intention to treat analysis is related to methodological quality. *BMJ.* 2000;320:1007–1008.
- Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomized controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ.* 1996;312:742–744.
- Tiruvoipati R, Balasubramanian SP, Atturu G, et al. Improving the quality of reporting randomized controlled trials in cardiothoracic surgery: the way forward. *J Thorac Cardio Surg.* 2006. In press.