



■ ANNOTATION

The design and assessment of prospective randomised, controlled trials in orthopaedic surgery

I. Boutron,
P. Ravaut,
R. Nizard

From Université Paris
VII Denis Diderot,
Paris, France

Randomised controlled trials represent the gold standard in the evaluation of outcome of treatment. They are needed because differences between treatment effects have been minimised and observational studies may give a biased estimation of the outcome. However, conducting this kind of trial is challenging. Several methodological issues, including patient or surgeon preference, blinding, surgical standardisation, as well as external validity, have to be addressed in order to lower the risk of bias. Specific tools have been developed in order to take into account the specificity of evaluation of the literature on non-pharmacological intervention. A better knowledge of methodological issues will allow the orthopaedic surgeon to conduct more appropriate studies and to better appraise the limits of his intervention.

In 1996, Richard Horton,¹ editor of *The Lancet*, initiated a debate on the quality of surgical research. Randomised, controlled trials (RCTs) accounted for only 7% of the original research published in journals on general surgery, despite this design of study producing a high level of evidence.^{2,3} Recently, an assessment of orthopaedic journals found that only 11.3% of the published articles were graded as level I⁴ representing the best level of evidence according to the Oxford Centre for Evidence based Medicine.⁵ These alarming results have triggered much discussion about the difficulties in assessing surgical procedures.⁶⁻⁸ This paper reviews the advantages, limitations and main pitfalls in conducting and assessing RCTs in orthopaedic surgery.

Do we need systematic randomised controlled trials in orthopaedic surgery?

Although RCTs are widely held to be the optimum level of evaluation, they may be unnecessary, inappropriate or inadequate.⁹ They are inappropriate when the effect of an intervention is dramatic, and when the likelihood of unknown confounding factors is small and can be ignored.¹⁰⁻¹² Consequently, observational studies are more widely used to demonstrate the effectiveness of procedures such as surgery for fracture of the hip.

However, with the evolution and improvements in health care, differences between treatments have been minimised, and observational studies may give a biased estimation of their

effect. There are several examples of pharmacological and surgical treatments which were initially deemed to be efficacious on the basis of observational studies and physiological considerations, but were then found to be less useful, without value, or even detrimental in subsequent RCTs.^{13,14} For example, initially the use of intramedullary nails for extracapsular fracture of the hip was found, on the basis of observational studies, to be effective, but a meta-analysis of RCTs showed significantly more peri- and post-operative complications when compared with treatment by sliding screws.¹⁵

Randomised controlled trials are inappropriate for the accurate measurement of infrequent adverse outcomes, the assessment of interventions designed to prevent rare events and for the evaluation of outcomes which are far in the future. For the determination of the long-term or rare outcome, the number of patients required in the trial and the difficulties of follow-up are immense. Other designs, such as the use of an arthroplasty register¹⁶⁻¹⁹ or administrative data, probably better retrieve the relevant information.

Problems in conducting randomised controlled trials in orthopaedic surgery

One of the major barriers is the lack of a regulatory agency, such as the US Food and Drug Administration (Rockville, Maryland), for surgery. Unlike drugs, which must be carefully investigated before their release, new surgical

n I. Boutron, MD, PhD,
Research Fellow
■ P. Ravaut, MD, PhD,
Professor of Clinical
Epidemiology
Département d'Epidémiologie,
Biostatistique et Recherche
Clinique, Hôpital Bichat, rue
Henri, Huchard, Paris, France.

■ R. Nizard, MD, PhD,
Professor of Orthopaedic
Surgery
Service d'Orthopédie
Hôpital Lariboisière, 2 rue
Ambroise Paré, 75475 Paris,
Cedex 10, France.

Correspondence should be sent
to Professor R. Nizard; e-mail:
remy.nizard@noos.fr

©2007 British Editorial Society
of Bone and Joint Surgery
doi:10.1302/0301-620X.89B7.
19440 \$2.00

J Bone Joint Surg [Br]
2007;89-B:858-63.

procedures can be performed with few restrictions,^{20,21} reducing the incentive for surgeons to perform RCTs.

Many surgeons and other key personnel are reluctant to participate in RCTs.²² Although surgeons may agree that the best treatment is unknown, each may prefer one method over another, which would prevent participation in a trial.^{8,20} The recruitment of patients may also present difficulties. Their preference is linked to the disadvantages or risks as patients compared with the supposed benefits, and this may limit the number willing to take part in an RCT. Such difficulties are more likely when patients have access to their preferred treatment outside the trial.

Difficulty in funding is a barrier to performing an RCT.^{20,23} Unlike with drugs, surgical trials rely mainly on funding from academia, and grants for assessing surgical procedures are lacking. For example, the rate of success of proposals related to surgery made to the National Institutes of Health, the major source of biomedical funding in the USA, is significantly lower than that for those not involving surgery. The mean amounts awarded for funded research proposals are consistently lower for surgeons than for non-surgeons, with a range of differences of between 5% and 27%.²⁴ However, these difficulties may be self-inflicted since the granting of funds may be influenced by the poor quality of some of the surgical research.

The quality of published reports of orthopaedic randomised controlled trials

If an RCT is to be considered as the most powerful method of demonstrating the effectiveness and risks of a treatment, the planning, conduct and reporting of the results must be of a high standard to allow for accurate assessment.²⁵ In order to avoid bias, it is necessary to have: an adequate method of generation of a randomised sequence; concealment of allocation, which prevents knowledge of the assignment of treatment, thus shielding those who enrol participants from being influenced by this knowledge; blinding, and an analysis of an intention-to-treat in which all randomised subjects are analysed in the groups to which they have been allocated.^{26,27} These standards are not always adequately reported in surgical trials. A systematic review of reports of RCTs in surgery published in 2004 showed that allocation of treatment was concealed in only 24.7% of the reports, a blinded outcome assessment was disclosed in only 17% and an intention-to-treat analysis described in only 35.5%.²⁸

Bhandari et al²⁹ assessed the quality of the reported results of RCTs published in the *Journal of Bone and Joint Surgery* between 1988 and 2000 and gave drug trials a significantly higher mean quality score than surgical trials. Only 7% of surgical trials had at least one author with biostatistical or methodological expertise, and less than half of the reports described concealed randomisation or blinded assessment of outcome.²⁹ A systematic review of the literature on fractures showed that 99% of the studies were monocentric trials.³⁰ The studies³⁰ adhered to about one-

third of the items in the CONSORT checklist,³¹ and fewer than 5% provided details which allowed readers to determine whether the patients, clinicians and outcome assessors were blinded in terms of allocation. These findings are consistent with other reviews assessing the quality of surgical trials performed in other fields.³²⁻³⁵

The main methodological issues in surgical randomised controlled trials

The methodological standards of RCTs have mainly been developed from the assessment of pharmacological treatments. Some of these concepts, such as the method of randomisation, concealment of allocation and intention-to-treat analysis, are applicable to surgical trials. However, some of these methodological issues are influenced by the high preference of patients and surgeons for one treatment, the difficulties in blinding, the complexity of the intervention, the expertise of the surgeon and the volume of treatment carried out in the centre. These issues require specific attention when planning, conducting and reporting the results of an RCT.

In surgical trials, patients may have strong preferences for the treatment under evaluation, especially if it is to be compared with a pharmacological or conservative approach. This issue is critical for surgical trials because regulatory controls for surgical procedures are not as strict as those for pharmacological treatments and patients can be treated by a new surgical procedure outside the framework of the RCT. Consequently, patients may decline to consent to randomisation.

When blinding is not feasible, patients who are randomised to an intervention which they have not preferred may feel resentful, and this may lead to bias.³⁶ The Spine Outcomes Research Trial³⁷ which compared discectomy and conservative care for patients with disc-related pain and neurological symptoms, had important pitfalls. The strong preferences of patients and surgeons, the subjective main outcome and the lack of blinding resulted in a high proportion of patients (45% to 60%) who 'crossed over' treatment strategies with a significant amount of missing data (24% to 27%), thus preventing proper conclusions to be drawn.

To overcome these problems, the Zelen or postrandomisation consent design, and the modified Zelen design with a two-stage consent procedure, may be used.^{38,39} In the modified Zelen design, initially the patients are informed and sign a consent form to participate in a cohort study. They are then told about the treatment to which they have been allocated. This design could enhance recruitment in surgical trials and limit the risk of bias when patients have a strong preference for one procedure over another.^{39,40}

The blinding of patients, providers of health care and assessors of outcome is essential to avoid performance bias, with unequal administration of co-interventions in each part, and ascertainment bias when there is unequal evaluation of methods undergoing comparison. Lack of blinding

may lead to bias in estimating the effects of treatment.^{27,41-43} Blinding is more difficult to achieve in surgical RCTs⁴⁴ and some surgeons have proposed innovative methods and strategies to overcome this.^{45,46}

Placebos of surgery have been suggested as in pharmacological trials. These rely on sham procedures consisting mainly of simulating the intervention and 'standardising' the post-operative care. For example, Moseley et al⁴⁷ performed simulated arthroscopic surgery, in which surgeons made a small incision, asked for all the instruments and manipulated the knee as if arthroscopy was being performed, but inserted no instruments. However, the ethical problems linked to the risk of sham interventions when assessing surgery are still not resolved,^{8,21,48} and patients may be reluctant to undergo such procedures. In the arthroscopic trial only about 40% of the eligible patients agreed to participate.⁴⁷ Therefore the external validity of such trials is questionable.

Other proposals have been suggested to limit the risk of performance bias. Contact between patients and non-blinded surgeons may be avoided by having blinded health-care providers follow patients and prescribe co-interventions. In trials comparing two surgical procedures, the risk of performance bias may be limited by asking surgeons to document the intended optimum intervention before randomisation.⁴⁹ Blinding patients by using the Zelen or the modified Zelen design, although rarely proposed, may also be a possible solution.

The risk of ascertainment bias when the provision of the intervention cannot be blinded merits discussion. Some evidence has suggested that a lack of blinding in RCTs may render them less prone to bias when the main outcome is objective (e.g. death).⁴² However, when the main outcome is subjective, such as pain or range of movement, special attention should be paid to blinding the outcome assessor and the use of a prospective, randomised open blinded endpoint (PROBE) study.⁵⁰ A systematic review of the methods of blinding in non-pharmacological trials highlighted that for most outcomes a blinded outcome assessment is feasible.^{45,46} However, with patient-reported outcomes, if the patients cannot be blinded there is a risk of ascertainment bias,^{45,46} but this can probably be limited if the individual interviewing the patients is independent and blinded.

Surgical interventions are complex and consist of several components, each of which influences the effect of the treatment. Assessment of treatment involves the evaluation of the effect of surgery, of pharmacological treatment, anaesthesia, rehabilitation, orthoses, rest etc. Consequently, these interventions must be described and standardised in order to be administered consistently to all patients and reproduced in clinical practice. There may also be a discrepancy between the intended intervention as described in the protocol and that which is actually administered.²¹ The surgeons must be aware that adherence to the planned procedure is important in order to allow an adequate application of the results in practice.⁵¹ A systematic review

of surgical RCTs showed that if the intended surgical procedure was described in most articles (87.3%), other important components such as the management of anaesthesia (35.4%) and pre-operative (15.2%) and post-operative care (49.4%), were lacking.²⁸ Furthermore, the description of the actual operation which was performed was given in less than half of the reports.²⁸

The expertise of the surgeon and the volume of work carried out in the particular centre have considerable influence on the success of a surgical treatment.⁵²⁻⁶² Surprisingly, in surgical RCTs, these issues are never accounted for in the planning of the trial or analysis of the results. For example, in a systematic review of RCTs in surgery, the institution and its volume of work were described in only 7% and 3% of the articles, respectively, and the number of care providers was given in only one-third.²⁸

Surgical expertise and the volume of work may introduce bias in RCTs and have a considerable influence on their external validity. Variation of expertise in one arm of the treatment compared with another implies a bias against the more difficult procedure. Devereaux et al⁶³ have advocated the use of an alternative design of RCT in which patients are randomised to surgeons with the most advanced expertise in the procedure. This should avoid the bias of differential expertise between treatment groups, increase the rate of participation of patients, and decrease the risk of contamination and differential co-interventions across the treatment groups.⁶⁴ However, this approach has never been used in elective orthopaedic surgery, only in emergency surgery in which patient choice is limited.

Although the determination of, and reporting on the expertise of surgeons in a procedure has been advocated, the evaluation of expertise depends on the procedure assessed.^{65,66} Some authors^{65,66,67} have focused on the qualifications of the surgeons, the years in practice, specific training before participating in the trial, the number of procedures performed and the learning curve or rate of complications.

In individual RCTs in surgery, variation in the outcome will be minimal for patients treated by the same surgeon.⁶⁷ The assumption that the observed outcomes of participants are independent is false, and observations for participants treated by the same surgeon may be clustered.⁶⁷ This type of clustering must be taken into account when calculating sample size and in the statistical analyses since it is likely to inflate the standard error and reduce the effective sample size, thus reducing the power of the trial.⁶⁷

Reporting guidelines

The CONSORT³¹ statement published in 1996 and revised in 2001³¹ attempted to improve the standard of RCTs in surgery.³⁵ However, important specific issues such as the method of blinding, the expertise of the surgeon, the volume of work of the centre and the different components of the intervention intended and actually administered were insufficiently reported.²⁸

Table I. Checklist of items for assessing the quality of our randomised controlled trials

| | Yes | No | No, blinding not feasible | No, although blinding feasible | Unclear |
|---|-----|----|---------------------------|--------------------------------|---------|
| 1. Was the generation of allocation sequences adequate? | | | | | |
| 2. Was the treatment allocation concealed? | | | | | |
| 3. Were details of the intervention administered to each group made available?* | | | | | |
| 4. Were care providers' experience or skill [†] in each arm appropriate? [‡] | | | | | |
| 5. Was participant (i.e. patient) adherence assessed quantitatively? [§] | | | | | |
| 6. Were the participants adequately blinded? | | | | | |
| 6.1 If participants were not adequately blinded | | | | | |
| 6.1.1 Were all other treatments and care (i.e. co-interventions) the same in each randomised group? | | | | | |
| 6.1.2 Were withdrawals and loss to follow-up the same in each randomised group? | | | | | |
| 7. Were care providers or persons caring for the participants adequately blinded? | | | | | |
| 7.1 If care providers were not adequately blinded | | | | | |
| 7.1.1 Were all other treatments and care (i.e. co-interventions) the same in each randomised group? | | | | | |
| 7.1.2 Were withdrawals and loss to follow-up the same in each randomised group? | | | | | |
| 8. Were outcome assessors adequately blinded to assess the primary outcomes? | | | | | |
| 8.1 If outcome assessors were not adequately blinded, were specific methods used to avoid ascertainment bias (systematic differences in outcome assessment)? [¶] | | | | | |
| 9. Was the follow-up schedule the same in each group?* | | | | | |
| 10. Were the main outcomes analysed according to the intention-to-treat principle? | | | | | |

* the answer should be 'yes' for this item if these data were either described in the report or made available for each arm (reference to a preliminary report, online addendum etc.)

† care provider experience or skill will be assessed only for therapist-dependent interventions (i.e. interventions in which the success of the treatment is directly linked to the care providers' technical skill). For other treatment, this item is not relevant and should be removed from the checklist or answered 'unclear'

‡ appropriate experience or skill should be determined according to published data, preliminary studies, guidelines, run-in period, or a group of experts and should be specified in the protocol for each study arm before the beginning of the survey

§ treatment adherence will be assessed only for treatments necessitating iterative interventions (e.g. physiotherapy which requires several sessions, in contrast to a one-shot treatment such as surgery). For one-shot treatments, this item is not relevant and should be removed from the checklist or answered 'unclear'

¶ the answer should be 'yes' for this item, if the main outcome is objective or hard, or if outcomes were assessed by a blinded or at least an independent end-point review committee or were assessed by an independent outcome assessor trained to perform the measurements in a standardised manner, or if the outcome assessor was blinded to the purpose of the study and hypothesis

** this item is not relevant for trials in which follow-up is part of the question. For example, it is not relevant for a trial assessing frequent vs less frequent follow-up for recurrence of cancer. In these situations, this item should be removed from the checklist or answered 'unclear'

An extension of the CONSORT statements for trials assessing non-pharmacological treatments such as surgery has recently been discussed. The publication of this extension should improve the reporting of these important specific issues in RCTs of surgery and allow for adequate appraisal of the internal and external validity of publications in this field.

Appraisal of the surgical literature

The assessment of the quality of reports of trials is particularly important for clinicians in appraising the literature and systematic reviews.⁶⁸⁻⁷⁰ Problems with the methods

employed can influence the estimation of the effect of treatment,^{26,27} and if the raw material is flawed, then the conclusions of systematic reviews are more likely to compound these biases. Most existing methods have been developed and validated in the context of pharmacological treatments^{71,72} and are not suitable for surgical trials because they do not take into account specific issues in assessing procedures, such as the influence of the providers of care,⁷³⁻⁷⁵ standardisation, the feasibility of blinding and the risk of bias in unblinded trials. A checklist of items, the Checklist to Evaluate a Report of Non Pharmacological Trial (CLEAR NPT) (Table I), was consequently developed

using the Delphi technique⁷⁶ to assess the quality of reports of trials evaluating non-pharmacological treatments such as surgery.⁷⁷ This checklist is simple and quick to complete and should help clinicians to appraise the medical literature critically, reviewers to assess the quality of reports included in systematic reviews and researchers to plan clinical trials in surgery.

Conclusions

The assessment of surgical procedures raises difficulties. The standards of therapeutic evaluation developed to assess pharmacological treatments are not adapted for assessing surgery. Consequently, RCTs in surgery require the use of specific reporting guidelines and quality tools, with the development of new designs and methods of blinding in order to allow for unbiased estimates of the effect of treatment.

References

- Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996;347:984-5.
- Jaeschke R, Sackett DL. Research methods for obtaining primary evidence. *Int J Technol Assess Health Care* 1989;5:503-19.
- Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet* 1996;348:596-8.
- Obrensky WT, Pappas N, Attallah-Wasif E, Tornetta P 3rd, Bhandari M. Level of evidence in orthopaedic journals. *J Bone Joint Surg [Am]* 2005;87-A:2632-8.
- No authors listed. Oxford Centre for Evidence based Medicine. http://www.cebm.net/levels_of_evidence.asp levels (date last accessed 9 July 2007).
- Weil RJ. The future of surgical research. *PLoS Med* 2004;1:13.
- Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA* 2003;290:1062-70.
- McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002;324:1448-51.
- Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-18.
- Concato J, Shah N, Horwitz RJ. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.
- Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003;327:1459-61.
- Glasiou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary?: picking signal from noise. *BMJ* 2007;334:349-51.
- Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999;52:487-97.
- Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. *Ann Surg* 2006;244:668-76.
- Parker MJ, Handoll HH. Gamma and other cephalocondylic intramedullary nails versus extramedullary implants for extracapsular hip fractures in adults. *Cochrane Database Syst Rev* 2005;4:93.
- Eskelinen A, Remes V, Helenius I, et al. Total hip arthroplasty for primary osteoarthritis in younger patients in the Finnish Arthroplasty Register: 4,661 primary replacements followed for 0-22 years. *Acta Orthop* 2005;76:28-41.
- Victor J. Do we need a national arthroplasty register? *Acta Orthop Belg* 2006;72:521-3.
- Engesaeter LB, Espehaug B, Lie SA, Furnes O, Havelin LI. Does cement increase the risk of infection in primary total hip arthroplasty: revision rates in 56,275 cemented and uncemented primary THAs followed for 0-16 years in the Norwegian Arthroplasty Register. *Acta Orthop* 2006;77:351-8.
- Eskelinen A, Remes V, Helenius I, et al. Uncemented total hip arthroplasty for primary osteoarthritis in young patients: a mid- to long-term follow-up study from the Finnish Arthroplasty Register. *Acta Orthop* 2006;77:57-70.
- McLeod RS. Issues in surgical randomized controlled trials. *World J Surg* 1999;23:1210-14.
- McLeod RS, Wright JG, Solomon MJ, et al. Randomized controlled trials in surgery: issues and problems. *Surgery* 1996;119:483-6.
- Seiler CM, Kellmeyer P, Kienle P, et al. Assessment of the ethical review process for non-pharmacological multicentre studies in Germany on the basis of a randomised surgical trial. *J Med Ethics* 2007;33:113-18.
- Bhandari M, Pape HC, Giannoudis PV. Issues in the planning and conduct of randomised trials. *Injury* 2006;37:349-54.
- Rangel SJ, Efron B, Moss RL. Recent trends in National Institutes of Health funding of surgical research. *Ann Surg* 2002;236:277-86.
- Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about prognosis. *J Bone Joint Surg [Am]* 2001;83-A:1555-64.
- Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
- 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;273:408-12.
- Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann Surg* 2006;244:677-83.
- Bhandari M, Richards RR, Sprague S, Schemitsch EH. 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-A:388-96.
- Bhandari M, Guyatt GH, Lochner H, Sprague S, Tornetta P 3rd. Applications of the consolidated standards of reporting trials (CONSORT) in the fracture care literature. *J Bone Joint Surg [Am]* 2002;84-A:485-9.
- Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomised trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
- Balasubramanian SP, Wiener M, Alshameeri Z, et al. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg* 2006;244:663-7.
- Manterola C, Pineda V, Vial M, Losada H. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann Surg* 2006;244:827-32.
- Ellis C, Hall JL, Khalil A, Hall JC. Evolution of methodological standards in surgical trials. *ANZ J Surg* 2005;75:874-7.
- Anyanwu AC, Treasure T. Surgical research revisited: clinical trials in the cardiothoracic surgical literature. *Eur J Cardiothorac Surg* 2004;25:299-303.
- King M, Nazareth I, Lampe F, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. *JAMA* 2005;293:1089-99.
- Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: a randomized trial. *JAMA* 2006;296:2441-50.
- Torgerson DJ, Roland M. What is Zelen's design? *BMJ* 1998;316:606.
- Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. *Contemp Clin Trials* 2006;27:305-19.
- Andrew M, Vegh P, Caco C, et al. A randomized, controlled trial of platelet transfusions in thrombocytopenic premature infants. *J Pediatr* 1993;123:285-91.
- Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;163:493-501.
- Wood L, Egger M, Gluud LL, et al. The association of allocation concealment and blinding with estimated treatment effect varies according to type of outcome: a combined analysis of meta-epidemiological studies. *Cochrane Colloquium*. Abstracts. Dublin, 2006.
- Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? *J Bone Joint Surg [Am]* 2007;89-A:550-8.
- Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in non-pharmacological than pharmacological trials. *J Clin Epidemiol* 2004;57:543-50.
- Boutron I, Estellat C, Guittet L, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med* 2006;3:425.
- Boutron I, Guittet L, Estellat C, et al. Reporting methods of blinding in randomized controlled trials assessing non pharmacological treatments. *PLoS Med* 2007;4:61.
- Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81-8.
- Russell I. Evaluation new surgical procedures. *BMJ* 1995;311:1243-4.
- Puskas JD, Williams WH, Mahoney EM, et al. Off-pump vs conventional coronary artery bypass grafting: early and 1-year graft patency, cost, and quality-of-life outcomes. *JAMA* 2004;291:1841-9.
- Hansson L, Hedner T, Dahlof B. Prospective randomized open blinded end-point (PROBE) study: a novel design for intervention trials. *Blood Press* 1992;1:113-19.

51. Lilford R, Braunholtz D, Harris J, Gill T. Trials in surgery. *Br J Surg* 2004;91:6-16.
52. Doro C, Dimick J, Wainess R, Upchurch G, Urquhart A. Hospital volume and inpatient mortality outcomes of total hip arthroplasty in the United States. *J Arthroplasty* 2006;21(Suppl 2):10-16.
53. Espehaug B, Havelin LI, Engesaeter LB, Vollset SE. The effect of hospital-type and operating volume on the survival of hip replacements: a review of 39,505 primary total hip replacements reported to the Norwegian Arthroplasty Register, 1988-1996. *Acta Orthop Scand* 1999;70:12-18.
54. Jain N, Pietrobon R, Hocker S, et al. The relationship between surgeon and hospital volume and outcome for shoulder arthroplasty. *J Bone Joint Surg [Am]* 2004;86-A:496-505.
55. Jain NB, Hocker S, Pietrobon R, et al. Total arthroplasty versus hemiarthroplasty for glenohumeral osteoarthritis: role of provider volume. *J Shoulder Elbow Surg* 2005;14:361-7.
56. Kreder HJ, Grosso P, Williams JI, et al. Provider volume and other predictors of outcome after total knee arthroplasty: a population study in Ontario. *Can J Surg* 2003;46:15-22.
57. Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. *J Arthroplasty* 1995;10:133-40.
58. Lyman S, Jones EC, Bach PB, Peterson MG, Marx RG. The association between hospital volume and total shoulder arthroplasty outcomes. *Clin Orthop* 2005;432:132-7.
59. Martineau P, Filion KB, Huk OL, et al. Primary hip arthroplasty costs are greater in low-volume than in high-volume Canadian hospitals. *Clin Orthop* 2005;437:152-6.
60. Mitsuyasu S, Hagihara A, Horiguchi H, Nobutomo K. Relationship between total arthroplasty case volume and patient outcome in an acute care payment system in Japan. *J Arthroplasty* 2006;21:656-63.
61. Sharkey PF, Shastri S, Teloken MA, et al. Relationship between surgical volume and early outcomes of total hip arthroplasty: do results continue to get better? *J Arthroplasty* 2004;19:694-9.
62. Soohoo NF, Zingmond DS, Lieberman JR, Ko CY. Primary total knee arthroplasty in California 1991 to 2001: does hospital volume affect outcomes? *J Arthroplasty* 2006;21:199-205.
63. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ* 2005;330:88.
64. Antolovic D, Kienle P, Knaebel HP, et al. Totally laparoscopic versus conventional ileoanal pouch procedure: design of a single-centre, expertise based randomised controlled trial to compare the laparoscopic and conventional surgical approach in patients undergoing primary elective restorative proctocolectomy. *BMC Surg* 2006;6:13.
65. Ramsay CR, Grant AM, Wallace SA, et al. Assessment of the learning curve in health technologies: a systematic review. *Int J Technol Assess Health Care* 2000;16:1095-108.
66. Ramsay CR, Grant AM, Wallace SA, et al. Statistical assessment of the learning curves of health technologies. *Health Technol Assess* 2001;5:1-79.
67. Lee KJ, Thompson SG. Clustering by health professional in individually randomised trials. *BMJ* 2005;330:142-4.
68. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
69. Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001;54:651-4.
70. Berlin JA, Rennis D. Measuring the quality of trials: the quality of quality scales. *JAMA* 1999;282:1083-5.
71. 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.
72. Kattrak P, Bialocerkowski AE, Massey-Westropp N, Kumar VS, Grimmer K. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 2004;4:22.
73. 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;230:404-11.
74. Khuri SF, Daley J, Henderson W, et al. Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. *Ann Surg* 1999;230:414-29.
75. Soljak M. Volume of procedures and outcome of treatment. *BMJ* 2002;325:787-8.
76. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 311:376-80.
77. Boutron I, Moher D, Tugwell P, et al. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58:1233-40.