

The Reporting of Randomized Clinical Trials Using a Surgical Intervention Is in Need of Immediate Improvement

A Systematic Review

Isabelle Jacquier, MD,*† Isabelle Boutron, MD, PhD,* David Moher, PhD,‡ Carine Roy, MSc,* and Philippe Ravaut, MD, PhD*

Objective: To assess the reporting of surgical interventions, care providers, and number of centers in randomized clinical trials.

Methods: Systematic review was performed to assess reports of randomized controlled trials assessing surgical procedure published in 2004. A standardized abstraction form was used to extract data.

Results: A total of 158 articles were included. Details on the intervention intended, such as the surgical procedure, were reported in 138 (87.3%) articles, anesthetic management in 56 (35.4%), preoperative care in 34 (15.2%), and postoperative care in 78 (49.4%). How the experimental surgical intervention was carried out was reported in 64 articles (40.5%). Most trials were conducted in single centers ($n = 109$, 69.0%). The setting was reported in only 11 articles, and the volume of interventions performed was only reported in 5. Selection criteria were reported for care providers in 64 articles (40.5%). The number of care providers performing the intervention was reported in 51 articles (32.2%). The quality of reporting was low as assessed by CLEAR NPT (a 10-items checklist specifically developed to assess the reporting quality of RCTs assessing nonpharmacologic treatment).

Conclusions: Inadequate reporting on the management of the surgical procedure, care providers, and surgery center may introduce bias in RCTs of surgical interventions, making their results questionable. We recommend extending the CONSORT Statement to surgical interventions.

(*Ann Surg* 2006;244: 677–683)

From the *Assistance Publique Hôpitaux de Paris (APHP), Groupe Hospitalier Bichat-Claude Bernard, Département d'Epidémiologie Biostatistique et Recherche Clinique, Paris, France; and Université Paris 7, Paris, France; and †Assistance Publique–Hôpitaux de Paris (APHP), Hôpital St. Antoine, Service d'O.R.L. et de Chirurgie Cervico-faciale, Paris, France; and Assistance Publique–Hôpitaux de Paris (APHP), Hôpital Tenon, Service d'O.R.L. et de Chirurgie Cervico-faciale, Paris, France; and ‡Children's Hospital of Eastern Ontario Research Institute and Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Reprints: Isabelle Boutron, MD, Département d'Epidémiologie Biostatistique et Recherche Clinique, INSERM U738, Groupe Hospitalier Bichat-Claude Bernard, 46 rue Henri Huchard, 75018 Paris, France. E-mail: isabelle.boutron@bch.ap-hop-paris.fr.

Copyright © 2006 by Lippincott Williams & Wilkins

ISSN: 0003-4932/06/24405-0677

DOI: 10.1097/01.sla.0000242707.44007.80

Randomized clinical trials (RCTs) are usually considered the gold standard for therapeutic evaluation.¹ However, surgical trials, because the intervention is multifactorial, separates them from the vastly more common pharmaceutical intervention,² pose special challenges in their conduct and reporting. Some examples, include blinding, influence of care providers' and centers' volume, present special difficulties for randomized trials.³

A clear definition and description of the treatment assessed are a prerequisite of therapeutic evaluation.⁴ As an example, the revised CONSORT statement⁵ recommends reporting "precise details of the interventions intended for each group and how and when they were actually administered." Contrary to pharmacologic treatments, surgical procedures can be very difficult to describe with any degree of precision. Indeed, surgical interventions are complex and involve several components such as preoperative care, the anesthetic procedure, the main surgical intervention(s), and postoperative care (eg, rehabilitation, nursing cares, physiotherapy). Further, several care providers (eg, surgeons, anesthesiologists, nurses, physiotherapists) actively participate in the care of the patient. As such, separating their unique contributions to the "success" of an intervention can be difficult to describe.

Moreover, contrary to pharmacologic treatment, in which the effect of the healthcare professional can be regarded as secondary, in general, in surgery, the healthcare professional is an integral part of the intervention. The success of the intervention depends on care providers' skill, experience, and training. Variation between care providers' skills in each arm of the trial can be confounded with the treatment effect.^{6–9} In addition, the setting and hospital volume of interventions can also influence the results of surgery.^{10–13}

To our knowledge, reporting the description of the experimental intervention in RCTs of surgery has not been reported previously,¹⁴ whereas variation in the quality of administration of intervention may explain some variability in the estimates of effects between trials in systematic reviews.¹⁵

The aim of the present study was to evaluate the quality of reporting of RCTs assessing a surgical intervention and how data on the experimental surgical interventions, and care providers are reported.

METHODS

Search Strategy and Selection of Reports

We searched Medline and the Central Cochrane Register of Clinical Trials (issue 1, 2005) for trials indexed in 2004. The search strategy is detail in Table 1. Retrieved citations were assessed by one of us (I.J.), who screened the titles and abstracts to identify the relevant studies.

Eligibility Criteria

Reports were included only if the study design was identified as an RCT assessing a surgical procedure and published as a full text article. Reports of RCTs assessing pharmacologic treatment adjuvant therapy, technical interventions (dental surgery, endovascular interventions, gastroscopy, colonoscopy, hysteroscopy, etc), implantable devices, surrounding surgical treatments (dressing, contention device, anesthesia, etc), and diagnostic intervention (biopsy, swab, etc) were excluded. Reports were screened for duplicate publication, and only the more recent report was selected.

Among the reports identified, we selected a random sample of reports. An independent statistician (C.R.) obtained a computer-generated list to select 200 journal articles for further evaluation and data extraction.

Data Extraction

From a review of the relevant literature on surgical RCTs, we generated a data collection form that was iterated among the research team. Before data extraction, as a calibration exercise, 2 members of the team (I.B., I.J.) independently evaluated a separate set of 10 reports. A meeting followed in which the ratings were reviewed and any disagreements were resolved by consensus. One reviewer (I.J.) independently completed all the data extractions. A second member of the team reviewed a random sample of 30 articles as a quality assurance exercise. The data abstraction form is available upon request.

Data were obtained on the surgical area (eg, digestive, cardiovascular, orthopedic) and control treatments (eg, placebo, usual care, waiting list, or active control treatment). Funding sources (ie, no manufacturer support, manufacturer support, both kinds of support, or unclear), sample size, and journal impact factor (2003) were also recorded.

TABLE 1. Search Strategy

Medline
Search "surgery [MeSH Terms]"
Limits
Publication Date from January 1, 2004 to December 31, 2004
Only items with abstracts
English
Randomized Controlled Trial
"Humans"
Cochrane Central Register of Controlled Trials (CENTRAL) (issue 1, 2005)
Search "surgery [key words]"
Limits
Date from 2004 to 2004

Reporting the Interventions

We recorded whether details on the interventions intended for each group (ie, interventions described in the methods) was reported and how they were reported (ie, reference article, description of at least some components of the procedure, description of the surgical procedure, anesthetic management, preoperative and postoperative care). We also recorded whether specific methods to standardize the procedure before the trial began were reported (ie, reporting of early developmental phases preceding the standardization of the intervention, use of protocol guidelines, use of video of the procedure).

The reporting, a posteriori, of how the surgical procedure, anesthetic management and preoperative and postoperative care were actually administered was recorded, as was the overall duration of the surgical procedure.

Finally, we retrieved any information reported on care providers' compliance with the planned procedure, use of a protocol to intervene if a poor or inappropriate intervention was identified and any quality assurance committee charged to monitor the interventions.

Reporting on Center and Care Providers

Data on the number of centers involved, the reporting of center volume of similar interventions performed, and the reporting of the setting were recorded. Additionally, the following data on care provider were recorded: 1) reporting of selection criteria for care providers (ie, care providers reported as "experienced," trained, having performed a specific number of interventions, years of practice, level of complication or one of the authors); 2) baseline data on care providers, number of care providers performing the experimental intervention and number of patients treated by each care provider; and 3) reporting of details about attending care providers.

Study Quality

The quality of reporting was assessed using CLEAR NPT¹⁶ (a 10-item checklist specifically developed to assess the reporting quality of RCTs assessing nonpharmacologic treatment [NPT]). These items focus on the reporting of generation of allocation sequence, allocation concealment, details of the intervention administered in each group, care providers' skill, patients adherence, blinding of patients, care providers and outcome assessors, follow-up schedule, and intent-to-treat analysis. In unblinded trials, this checklist focuses on the risk of performance and ascertainment bias.

We also assessed whether the groups were described as being similar at baseline regarding the most important prognostic factor and whether eligibility criteria were specified.

Statistical Analysis

We computed descriptive statistics for continuous variables: means, standard deviation (SD), medians, interquartile range (IQR), and minimum and maximum values. Categorical variables were described with frequencies and percentages.

All data analyses were performed with use of the SAS system for Windows, Release 9.1 (SAS Institute, Cary, NC).

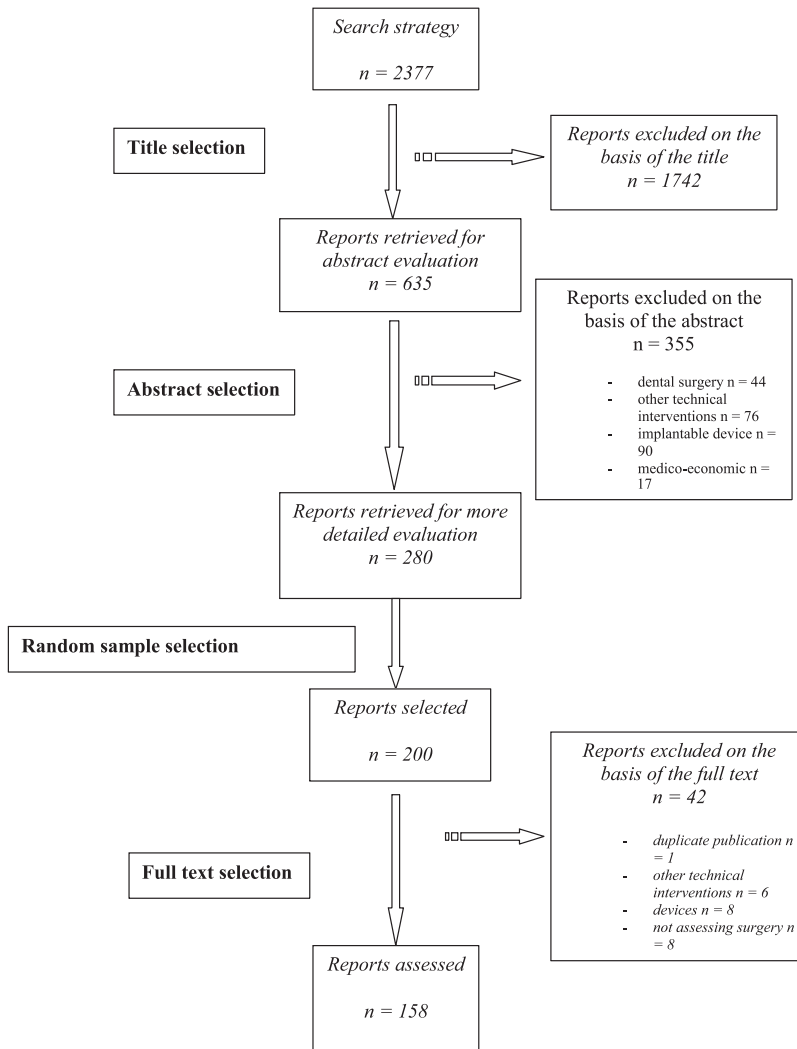


FIGURE 1. Selection of reports.

RESULTS

Articles Selected

The electronic search yield 2377 citations, of which 280 reports were selected for detailed evaluation. Of these, 158 reports were included in our final analysis. The flow of articles through the study is reported in Figure 1.

Characteristics of the Included Articles

The RCTs predominantly reported on digestive and endocrine surgery (n = 51, 32.3%); gynecologic, obstetric, and urogenital surgery (n = 31, 19.6%), orthopaedic surgery (n = 30, 19.0%); and cardiothoracic surgery (n = 19, 12.0%). The mean (SD) and median (IQR) journal impact factors were 3.3 (5.4) and 2.0 (1.4–3.2), respectively. Twenty-one reports (13.3%) had a high impact factor (>5).

The funding source was not reported in most articles (n = 97, 61.4%). The source of funding was described as public in 43 (27%) articles and private (partially or totally) in 18 (11.4%).

The mean (SD) and median (IQR) sample size was 166 (223.8) and 80 (45–185), respectively. All trials were parallel group RCTs.

The control treatment was described as active in 149 reports, usual care in 7, and active and usual care in 2. The active control treatment was described as a surgical intervention in 136 reports (86.1%). No placebo control interventions were reported.

The statistical analysis regarding the main outcome was described as significant in 57 (36%) reports, not significant in 77 (48.7%), and unclear in 24 (15.2%).

Reporting the Surgical Interventions

Details on the interventions intended (ie, interventions as they were described in Methods) for the experimental surgical procedure was described in 140 reports (88.6%).

Details on the interventions intended were provided by reporting some components of the procedure (n = 96, 60.8%), a reference to another article (n = 4, 2.5%), and both a reference and description of the procedure (n = 40, 25.3%).

TABLE 2. Reporting the Experimental and Control Surgical Interventions

Reporting	Experimental Intervention (n = 158)		Control Surgical Intervention (n = 135)	
	n	%	n	%
Details of the intended intervention	140	88.6	135	100.0
Surgical procedure	138	87.3	116	95.1
Anesthetic management	56	35.4	48	39.7
Preoperative care	24	15.2	20	16.5
Postoperative care	78	49.4	71	58.7
How the intervention was actually administered	64	40.5	52	38.5
Surgical procedure	59	37.3	47	34.8
Anesthetic management	10	6.3	10	7.4
Preoperative care	2	1.3	2	1.5
Postoperative care	15	9.5	15	11.1

These details concerned the surgical procedure in 138 articles (87.3%), the anesthetic management in 56 (35.4%), preoperative care in 24 (15.2%), and postoperative care in 78 (49.4%) (Table 2). The results are similar for reporting the surgical procedure performed in the control group (Table 2). Use of a protocol guideline was reported in one article and use of a video of the surgical procedure to standardize the procedure was reported in one paper. A developmental phase preceding standardization was reported in 3 articles (2%): one study involved dogs, one cadaver, and one retrospective study.

Details on how the surgical experimental intervention was actually administered were reported in 64 articles (40.5%). These details concerned the surgical procedure (n = 59, 37.3%), anesthetic management (n = 10, 6.3%), preoperative care (n = 2, 1.3%), and postoperative care (n = 15, 9.5%) (Table 2). The overall duration of the procedure was reported in 62 articles (39.2%). The results are similar for reporting the surgical procedure performed in the control group (Table 2).

Care providers' compliance with the planned procedure was reported in only 4 (2.5%) articles. The evaluation of care providers' qualitative compliance by a quality assurance committee was reported in 2 articles, with the development of a standard treatment protocol and the determination of minimum competence. A quality control procedure was described as being used to monitor care providers' adherence in 1 report.

Reporting on Centers and Care Providers

Most trials were conducted at a single center (n = 109, 69.0%), 12 (7.6%) in 2 centers and 30 (19.0%) in more than 2 centers. In 4 (2.5%) articles, the trial was reported as being multicenter but the exact number of centers was not provided. The number of centers could not be determined in 3 reports (1.9%). The setting (ie, primary, secondary, tertiary, academic) was reported in 11 articles, and the volume of interventions performed in centers was reported in only 5 articles.

At least some data on surgeons were reported in 82 articles (52.0%). The number of surgeons performing the experimental surgical procedure was reported in 51 reports (32.3%), with only 1 surgeon performing the procedure in 28 reports (17.7%). Selection criteria for surgeons were reported in 64 articles (40.5%). Surgeons were mainly described as "experienced" (n = 30, 19.0%), by use of authors' names (n = 28, 17.7%) or by reporting the number of experimental interventions performed before the beginning of the trial (n = 18, 11.4%) (Table 3). The intervention was reported as being performed under the supervision of an experienced surgeon in 10 reports (6.3%). Specific training of surgeons was reported in 2 articles (ie, the surgeons reviewed a videotape on the procedure; the surgical procedures were performed according to protocol guidelines) and surgeons' skill was reported as being assessed on a video reviewed by an executive committee in 2 reports.

Details on attending care providers such as anesthesiologists and intensive care providers were never reported.

Study Quality

Details of the results of the items of the CLEAR NPT are described in Table 4. The treatment allocation was concealed in only 24.7% (n = 39) of the reports. Blinding was adequately reported for participants in only 8.2% of the reports, the outcome assessor in 17.1% and never for care providers. When blinding was not reported for patients or care providers, cointerventions were considered as similar in more than half reports and the number of withdrawals and lost to follow-up was similar in about 60% of the reports. An intent-to-treat analysis was described in only 35.5% (n = 56) of the reports. The follow-up schedule was described as being the same in each group in 61.4% (n = 97) of the reports. When outcome assessors were not reported as adequately blinded, specific methods to avoid ascertainment bias such as

TABLE 3. Reporting of Descriptive Information on Surgeons (n = 158)

	n	%
Selection criteria reported for surgeons		
Selection criteria for surgeons reported	64	40.5
Surgeons reported as "experienced"	30	19.0
Surgeons identified as an author	28	17.7
Years of practice	1	0.6
No. of interventions performed	18	11.4
Level of complication of interventions	1	0.6
Under supervision of a senior surgeon	10	6.3
Specific training	2	1.3
Video training	2	1.3
Details on surgeons performing the surgical procedure		
No. of surgeons performing the procedure reported	51	32.3
No. of patients treated by each surgeon reported	33	20.9
Details on important attending care providers reported	0	0.0

TABLE 4. Number (%) of Reports Providing Information for Each Item of the CLEAR NPT in Reports of Surgery (n = 158)

Items of CLEAR NPT	Yes	
	n	%
1. Was the generation of allocation sequences adequate?	65	41.1
2. Was the treatment allocation concealed?	39	24.7
3. Were details of the intervention administered to each group made available?	120	76.0
4. Were care providers' experience or skill in each arm appropriate?	38	24.0
5. Was patient adherence assessed?	0	0.0
6. Were participants adequately blinded?	13	8.2
6a. If participants were <i>not adequately blinded</i> , were all other treatments and care (ie, cointerventions) the same in each randomized group?	82	56.5
6b. If participants were <i>not adequately blinded</i> , were the number of withdrawals and lost to follow-up the same in each randomized group?	88	60.7
7. Were care providers or persons caring for the participant adequately blinded?	0	0.0
7a. If care providers were not adequately blinded, were all other treatments and care (ie, cointerventions) the same in each randomized group?	93	58.8
7b. If care providers were not adequately blinded, were the number of withdrawals and lost to follow-up the same in each randomized group?	94	59.5
8. Were outcome assessors adequately blinded to assess the primary outcomes?	27	17.1
8a. If outcome assessors were not adequately blinded, were specific methods used to avoid ascertainment bias (systematic differences in outcome assessment)?	16	12.2
9. Was the follow-up schedule the same in each group?	97	61.4
10. Were the main outcomes analyzed according to the intent-to-treat principle?	56	35.5

an independent adjudication committee, use of hard main outcomes (death) were reported in 16 reports.

The groups were described as being similar at baseline regarding the most important prognostic factor in 83.0% of the reports, and the eligibility criteria were specified in 88.6%.

DISCUSSION

This study assessed the reporting of different components of surgical interventions, care providers, and reporting quality in a random sample of surgical RCTs indexed in 2004.

Our results suggest that descriptive data are reported on the surgical procedure intended in most articles. However, reporting of how the procedure was carried out, details on the surrounding management, such as preoperative and postoperative care, anesthesia, and intensive care, were lacking. Further, the reporting on surgeons was poor and data on

centers were almost never reported. Our results suggest that the quality of reporting of surgical RCTs is suboptimal and in need of immediate remedy.

Some studies have assessed the quality of reports assessing surgery,^{17,18} the application of the CONSORT Statement,¹⁴ while others have focused on the reporting of the surgical procedure.^{19,20} We think this is the first study assessing a representative sample of recently published trials that focused on several characteristics important in surgical trials.

Surgical Intervention

Simpler interventions, such as pharmaceuticals, are likely easier to describe. It is far more difficult to describe NPT such as surgery. Surgical interventions are usually complex, multifactorial, including care by several health professionals and several treatments.³ Each component is difficult to standardize, and there may be important subtle differences between the intervention intended and the one actually received. Further each component can influence treatment effect. For example, medical and surgical comanagement after elective hip and knee arthroplasty reduce postoperative complication rate.²¹

To enhance adequate reporting, reporting standards have been advocated for specific surgical procedures such as endovascular aortic aneurysm repair.²² Generic reporting guidelines for all nonpharmacologic treatments are currently in progress. These guidelines recommend the reporting of all the components involved in the intervention, such as preoperative care, the principal procedure (ie, procedure the surgeons believe to contribute most to the treatment), intraoperative and postoperative care, and device configurations when needed. The advantage of such guidelines is that we will provide a unique system of reference for authors and editors. However, such guidelines will not take into account the specificities of each surgical procedure. Consequently, reporting guidelines regarding some components of the procedure, such as anesthetic management, rehabilitation, and care providers' volume of procedures, might not be relevant for all surgical procedures.

Centers and Care Providers

There is abundant evidence that hospitals with a larger volume of activity tend to have better outcomes and that surgeons' volume of work could also be a determinant of outcome.^{7,23–27} For example, RCTs performed to assess carotid endarterectomy²⁸ were conducted in large-volume hospitals by surgical teams with low perioperative complication rate. The results of this procedure performed in clinical practice by surgical teams with less experience shows an in-hospital mortality rate 10-fold higher. Further, when assessing a new surgical procedure, the learning curve should be taken into account.^{3,29–32} For example, surgeons' use of laparoscopic surgical procedures without previous training have been associated with an increased rates of bile duct injury.³³

Most trials were conducted in single centers, comparing 2 surgical procedures, involving few surgeons. The general applicability of results of these trials to clinical practice should be made cautiously. Only 40% of the articles reported

criteria on selection of surgeons. These criteria were mainly surgeons' names or reporting that the surgeons were "experienced," with no details on what "experienced" meant. Lack of reporting these data is a barrier to appraising the risk of bias and gauging the generalizability of results. Indeed, the relative skill of surgeons may be a confounding influence in the purported effectiveness of surgical RCTs. Surgical skill and/or experience might also influence the frequency of harms.^{9,34} Randomizing between a familiar and an unfamiliar surgical procedure might therefore introduce bias into this evaluation.^{9,35,36}

Consequently, precisely reporting surgeons' skill such as qualification, experience in the procedure or in similar procedures, years of practice and specific training performed before joining the trial should be encouraged. In addition, the reporting of the volume of the experimental surgical procedure or a similar procedure in a center is recommended.¹²

Quality

The quality of reporting was assessed by using CLEAR NPT a checklist of items that takes into account the difficulties of assessing NPT.^{37–39} Our results highlight the lack of adequate quality of surgical reports for items common to all RCTs but also for items specific to NPTs. Treatment allocation was described as concealed in only 24.7% of the reports, blinding of outcome assessor in only 17.1%, and an intent-to-treat analysis in only one third. These results are consistent with those of Bhandari et al.^{14,39} Further data on care providers' experience or skill, specific methods to avoid ascertainment bias, follow-up schedule in each group were insufficiently reported. Low-quality reporting of RCTs is associated with increased bias, in the form of exaggerated and possibly spurious estimates of treatment effects,⁴⁰ making their utility of limited use in clinical practice.

Use of the CONSORT Statement is associated with increases in quality of reporting.^{41,42} Partly because of these results, the CONSORT Statement has been extended to other designs,⁴³ data,⁴⁴ and interventions.⁴⁵ Given the results reported here, it might be prudent to consider extending CONSORT, specifically the checklist, for trials involving NPT interventions.

Limitations of This Study

This study was limited because we only assessed reports of RCTs, not the trials themselves, and failure to report is not necessarily equivalent to failure to actually carry out the procedures.⁴⁶ The data appear inconclusive on this point with more recent results suggesting that inadequate reporting is associated with an increased risk of bias in estimating the "true" estimate of effectiveness.⁴⁷ More generally, busy clinicians, wanting to use evidence-based methods to inform their practice, will likely have to rely on reports of RCTs for the foreseeable future.

CONCLUSION

Despite the recommendations of the CONSORT Statement to report details of the intervention intended and how it was administered, this study highlights the low reporting

involving surgical interventions. An extension of the CONSORT Statement to NPT might help reduce these problems.

REFERENCES

1. Prescott RJ, Counsell CE, Gillespie WJ, et al. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technol Assess.* 1999;3:1–143.
2. Chan A-W, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet.* 2005;365:1159–1162.
3. McCulloch P, Taylor I, Sasako M, et al. Randomised trials in surgery: problems and possible solutions. *BMJ.* 2002;324:1448–1451.
4. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ.* 2000;321:694–696.
5. 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.
6. Stark J, Gallivan S, Lovegrove J, et al. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet.* 2000;355:1004–1007.
7. Rothwell PM, Warlow CP. Interpretation of operative risks of individual surgeons: European Carotid Surgery Trialists' Collaborative Group. *Lancet.* 1999;353:1325.
8. Devereaux PJ, McKee MD, Yusuf S. Methodologic issues in randomized controlled trials of surgical interventions. *Clin Orthop.* 2003;413:25–32.
9. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ.* 2005;330:88.
10. McGrath PD, Wennberg DE, Dickens JD Jr, et al. Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. *JAMA.* 2000;284:3139–3144.
11. Soljak M. Volume of procedures and outcome of treatment. *BMJ.* 2002;325:787–788.
12. Urbach DR, Baxter NN. Does it matter what a hospital is 'high volume' for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *Qual Saf Health Care.* 2004;13:379–383.
13. Killeen SD, O'Sullivan MJ, Coffey JC, et al. Provider volume and outcomes for oncological procedures. *Br J Surg.* 2005;92:389–402.
14. Bhandari M, Guyatt GH, Lochner H, et al. Application of the Consolidated Standards of Reporting Trials (CONSORT) in the Fracture Care Literature. *J Bone Joint Surg Am.* 2002;84:485–489.
15. Herbert RD, Bo K. Analysis of quality of interventions in systematic reviews. *BMJ.* 2005;331:507–509.
16. Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *J Clin Epidemiol.* 2005;58:1220–1226.
17. Schumm LP, Fisher JS, Thisted RA, et al. Clinical trials in general surgical journals: are methods better reported? *Surgery.* 1999;125:41–45.
18. Solomon MJ, McLeod RS. Surgery and the randomised controlled trial: past, present and future. *Med J Aust.* 1998;169:380–383.
19. Solomon MJ, Laxamana A, Devore L, et al. Randomized controlled trials in surgery. *Surgery.* 1994;115:707–712.
20. Hall JC, Platell C, Hall JL. Surgery on trial: an account of clinical trials evaluating operations. *Surgery.* 1998;124:22–27.
21. Huddleston JM, Long KH, Naessens JM, et al. Medical and surgical comanagement after elective hip and knee arthroplasty: a randomized, controlled trial. *Ann Intern Med.* 2004;141:28–38.
22. Chaikof EL, Blankensteijn JD, Harris PL, et al. Reporting standards for endovascular aortic aneurysm repair. *J Vasc Surg.* 2002;35:1048–1060.
23. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med.* 2002;137:511–520.
24. Begg CB, Riedel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med.* 2002;346:1138–1144.
25. Martling A, Cedermark B, Johansson H, et al. The surgeon as a prognostic factor after the introduction of total mesorectal excision in the treatment of rectal cancer. *Br J Surg.* 2002;89:1008–1013.

26. Cowan JA Jr, Dimick JB, Thompson BG, et al. Surgeon volume as an indicator of outcomes after carotid endarterectomy: an effect independent of specialty practice and hospital volume. *J Am Coll Surg*. 2002;195:814–821.
27. Harmon JW, Tang DG, Gordon TA, et al. Hospital volume can serve as a surrogate for surgeon volume for achieving excellent outcomes in colorectal resection. *Ann Surg*. 1999;230:404–411; discussion 411–413.
28. North American Symptomatic Carotid Endarterectomy Trial. Methods, patient characteristics, and progress. *Stroke*. 1991;22:711–720.
29. Darzi A, Smith S, Taffinder N. Assessing operative skill: needs to become more objective. *BMJ*. 1999;318:887–888.
30. Darzi A, Mackay S. Skills assessment of surgeons. *Surgery*. 2002;131:121–124.
31. Darzi A, Mackay S. Assessment of surgical competence. *Qual Health Care*. 2001;10(suppl 2):64–69.
32. Darzi A, Datta V, Mackay S. The challenge of objective assessment of surgical skill. *Am J Surg*. 2001;181:484–486.
33. Moore MJ, Bennett CL. The learning curve for laparoscopic cholecystectomy: the Southern Surgeons Club. *Am J Surg*. 1995;170:55–59.
34. Lawrence K, McWhinnie D, Collin J, et al. Surgical evaluation. *Br J Surg*. 1994;81:1390–1392.
35. Bonenkamp JJ, Songun I, Hermans J, et al. Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet*. 1995;345:745–748.
36. Lilford R, Braunholtz D, Harris J, et al. Trials in surgery. *Br J Surg*. 2004;91:6–16.
37. Boutron I, Tubach F, Giraudeau B, et al. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA*. 2003;290:1062–1070.
38. Boutron I, Tubach F, Giraudeau B, et al. Blinding was judged more difficult to achieve and maintain in non-pharmacological than pharmacological trials. *J Clin Epidemiol*. 2004;57:543–550.
39. Bhandari M, Richards RR, Sprague S, et al. The quality of reporting of randomized trials in the Journal of Bone and Joint Surgery from 1988 through 2000. *J Bone Joint Surg Am*. 2002;84:388–396.
40. Egger M, Juni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*. 2003;7:1–76.
41. 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.
42. Devereaux PJ, Manns BJ, Ghali WA, et al. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials*. 2002;23:380–388.
43. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004;328:702–708.
44. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141:781–788.
45. Gagnier JJ, Boon H, Rochon P, et al. Improving the quality of reporting of randomized controlled trials evaluating herbal interventions: an extension of the CONSORT Statement. *Submitted*.
46. Devereaux PJ, Choi PT-L, El-Dika S, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol*. 2004;57:1232–1236.
47. Pildal J, Chan AW, Hrobjartsson A, et al. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ*. 2005;330:1049.