Review

Assessment of risk of bias in randomized clinical trials in surgery

K. S. Gurusamy[1], C. Gluud[2], D. Nikolova[2] and B. R. Davidson[1]

[1]University Department of Surgery, Royal Free Hospital and University College School of Medicine, London, UK, and [2]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

Correspondence to: Mr K. S. Gurusamy, c/o Professor B. R. Davidson, 9th Floor, Royal Free Hospital, Pond Street, London NW3 2QG, UK (e-mail: kurinchi2k@hotmail.com)

Background: **Meta-analysis of randomized clinical trials (RCTs) with low risk of bias is considered the highest level of evidence available for evaluating an intervention. Bias in RCTs may overestimate or underestimate the true effectiveness of an intervention.**

Methods: **The causes of bias in surgical trials as described by The Cochrane Collaboration, and the methods that can be used to avoid them, are reviewed.**

Results: **Blinding is difficult in many surgical trials but careful trial design can reduce the bias risk due to lack of blinding. It is possible to conduct surgical trials with low risk of bias by using appropriate trial design.**

Conclusion: **The risk of providing a treatment based on a biased effect estimate must be balanced against the difficulty of conducting trials with very low risk of bias. Better understanding of the risk of bias may result in improved trials with a closer estimate of the true effectiveness of an intervention.**

Paper accepted 16 December 2008

Published online in Wiley InterScience (www.bjs.co.uk). **DOI:** 10.1002/bjs.6558

Introduction

Meta-analysis of randomized clinical trials (RCTs) with low risk of bias is considered the highest level of evidence available for evaluating the beneficial and harmful effects (effectiveness and harm) of an intervention[1]. Such analysis is an important factor in determining the grade of recommendation within clinical guidelines[2]. Meta-analysis can provide answers to important clinical questions and can be used to guide the management of individual patients[3]. Bias is defined as 'predisposition towards' or 'prejudice' by the Oxford English Dictionary[4]. Thus, the risk of bias in a RCT can be defined as the risk of predisposition towards the experimental intervention group or the control group. Bias in RCTs may overestimate or underestimate the true benefits and harms of an intervention (bias in effect estimate)[5–8].

Various scales and checklists have been suggested for the assessment of risk of bias[9,10]. It has, however, become evident that risk of bias should not be assessed by scales. Rather, certain design components ought to be evaluated in each trial[11]. Recently, *The Cochrane Handbook for Systematic Review of Interventions*[11] as well as The Cochrane Hepato-Biliary Group module[12], which provide the guidelines for preparation of reviews registered with The Cochrane Hepato-Biliary Group, were updated. Both sources recommend using various domains or components for assessment of bias risk. The main components are randomization, blinding, bias due to incomplete outcome data, bias due to selective outcome reporting, baseline imbalance bias, early stopping bias, academic bias and source of funding bias. The summary of the different domains of bias risk assessment and the risks assessed by each component are outlined in *Table 1*. Further details of these domains and biases, and the methods that may be used to avoid them, are provided in this review. Situations in which it is impossible or difficult to achieve low risk of bias in some components, and the dilemma that a researcher faces during trial design to achieve low risk of bias in these components, are also considered.

In parallel trials, each participant is randomly allocated to an intervention or control group. In crossover trials, all the participants receive both treatments (intervention and control) in sequence with a 'washout' period between the two treatments in order to allow the effect of one

Table 1 Summary of domains and the bias related to each domain

Domain assessed	Bias risk assessed	Description
Randomization	Selection bias	Systematic differences between baseline characteristics of the groups that are compared
Blinding	Detection bias	Systematic differences between groups in how outcomes are determined
	Performance bias	Systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest
Incomplete outcome data	Attrition bias	Systematic differences between groups in withdrawals from a study
Selective outcome reporting	Reporting bias	Systematic differences between reported and unreported findings
Baseline imbalance	May indicate selection bias	An imbalance in baseline characteristics may indicate flaws in randomization
	May raise questions related to effect estimate	An imbalance in prognostic factors leads to the question of whether the effect estimate is because of the treatment effect or because of the difference in prognostic factors
Early stopping	Early stopping bias	Trials are stopped at a point when the treatment effect is high at random when an insufficient number of outcome measures have been achieved
Academic	Academic bias	A bias towards finding the same result if the clinical trial is repeated in a new group of patients
Source of funding	Detection bias Performance bias Reporting bias Early stopping bias	See descriptions above. The results tend to become favourable to the sponsor's product

treatment to stop. The order in which the treatments are given is randomized. In cluster randomized trials, a group of patients is randomly allocated to intervention or control. For example, general practitioners may be randomized to intervention or control. All patients registered with a general practitioner allocated to intervention will receive the same interventional treatment. Patients visiting general practitioners allocated to control will all act as controls. Crossover cluster randomized trials are also possible. As most surgical trials are of parallel design, only the guidelines pertaining to this design are discussed here.

Randomization process

Only a quarter of surgical trials report the randomization process[13]. This process consists of two separate components, generation of the allocation sequence and allocation concealment. The primary aim of randomization is to ensure that the same sort of participants receive each intervention. If future allocation can be predicted, selection bias may result[11]. Take for example a trial comparing surgical resection and chemoradiotherapy for locally advanced pancreatic cancer. If future assignment can be predicted because of improper sequence generation, such as alternation (first patient receives resection, second patient receives chemoradiotherapy, third patient receives resection, fourth patient receives chemoradiotherapy, and so on) or improper allocation concealment, such as use of an open allocation list (an open list that contains the sequence in which the patients will be allocated to different treatments), patients who may not be suitable for resection because of involvement of a vital structure may be enrolled into the trial if the next 'turn' is chemoradiotherapy. This may overestimate the survival benefit of surgery. Thus, it is necessary to avoid any anticipation of future assignments. This can be achieved by an appropriately conducted randomization. The commonly used methods include sequence generation using a computer or a table of random numbers. Concealment of the sequence can be achieved by using opaque, sequentially numbered, sealed envelopes or by having the sequence held by a third party not involved in the treatment or assessment of the patient[11]. In fact, the current recommendation is to avoid envelope randomization (which can easily lead to violation of the allocation sequence) and conduct centralized randomization using a third party[11].

A special mention must be made of blocked randomization in unblinded trials. Blocked randomization involves dividing the sample into different blocks to ensure a particular allocation ratio[11] (such as 50 per cent intervention and 50 per cent control or 66 per cent intervention and 34 per cent control). For example, if a total sample size of 20 patients is required in equal allocation ratio (50 per cent intervention and 50 per cent control), they can be divided into five blocks of four patients each. Of these four patients, two can be allocated to intervention and two to control by

ensuring that the randomization list contains two interventions and two controls for each sequence of four patients. However, when there is a lack of blinding, it is possible to predict the future assignments with certainty in between 25 and 50 per cent of all assignments. The ability to predict future assignments decreases with increasing block size; yet, this may constitute a significant proportion of assignments. In a situation where blinding is not possible when blocked randomization is required, the use of blocks of variable size and large block sizes will decrease the risk in allocation concealment.

Another note of caution concerns the minimization process. Minimization has been suggested as an alternative to randomization[14]. The purpose of minimization is to balance the intervention and control groups for important prognostic factors[14]. The minimization process has been explained in a simple and detailed manner by Scott and colleagues[14]. The use of minimization is considered to be of low bias risk in the sequence generation domain[11]. However, in unblinded single-centre trials, it is possible to predict allocation easily if the method of minimization is known. This may introduce bias.

Blinding of participants, treating personnel and outcome assessors

Blinding refers to the process by which study participants, treating personnel and outcome assessors are kept unaware of intervention allocations after inclusion and randomization of participants into a trial[12]. Lack of patient blinding may result in differences in the measurement of patient-reported outcomes (such as quality of life and pain)[12]. Take, for example, a patient receiving wound infiltration with local anaesthetic in a trial assessing the role of such analgesia in laparoscopic cholecystectomy. The patient may not complain of wound pain if he or she already knew that local anaesthetic wound infiltration had been used. Such a lack of patient blinding might result in a biased measurement of pain. If the nurse who administers analgesia knows that the patient had local anaesthetic wound infiltration, there may be an inclination to administer less analgesia. A small difference in the way questions are asked ('do you need pain relief?' or 'don't you need pain relief?' carry different weights in recommending pain relief to the patient) or a small comment ('I am surprised that you need so much pain relief!') may result in a change in outcomes. Such a lack of observer blinding might result in a biased measurement of pain (detection bias).

Blinding has also been recommended to decrease the difference in the care provided or performance bias[11]. Take for example a surgeon involved in a trial assessing the use of routine abdominal drainage after liver resection. If the surgeon knew already that a drain was not going to be placed, he may take more care in dissection or haemostasis than if he knew that a drain was going to be used. This may indirectly benefit the group without drainage. Such a lack of surgeon blinding might result in bias. Another example would be a surgeon performing early laparoscopic cholecystectomy for acute cholecystitis. The threshold for further investigation of abdominal pain may be different to that in a patient undergoing delayed (elective) laparoscopic cholecystectomy for acute cholecystitis, resulting in a higher rate of complications being diagnosed in the early laparoscopic cholecystectomy group. Thus, lack of surgeon blinding can result in bias.

It is possible to achieve satisfactory blinding of all three groups (patients, doctors and nurses, and outcome observers) by the use of placebos in trials assessing pharmacological interventions. However, successful blinding of all three groups is difficult or impossible in surgical trials. In many situations, such as comparisons involving surgery in one group only, or a difference in the timing of surgery, blinding the patient is difficult. Examples are trials comparing laparoscopic cholecystectomy with no cholecystectomy for gallbladder dyskinesia, and comparing early with delayed laparoscopic cholecystectomy.

The only way in which blinding of patients could be achieved in the first situation would be to perform sham operations in the no-cholecystectomy group. In the second example, all the patients would have to undergo sham operations (the early group would have a delayed sham operation and the delayed group would have an early sham operation). The sham operation would involve a skin-deep umbilical scar, a skin-deep upper abdominal scar and two subcostal scars under local anaesthetic. The patients would have to be sedated for about 30–45 min. The treatment group would also need to have local anaesthetic wound infiltration so that it would not be possible to identify the group by finding out if there was local anaesthetic wound infiltration or not. Routine antiemetics would be used to prevent patients identifying the groups because of the nausea induced by pain and anaesthetic agents. Similarly, routine analgesia would be needed in both groups for the first few days after operation to avoid patients identifying the groups to which they belong. As the incisions are only skin deep (without incising the peritoneum), it is unlikely that future definitive laparoscopic surgery would affected by the sham operation. However, patients may prefer not to have scars that did not result in a definitive treatment of their symptoms. Furthermore, the risks of sham surgery include both minor complications, such as wound infection

and postoperative pain, and serious complications, such as aspiration and myocardial infarction. Although the risk of serious complications is low in this scenario, there might be other instances when the sham operations would carry a high risk of serious complications. It is obviously unethical to subject patients to sham operations with risks of complications.

Blinding of the surgeon is clearly impossible in many surgical trials. However, in trials involving assessment of abdominal drainage, the drain can be placed by a second surgeon who is not otherwise involved in the operation. This has to be combined with outcome assessor blinding to decrease the risk of bias owing to lack of blinding. This could also be applied to other situations, such as the assessment of ischaemic preconditioning in liver resection or in liver transplantation. Surgeon blinding, though problematic, is achievable in a number of situations.

Less than half of surgical trials report blinded assessment of outcomes[13]. This is because the surgeon who performed the operation is also often the outcome assessor. Although other members of the team, such as nurses or research assistants, can assess some outcomes, they cannot assess outcomes that involve alteration of patient management based on the outcome assessed. For example, the threshold for requesting ultrasonography for pain in patients undergoing early laparoscopic cholecystectomy may be different from that in patients undergoing delayed laparoscopic cholecystectomy because of the perceived increased risk of bile duct injury with the former technique. This may result in the detection of small intra-abdominal collections, which may not otherwise have been diagnosed. By involving another surgeon (not involved in the operation and blinded to the intervention provided) in the postoperative assessment of the patient, it is possible to achieve assessor blinding. However, this would require a very good understanding and working relationship between the surgeons if significant postoperative morbidity is to be avoided.

It may be difficult or impossible to achieve blinding of one or more groups. Fortunately, the bias due to lack of blinding can be minimized by using objective outcomes[8], which cannot be easily influenced by the patient or the investigator (such as all-cause mortality, recognized laboratory tests[8], prespecified criteria for investigation, prespecified criteria for treatment based on laboratory tests, or radiological investigations interpreted by doctors blinded to the groups). Currently, there is no strong evidence that lack of blinding increases bias in trials using objective outcomes[8]. Only 65 per cent of surgical trials, however, define the main outcomes[13]. In any event, the definition of these outcomes must be decided before the trial starts (when no universal definition exists for the outcomes) in order to avoid any potential bias resulting from the definition of outcomes after the observations have been recorded.

Despite all the precautions to decrease bias resulting from lack of blinding, this is sometimes unachievable. Consider, for example, laparoscopic cholecystectomy *versus* no cholecystectomy for gallbladder dyskinesia. Not performing a cholecystectomy does not result in serious morbidity as the main reason for cholecystectomy is relief of symptoms. This is a subjective measure in a circumstance in which it would be difficult to blind the patients. A sham operation would be the only way to achieve blinding of patients. Thus, one must balance the risk of providing treatment based on biased trials with an acceptance of a more complex trial design to obtain a result as free as possible from bias introduced by the lack of blinding.

Incomplete outcome data

This occurs as a result of exclusion of patients from the analysis after randomization (attrition bias), particularly if the reasons for exclusion are related to the treatment[15]. The easiest way to avoid this bias is to avoid withdrawals after randomization. This is usually feasible in surgical trials in which short-term outcomes are evaluated, and the interval between randomization and treatment is short. For example, in a trial assessing the routine use of abdominal drainage in laparoscopic cholecystectomy, randomization can be performed towards the end of surgery in order to avoid postrandomization dropouts owing to conversion to open cholecystectomy.

Attrition bias is common in surgical trials. For example, in a trial comparing low-pressure pneumoperitoneum with standard-pressure pneumoperitoneum for laparoscopic cholecystectomy, if the patients who underwent conversion to open cholecystectomy are excluded without mentioning the reasons for conversion, it is not safe to make assumptions about the safety of performing laparoscopic cholecystectomy using low-pressure pneumoperitoneum. The reason for conversion to open cholecystectomy may be a bile duct injury. It is also important to note that all postrandomization dropouts do not result in attrition bias. For example, in a trial comparing day-care laparoscopic cholecystectomy with overnight laparoscopic cholecystectomy, some patients may be excluded from analysis after randomization because they did not undergo the surgery as planned (some patients may have urgent surgery because of gallstone complications). This is not related to the treatment (provided that the inclusion and exclusion criteria are applied strictly while recruiting the

patients). Any conclusion from such a trial or a systematic review of such trials is applicable only to patients who underwent successful completion of elective laparoscopic cholecystectomy. The aim of the trial or the systematic review was to assess the safety and effectiveness of day-case surgery in elective laparoscopic cholecystectomy[16,17]. So there is unlikely to be attrition bias in spite of postrandomization dropouts. However, the assumption that there is no attrition bias in spite of postrandomization dropouts requires all the possible mechanisms for such dropouts to be known. This is not something about which one can ever be certain and so every effort to avoid postrandomization dropouts should be made.

Postrandomization drop-outs have a direct impact on the method of analysis used. In intention-to-treat analysis, all patients randomly allocated to one of the treatments in a trial are analysed together as representing that treatment, whether or not they completed, or indeed received, that treatment[18]. When there are postrandomization dropouts, it may or may not be possible to measure the outcomes. For example, in the trial comparing day-case with overnight elective laparoscopic cholecystectomy, one of the reasons for dropouts could be that the patient did not undergo laparoscopic cholecystectomy as planned, in which case the outcome postoperative pain cannot be measured. This is called missing data. Another reason for dropouts could be conversion to open cholecystectomy where the outcome postoperative pain can be measured. By following an intention-to-treat analysis, pain scores from these patients should also be included in the analysis. When data are missing, various imputation methods have been suggested[19].

An alternative to intention-to-treat analysis is available-case analysis. Here, data are analysed for every participant for whom the outcome was obtained[11], and the patients are analysed according to the group to which they were randomized. In the above example, if data on the patients who did not undergo laparoscopic cholecystectomy were not available, an available-case analysis may be more appropriate than intention-to-treat analysis with imputation, if the aim was to assess the safety and effectiveness of day-case laparoscopic cholecystectomy.

Two other analyses, per-protocol analysis (only participants who completed the trial and who received the allocated treatment are included in the analyses)[11] and treatment-received analysis (participants are analysed according to the interventions received irrespective of the allocation)[11] should be avoided in superiority trials (trials designed to identify superiority of one treatment over another). Take, for example, a trial comparing laparoscopic cholecystectomy with open cholecystectomy. If there were bile duct injuries (recognized at operation) in the laparoscopic cholecystectomy group, the procedure would usually be converted to open cholecystectomy. These patients with major complications would be excluded by following a per-protocol analysis. Worse still, they would be included under open cholecystectomy in a treatment-received analysis. An illustration of the four different methods of analysis from a meta-analysis of early *versus* delayed laparoscopic cholecystectomy for acute cholecystitis[20] is shown in *Tables 2* and *3*.

In non-inferiority trials (designed to investigate whether a treatment is not inferior to another treatment) or equivalence trials (designed to investigate whether a treatment is therapeutically similar to another treatment), per-protocol analysis may be better than intention-to-treat analysis in specific situations[21]. However, the ethical nature of non-inferiority and equivalence trials is questionable[22].

Selective reporting bias

Selective reporting bias is the bias that results from authors reporting only statistically significant beneficial outcomes and excluding statistically non-significant outcomes or

Table 2 Incomplete outcome data: conversion rates (to open cholecystectomy) in trials comparing early and delayed laparoscopic cholecystectomy[20]

	Conversion rate				
Study	ELC	DLC	Crossover*	Elective†	Dropouts (ELC *versus* DLC)
Davila 1999	1 of 27 (4)	6 of 36 (17)	4 of 5 (80)	2 of 31 (6)	Not reported
Johansson 2003	23 of 74 (31)	20 of 69 (29)	10 of 18 (56)	10 of 51 (20)	0 *versus* 2
Kolla 2004	5 of 20 (25)	5 of 20 (25)	n.a.	5 of 20 (25)	No dropouts
Lai 1998	11 of 53 (21)	11 of 46 (24)	2 of 8 (25)	9 of 38 (24)	0 *versus* 5
Lo 1998	5 of 48 (10)	9 of 45 (20)	2 of 9 (22)	7 of 36 (19)	1 *versus* 5
All studies	45 of 222 (20·3)	51 of 216 (23·6)	18 of 40 (45)	33 of 176 (18·8)	1 *versus* 12

Values in parentheses are percentages. *Those belonging to the delayed laparoscopic cholecystectomy (DLC) group who had worsening, non-resolution or recurrence of acute cholecystitis. †Those belonging to the DLC group who were successfully managed conservatively. ELC, early laparoscopic cholecystectomy; n.a., not applicable (no crossover).

Table 3 Incomplete outcome data: meta-analysis of conversion to open cholecystectomy in early and delayed laparoscopic cholecystectomy[20] using different methods of dealing with postrandomization dropouts

Type of analysis	Relative risk	Group favoured by trend
Intention-to-treat analysis (good outcome)	0·88 (0·62, 1·25)	ELC
Intention-to-treat analysis (poor outcome)	0·73 (0·52, 1·01)	ELC
Intention-to-treat analysis (best case for ELC)	0·71 (0·51, 0·99)*	ELC*
Intention-to-treat analysis (worst case for ELC)	0·90 (0·63, 1·28)	ELC
Available-case analysis	0·84 (0·59, 1·19)	ELC
Per-protocol analysis	1·03 (0·69, 1·54)	DLC
Treatment-received analysis	1·23 (0·85, 1·80)	DLC

Values in parentheses are 95 per cent confidence intervals. ELC, early laparoscopic cholecystectomy; DLC, delayed laparoscopic cholecystectomy. *Statistically significant.

harmful outcomes. It is not safe to assume that the authors did not measure this outcome. Selective reporting can lead to bias in the effect estimate (an incorrect assumption that a treatment is more effective or less effective than the true effectiveness) as neutral or negative results are not reported. It could be the complete absence of reporting of a particular outcome or it could take the form of incomplete details, which cannot be included in a meta-analysis[11]. Selective reporting can sometimes invalidate the results of a trial. For example, in a trial assessing the role of palliative liver resection for neuroendocrine liver metastases, if the relief of symptoms but not the surgery-related morbidity was reported, concerns about the safety and effectiveness of the treatment will remain. In fact, selective outcome reporting may overestimate intervention effects by up to 100 per cent[23].

Baseline imbalance

In large trials (say more than 400 patients[24]), it is expected that the randomization process (if performed properly) will result in patients matched for important characteristics in the randomized groups. In smaller trials, however, this may not be achieved. Baseline imbalance may be in the characteristics of patients or the experience of the surgeons. An imbalance in the baseline characteristics may raise doubts about the effectiveness and the applicability of the intervention. For example, in a trial comparing survival after liver resection for colorectal liver metastases, assume that there was an imbalance in the proportion of synchronous liver metastases (present at the time of diagnosis of colorectal cancer) and metachronous liver metastases (appearing some time after diagnosis of colorectal cancer). The prognosis following liver resection for metachronous disease is better than that for synchronous metastases[25], and the difference between the treatment groups may be due either to the treatment or to the prognostic difference between synchronous and metachronous liver metastases. Baseline imbalance may be due to improper randomization methods[11] or simply to the 'play of chance'. This is why stratification for a few important prognostic factors is used at randomization of small trials, typically with fewer than 400 patients[24]. Minimization is an alternative way to reduce baseline imbalance[14]. Statistical methods such as regression analysis may have to be used to adjust for any baseline imbalance that arises in spite of stratification or minimization.

Early stopping bias

Sample size calculation is reported in less than 25 per cent of RCTs[13,26,27]. It is not possible to determine if recruitment to the trial was stopped early unless sample size calculations are reported. Early stopping of a trial may result in bias in the effect estimate[28]. This is particularly true of trials stopped for beneficial effect of the intervention. In the course of a trial, the early results may favour the intervention, but may even out in the long run. However, early stopping of the trial (even with formal stopping results) stops the trial when the observed effectiveness is at its greatest[28]. This may not be the true effect estimated when all patients planned for enrolment become randomized. By including such trials in a meta-analysis, a bias can be introduced in the effect estimate.

Academic bias

If the same comparison is made in two trials (without any difference in the intervention, inclusion criteria, exclusion criteria or methods of assessment) by the same author, a bias may be introduced in the effect estimate. For example, if the author performs a pilot RCT and finds interesting results warranting a larger trial, the pilot study may have been stopped at a point where the observed effectiveness is at its greatest. Similar positive results may not be observed when the larger trial is carried out, leading the investigator to chase the expected result through multiple statistical analyses, including subgroup analyses. All researchers know that it is difficult to keep a proper academic disinterest regarding their own results. Publication of the trial protocol

before the trial starts in a trials register[29], for example at http://www.clinicaltrials.gov/, will help in the assessment of this type of bias.

Source of funding bias

Less than 50 per cent of surgical trials declare their source of funding[13]. If the funding came from a party that has a vested interest in the success of the intervention, there is a possibility of bias in the effect estimate or in interpretation of the data[30,31]. Funding may have been withdrawn at the point where the observed effectiveness is at its greatest, resulting in early stopping bias. Only positive results may be reported, resulting in selective reporting bias. Study design, including the use of inappropriate comparator treatment and lack of peer-reviewed publication of results, is another possible reason for source of funding bias[31]. However, not all trials sponsored by a party that has a vested interest in the success of the intervention are subject to this bias. Sometimes the authors report that an intervention is neither safe nor effective even if the work was sponsored by the manufacturer of the intervention. For example, an author of a trial assessing the role of a radiofrequency dissecting sealer in liver resection concluded that it resulted in more complications and was not effective in reducing transfusion requirements, despite the fact the trial was sponsored by the manufacturer of the radiofrequency dissecting device[32]. Publication of the protocol of the trial should help as in the identification of academic bias[29]. By comparing the trial report with the protocol, it should be possible to determine whether early stopping or selective reporting bias exists. The other domains (such as sequence generation, allocation concealment, blinding, missing outcomes and baseline imbalance) may also give a clue to the validity of such a trial.

Overview

It is possible to conduct surgical trials that will be classified as at low risk of bias as outlined in *The Cochrane Handbook for Systematic Review of Interventions*[11]. Blinding, however, remains a major challenge in surgical trials. Various strategies, such as sham operations, blinded assessment of outcomes or use of a second surgical team, may be adopted to minimize this risk of bias. A balance must be struck between the risk of providing a treatment based on a biased effect estimate with the difficulty of conducting trials with very low risk of bias. In situations where postrandomization dropouts cannot be avoided, an intention-to-treat analysis and available-case analysis are appropriate. Differential expertise bias[33] (surgeon better trained in control than intervention, for example, comparison of robot-assisted surgery with human-assisted surgery) and bias owing to use of composite outcomes (several endpoints)[34] may produce major errors. Various guidelines exist for the conduct and reporting of RCTs[35–41]. A better understanding of the risks of bias may result in an improved conduct and reporting of trials, with a closer estimate of the true effectiveness of an intervention as a result. Ultimately, this must benefit the patient.

Acknowledgements

The authors declare no conflict of interest.

References

1 Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; **323**: 334–336.
2 Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S *et al*. Grading quality of evidence and strength of recommendations. *BMJ* 2004; **328**: 1490.
3 Bhandari M, Devereaux PJ, Montori V, Cina C, Tandan V, Guyatt GH; Evidence-Based Surgery Working Group. Users' guide to the surgical literature: how to use a systematic literature review and meta-analysis. *Can J Surg* 2004; **47**: 60–67.
4 Oxford University Press. *Oxford English Dictionary – The Definitive Record of the English Language*. http://www.oed.com [accessed 7 September 2008].
5 Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001; **135**: 982–989.
6 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**: 609–613.
7 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**: 408–412.
8 Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG *et al*. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; **336**: 601–605.
9 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**: 1–12.
10 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995; **16**: 62–73.

11 Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0* (updated February 2008). http://www.cochrane-handbook.org [accessed 7 September 2008].
12 Gluud C, Nikolova D, Klingenberg S, Alexakis N, Als-Nielsen B, D'Amico G *et al*. The Cochrane Hepato-Biliary Group. About The Cochrane Collaboration (Cochrane Review Groups (CRGs)). In *The Cochrane Library* 2008; Issue (4).
13 Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery* 1996; **119**: 466–472.
14 Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. A review. *Control Clin Trials* 2002; **23**: 662–674.
15 Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *Int J Epidemiol* 2005; **34**: 79–87.
16 Gurusamy K, Junnarkar S, Farouk M, Davidson BR. Meta-analysis of randomized controlled trials on the safety and effectiveness of day-case laparoscopic cholecystectomy. *Br J Surg* 2008; **95**: 161–168.
17 Gurusamy KS, Junnarkar S, Farouk M, Davidson BR. Day-case *versus* overnight stay for laparoscopic cholecystectomy. *Cochrane Database Syst Rev* 2008; (3)CD006798.
18 Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992; **21**: 837–841.
19 Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials* 2008; **5**: 225–239.
20 Gurusamy KS, Samraj K. Early *versus* delayed laparoscopic cholecystectomy for acute cholecystitis. *Cochrane Database Syst Rev* 2006; (4)CD005440.
21 Matilde Sanchez M, Chen X. Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Stat Med* 2006; **25**: 1169–1181.
22 Garattini S, Bertele' V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007; **370**: 1875–1877.
23 Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007; **297**: 468–470.
24 Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol* 1999; **52**: 19–26.
25 Viganò L, Ferrero A, Lo Tesoriere R, Capussotti L. Liver surgery for colorectal metastases: results after 10 years of follow-up. Long-term survivors, late recurrences, and prognostic role of morbidity. *Ann Surg Oncol* 2008; **15**: 2458–2464.
26 Gluud C. The culture of designing hepato-biliary randomised trials. *J Hepatol* 2006; **44**: 607–615.
27 Slim K, Bousquet J, Kwiatkowski F, Pezet D, Chipponi J. Analysis of randomized controlled trials in laparoscopic surgery. *Br J Surg* 1997; **84**: 610–614.
28 Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. *J Clin Epidemiol* 2008; **61**: 241–246.
29 Wille-Jorgensen P, Gluud C. Prospective registration of clinical trials. *Colorectal Dis* 2006; **8**: 1.
30 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003; **290**: 921–928.
31 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; **326**: 1167–1170.
32 Lesurtel M, Selzner M, Petrowsky H, McCormack L, Clavien PA. How should transection of the liver be performed?: a prospective randomized study in 100 consecutive patients: comparing four different transection strategies. *Ann Surg* 2005; **242**: 814–822.
33 Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S *et al*. Need for expertise based randomised controlled trials. *BMJ* 2005; **330**: 88.
34 Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM *et al*. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007; **334**: 786.
35 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008; **148**: W60–W66.
36 Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008; **148**: 295–309.
37 Gagnier JJ, Boon H, Rochon P, Moher D, Barnes J, Bombardier C; CONSORT Group. Reporting randomized, controlled trials of herbal interventions: an elaborated CONSORT statement. *Ann Intern Med* 2006; **144**: 364–367.
38 Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG *et al*. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008; **5**: e20.
39 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K *et al*. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; **141**: 781–788.
40 Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; **285**: 1987–1991.
41 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of non-inferiority and equivalence randomized trials: an extensionof the CONSORT statement. *JAMA* 2006; **295**: 1152–1160.