



Review

Bias in Clinical Intervention Research

Lise Lotte Gluud

From the Copenhagen Trial Unit, Center for Clinical Intervention Research, Cochrane Hepato-Biliary Group, Copenhagen, Denmark.

Received for publication January 26, 2005; accepted for publication October 27, 2005.

Research on bias in clinical trials may help identify some of the reasons why investigators sometimes reach the wrong conclusions about intervention effects. Several quality components for the assessment of bias control have been suggested, but although they seem intrinsically valid, empirical evidence is needed to evaluate their effects on the extent and direction of bias. This narrative review summarizes the findings of methodological studies on the influence of bias in clinical trials. A number of methodological studies suggest that lack of adequate randomization in published trial reports may be associated with more positive estimates of intervention effects. The influence of double-blinding and follow-up is less clear. Several studies have found a significant association between funding sources and pro-industry conclusions. However, the methodological studies also show that bias is difficult to detect and appraise. The extent of bias in individual trials is unpredictable. A-priori exclusion of trials with certain characteristics is not recommended. Appraising bias control in individual trials is necessary to avoid making incorrect conclusions about intervention effects.

bias (epidemiology); evidence-based medicine; meta-analysis; publication bias; selection bias

Abbreviations: CI, confidence interval; ROR, relative odds ratio; SSRIs, selective serotonin reuptake inhibitors.

Bias in clinical trials may be described as systematic errors that encourage one outcome over others. The potential effect of bias is that investigators will come to the wrong conclusions about the beneficial and harmful effects of interventions. Several mechanisms may bias clinical trials, affecting the estimated intervention effects. Certain trial characteristics have been suggested as quality components that may be used when appraising the control of bias in clinical trials. Although they seem intrinsically valid, empirical evidence is necessary to make conclusions regarding the effects of the individual components.

The present paper is a narrative review of methodological studies of bias in clinical trials. The included studies were identified through manual and electronic searches. The manual searches included scanning of bibliographies, journals, and conference proceedings and correspondence with experts. The electronic searches were performed in the

Cochrane Library, the Cochrane Hepato-Biliary Group Controlled Trials Register, MEDLINE, and the Excerpta Medica Database (EMBASE). No limitations were used for year of publication or language. The keywords used in the electronic searches included “quality,” “bias,” “trials,” and “random*.” The searches were completed on February 21, 2004.

THE QUALITY OF BIAS CONTROL IN RANDOMIZED TRIALS

The quality of clinical trials may be defined as the confidence that the design, conduct, report, and analysis restrict bias in the intervention comparison (1, 2). At least 25 quality scales have been developed, but few have been validated using established criteria (3, 4). Scales are useful for summarizing the quality of trial cohorts, but the combination of

Correspondence to Lise Lotte Gluud, Department 7102, Copenhagen Trial Unit, Cochrane Hepato-Biliary Group, Rigshospitalet, DK-2100 Copenhagen, Denmark (e-mail: lgluud@ctu.rh.dk).

quality components may be problematic. The extent and direction of the effect of individual components may vary, and the conclusions of different scales often disagree (4). If a significant association between quality and intervention effects is found, separate analyses of the individual components are necessary. Therefore, the use of quality scales is not generally recommended for assessment of bias control in individual trials (5). Instead, separate quality components may be used (2–9).

Randomization

Evidence-based medicine involves identification of the most reliable research evidence (10, 11). The reliability of the evidence depends on several aspects, including the risk of bias associated with different research designs (12). Uncontrolled clinical observations provide reliable evidence if interventions have dramatic effects along the lines of the effect of insulin in diabetic ketoacidosis. When intervention effects are moderate or small, the human processing of data, unsystematic data collection, and the human capacity to overcome illnesses spontaneously limit the value of uncontrolled observations. Experimental models are essential for estimation of toxicity and pathophysiology. The main problem of assessing intervention effects in experimental models lies in the necessary extrapolation. Previous examples show how extrapolation may lead to the wrong conclusions. One example is β -blockers, which reduce mortality in congestive heart failure in spite of a negative inotropic effect (13). Thalidomide, which has species-specific teratogenic effects that were overlooked in experimental models (14), is another example.

Observational studies are important in the evaluation of rare adverse events (15–17). If a retrospective study design is used, bias due to periodical changes, recall bias, and differential measurement errors may occur (18–20). Bias related to confounding by indication may occur in prospective studies as well as retrospective studies. Confounding by indication occurs when patients are allocated to the intervention or control group on the basis of patient and investigator preferences, patient characteristics, and clinical history (21).

On average, the prognostic factors of patients in the experimental and control groups differ. The result may be selection bias, which occurs when prognostic factors are unevenly distributed between the experimental group and the control group. Selection bias often means that patients in the control group have less favorable outcomes than patients in the experimental group. The purpose of randomization is to reduce such bias by creating comparison groups that are similar with regard to known as well as unknown prognostic variables.

The effect of randomization may be deduced from comparisons of randomized trials and observational studies. A number of methodological studies include such analyses. One of these studies found considerable differences between estimated intervention effects in 15 of 22 comparisons of randomized and nonrandomized trials (22). Analyses of all comparisons showed that the effect estimates in the nonrandomized trials ranged from an underestimation of effect

TABLE 1. Cohort studies on the sample size of randomized controlled trials in different disease areas

Disease area (ref. no.)	No. of trials	Sample size	
		Median	Interquartile range
Hepatology (32)	235	52	28–88
Dermatology (33)	68	46	30–80
Gastroenterology (36)	385	54	24–110
Sclerosis (34)	73	28	17–43
Intensive care medicine (35)	173	30	20–64
Radiology (37)	130	61	27–104

of 76 percent to an overestimation of effect of 160 percent (22). In another methodological study, odds ratios generated by 168 observational studies and 240 randomized trials on 45 topics were compared (23). All trials and studies were included in a meta-analysis with binary outcomes. Overall, the observational studies tended to generate larger summary odds ratios, suggesting a more beneficial effect of the intervention.

Bias associated with nonrandom allocation was also analyzed in a review summarizing the results of 69 studies that compared randomized trials with observational studies (21). The review found that nonrandom allocation was related to overestimation as well as underestimation of treatment effects. The variation in the results of observational studies was increased because of haphazard differences in the case mix between groups. Four strategies for case-mix adjustment were subsequently evaluated by generating nonrandomized studies from two large randomized trials (21). Participants were resampled according to allocated treatment, center, and period. None of the strategies adequately adjusted for bias in historically or concurrently controlled studies. Logistic regression was found to increase bias due to misclassification and measurement errors in confounding variables and differences between conditional and unconditional odds ratio estimates of treatment effects.

The large randomized trial is one of our most reliable sources of evidence for assessment of intervention effects (6, 24–28). Disagreements often occur between meta-analyses and large trials and between large trials on the same topic (24–26, 29, 30). Discordance rates seem to depend on whether primary outcomes or secondary outcomes are assessed. In many cases, there are no large trials but several small trials with low statistical power (31). Table 1 lists six cohort studies on the sample size of randomized trials in different disease areas (32–37). The median sample size in these studies ranged from 28 patients to 61 patients, suggesting that relatively large intervention effects may have been overlooked. In these situations, it may be useful to combine the results of the individual trials in a meta-analysis in order to increase statistical power and evaluate bias (38–40).

Adequate randomization requires that the allocation of the next patient be unpredictable (41–43). If the next assignment is known, enrollment of certain patients may be prevented or delayed to ensure that they receive the treatment

TABLE 2. Cohort studies on reported randomization methods in randomized trials

Disease area (ref. no.)	No. of trials	% with adequate allocation sequence generation	% with adequate allocation concealment
Gynecology/obstetrics (48)	206	32	26
Hepatology (47)	166	28	23
Hepatology (32)	235	52	37
Dermatology (33)	68	1	34
Intensive care medicine (35)	173	27	7
Gastroenterology (36)	383	42	6
Orthodontics (49)	155	50	39

believed to be superior (44). Adequate randomization involves both generation of an allocation sequence and concealment of allocation. The simplest method of generating an allocation sequence is to give each patient an equal chance of receiving either treatment. A table of random numbers or a random number generator on a computer may be used. Modifications include block randomization to ensure equal numbers in comparison groups and stratified randomization, to keep comparison groups balanced for known prognostic factors (43). Irrespective of how the allocation sequence is generated, bias may be introduced if the allocation of the next patient is known. Therefore, the allocation sequence must be concealed. Adequate allocation concealment may be achieved through the use of independent centers or serially numbered identical sealed packages. The use of sealed envelopes is another popular method, but previous evidence demonstrates that envelopes may be transilluminated or opened before or after patients are excluded (45, 46). Whether sealed envelopes provide adequate allocation concealment is debatable.

Many trials that are described as randomized do not report randomization methods. Table 2 lists seven studies on reported randomization methods in different disease areas (32, 33, 35, 36, 47–49). The studies found that the reporting of randomization methods in different disease areas varies. The proportion of trials reporting adequate allocation sequence generation ranged from 1 percent to 52 percent (median, 37 percent). The proportion with adequate allocation concealment ranged from 2 percent to 39 percent (median, 25 percent). The variation between the studies may be related to the characteristics of the disease or classification randomization (32, 36, 50).

Sometimes, reported randomization methods do not reflect the methods that were actually used during the trial. In a study of 105 randomized trials, reported allocation concealment methods in full-text publications were compared with information obtained through direct communication with the authors (51). The study found that several trials had adequate allocation concealment methods that were not described in the published report (52). In a study on 56 randomized controlled trials conducted by the Radiation

Therapy Oncology Group (53), adequate allocation concealment was achieved in all trials but was reported for only 42 percent. On the other hand, a study comparing publications and protocol descriptions of allocation concealment in 102 trials found that most trials with unclear allocation concealment in the published paper also had unclear allocation concealment in their protocol (54). Accordingly, additional evidence on discrepancies between the conduct and reporting of randomized trials is needed.

Several previous studies have analyzed the association between reported randomization and intervention effects in trials from meta-analyses. Two of the initial studies included obstetrics and gynecology trials (2) or trials from various disease areas (9). Neither of the studies found a significant effect of the allocation sequence generation, but both found that inadequate allocation concealment was associated with significantly more positive (i.e., statistically significant) estimates of intervention effects. These findings suggest that inadequate allocation concealment may bias the results of clinical trials. On the other hand, the findings may also reflect bias due to selective publication of reports on small, low-quality trials with positive results (39, 55). The results of very large randomized trials rarely remain unpublished. Therefore, a subsequent study used similar methods but included large randomized trials ($n > 1,000$ participants) as a reference group (6). The study found that, on average, small trials without adequate allocation sequence generation or allocation concealment overestimated intervention benefits. Small trials with adequate randomization methods were not significantly different from the large trials.

The association between randomization and intervention effects was analyzed in subsequent cohort studies (5, 7, 8). On the basis of unsystematic searches in MEDLINE, manual searches of bibliographies and conference proceedings and correspondence with experts revealed six methodological studies comparing intervention effects in randomized trials that did or did not describe adequate randomization or double-blinding (56). The included studies analyzed randomized trials from meta-analyses assessing unwanted binary outcomes. All studies calculated the relative odds ratios by comparing summary odds ratios generated by groups of trials with or without adequate randomization or blinding. The relative odds ratios were combined in random-effects meta-analyses. As figure 1 shows, odds ratios were approximately 12 percent more positive in trials without adequate allocation sequence generation (relative odds ratio (ROR) = 0.88, 95 percent confidence interval (CI): 0.79, 0.99). Trials without adequate allocation concealment were approximately 21 percent more positive than trials with adequate allocation concealment (ROR = 0.79, 95 percent CI: 0.66, 0.93). The meta-analyses also showed considerable between-study heterogeneity that may have been related to the disease, the intervention, trial inclusion criteria, and classification of adequate randomization. The variation suggests that caution is necessary when making recommendations for quality assessments. Using certain components as exclusion criteria does not seem justified. The quality of individual trials and the quality of meta-analyses must be evaluated separately.

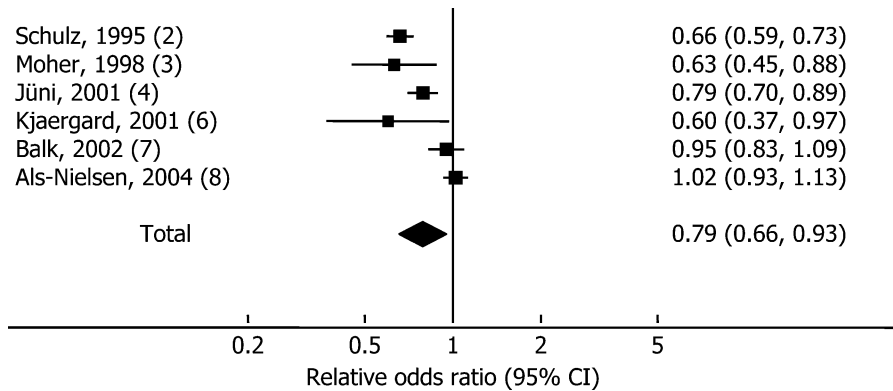


FIGURE 1. Forest plot of a random-effects meta-analysis of methodological studies calculating the relative odds ratio between groups of randomized trials with or without adequate allocation concealment. The squares show the point estimates for individual studies (horizontal bars, 95 percent confidence interval (CI)); the diamond shows the overall relative odds ratio from the meta-analysis.

Blinding

In clinical trials, the term “blinding” refers to keeping participants, health-care providers, data collectors, outcome assessors, or data analysts unaware of the assigned intervention (57). The purpose of blinding is to prevent bias associated with patients’ and investigators’ expectations (5). If interventions are compared with no intervention, an identical placebo may be used. The compared interventions must be identical in taste, smell, appearance, and mode of administration. Any difference may destroy the blinding (58–60). The terminology authors use to convey blinding status is open to various interpretations. Both physicians and textbooks vary in their definitions of single-, double-, and triple-blinding (57, 61). “Double-blinding” may refer to blinding of both participants and health-care providers, investigators, data collectors, judicial assessors, or data analyzers. Sometimes, the nature of an intervention precludes blinding of participants. Blinding of outcome assessors is generally possible and may theoretically be one of the most fundamental considerations.

In a study including 616 hepatobiliary randomized trials published during 1985–1996, 34 percent were described as double-blind (50). The variation may reflect not only that some interventions preclude double-blinding but also that certain trials may be performed without double-blinding, although blinding would have been feasible (62). A meta-analysis of methodological studies has analyzed the association between double-blinding and intervention effects (2, 5–9). The studies compared summary odds ratios in groups of trials that were or were not described as double-blind. Two of these studies found that trials not described as double-blind overestimated intervention effects compared with double-blind trials (2, 6). A random-effects meta-analysis of the six studies (56) found no significant association between the reported blinding and intervention effects (ROR = 0.86, 95 percent CI: 0.71, 1.05). The meta-analysis revealed considerable between-study variation that may have been related to the nature of the disease or the inter-

vention. The variation may reflect the fact that some interventions are difficult to blind. For example, if we perform double-blind trials on drugs associated with adverse events, blinding may be ineffective. The type of outcome may be equally important. Measurable clinical outcomes may be less prone to assessment bias than subjective outcomes. Therefore, trials evaluating the effect of drugs on, for example, mortality may be less susceptible to bias than trials evaluating the effect of drugs on pain. The effect of blinding is highly unpredictable, and separate analyses of the effect of blinding in individual trials and meta-analyses are necessary.

Follow-up

Protocol deviations are often related to prognostic factors and may lead to attrition bias (63–65). Attempts to obtain data on patients who have been lost to follow-up and clear descriptions of follow-up are important (66–68). Thirty percent of 235 randomized trials whose results were published in the journal *Hepatology* during 1981–1998 did not describe the numbers or reasons for dropouts and withdrawals (32). The association between the completeness of the reported follow-up and trial results was analyzed in two methodological studies (2, 6). One study found no significant differences in intervention effects generated by trials with losses to follow-up compared with trials with complete (explicit or assumed) follow-up (2). The other study found no significant association between reported follow-up and intervention effects (6). These results may suggest that reporting of follow-up was inadequate in several trials.

Intention-to-treat analyses include all patients, whereas per-protocol analyses exclude data from patients with protocol deviations (69). Therefore, intention-to-treat analyses must deal with missing data. Suggested strategies include carrying forward the last observed response or calculating the most likely outcome based on the outcome of other patients. Per-protocol analyses exclude patients with missing data from the analyses. If patients with missing data are mainly outliers, per-protocol analyses may increase

homogeneity and precision. On the other hand, if losses to follow-up are related to prognostic factors, adverse events, or lack of treatment response, per-protocol analyses may overestimate the intervention effects (65, 70). In general, the intention-to-treat analysis is the most reliable type for analyzing data from randomized trials. However, discrepancies between intention-to-treat and per-protocol analyses may provide important additional information, and using both analytical strategies may be considered.

In a cohort study of 199 randomized trials in which the authors stated that intention-to-treat analyses were used (68), few trials clarified the analytical strategies used, which seemed to vary considerably. Several analyses were described as intention-to-treat although patients were not analyzed as allocated. Therefore, the analytical strategy must be appraised carefully in individual trials, as the intention-to-treat approach may be inadequately applied.

COMPETING INTERESTS

The effect of competing interests is debated (71). On the one hand, industry funding has been associated with high quality as compared with trials without external funding (32, 50). On the other hand, financial interests may bias the interpretation of trial results. In a cohort study of 159 randomized trials, the interpretation of the results of individual trials was significantly more favorable towards experimental interventions if funding obtained from for-profit organizations was declared (72). The association was not related to quality or statistical power. Two systematic reviews found similar associations between funding and pro-industry conclusions (73, 74). The extent to which these findings reflect the quantitative trial results or the interpretation of trial results may be important. Therefore, a subsequent study with 370 drug trials analyzed the association between funding and conclusions reached after adjusting for the quantitative trial results and the occurrence of adverse events (75). Neither appeared to explain the association between funding and pro-industry conclusions.

The reason for the association between funding and pro-industry conclusions is not clear. Potential explanations include violation of the uncertainty principle (76), publication bias (55), and biased interpretation of trial results (75). The uncertainty principle means that patients should be enrolled in a trial only if there is substantial uncertainty about which of the treatments in the trial is most appropriate for the patient (77). Trials may be considered unethical if patients allocated to the control group are not offered a known effective intervention. A violation of the uncertainty principle may be related to selective sponsoring of trials with known beneficial effects (76).

In 1997, approximately 16 percent of 1,396 highly ranked scientific journals had policies on conflicts of interest (78). Less than 1 percent of the articles published in these journals contained disclosures of personal financial interests. The importance of disclosure of financial interests is increasingly being recognized, as demonstrated by the following examples of publication bias and neglect of harm.

- Selective serotonin reuptake inhibitors (SSRIs) have been prescribed for depression since the 1980s (79). The first trials of SSRIs were initiated in the late 1980s, but the results of several remained unpublished more than 10 years later (80, 81). During subsequent years, cases of suicidal behavior among children on SSRIs appeared (79). In 2003, a review of data from published and unpublished clinical trials prompted the Food and Drug Administration and the Committee on Safety of Medicines to warn against the use of SSRIs in children (82–85).
- The Vioxx Gastrointestinal Outcomes Trial suggested that rofecoxib was a safe alternative to antiinflammatory drugs for pain relief (86). From 2001 onward, independent researchers expressed their concerns about its potential cardiovascular effects (87, 88). The published report from the Vioxx Gastrointestinal Outcomes Trial did not address this question, but a review of unpublished safety data showed that rofecoxib increased the risk of serious cardiovascular thrombotic events (89, 90). In 2004, the manufacturer, Merck & Co., Inc. (Whitehouse Station, New Jersey), finally chose to remove Vioxx from the market (91).
- In 1995, the United Kingdom Department of Health commissioned a meta-analysis of individual patient data on the use of primrose oil supplementation for atopic dermatitis (92). For unknown reasons, the authors were not allowed to publish their work. G. D. Searle and Company (Pfizer, Inc., New York, New York), the company then responsible for marketing evening primrose oil in the United Kingdom, asked the authors to sign a written statement verifying that the contents of the report had not been leaked (92).

PUBLICATION BIAS AND RELATED BIASES

Prospective studies of clinical trials have analyzed the effect of unwanted results and publication of the findings of clinical trials (55, 93–97). Publication bias, with selective publication of the findings of trials with positive results, may lead to overestimation of treatment effects. One study examined the publication status of all studies submitted to an ethics committee over a 10-year period (96). Overall, studies with statistically significant results were more likely to be published than studies with nonsignificant results, and they had significantly shorter times to publication. In a similar study of 109 randomized efficacy trials, the time from completion to submission for publication was substantially shorter for trials with positive findings (98). In prospective analyses of clinical trial reports submitted for publication in the *Journal of the American Medical Association*, the chances of being published were not significantly different for trials with positive results and trials with negative results (99). Accordingly, publication bias may reflect a reluctance to submit reports on negative trials for publication rather than a reluctance to publish negative trial reports on the part of journals.

Selective or delayed publication of the findings of trials with unwanted results seems to be a widespread problem. Citation habits and similar dissemination biases may also

be important, but the evidence is not convincing (100–104). However, recent evidence suggests that selective reporting of statistically significant outcomes in published trial reports may be a common problem. Analyses of 102 protocols and publications (122 papers) showed that studies with positive outcomes had higher odds of being fully reported (97). Comparisons of publications and protocols revealed that at least one primary outcome was changed, introduced, or omitted in 62 percent of the trials. In a survey of authors from 519 randomized trials, 32 percent of responders denied the existence of unreported outcomes, despite evidence to the contrary in their publications (105). Incompletely reported outcomes were approximately twice as likely to be statistically nonsignificant as fully reported outcomes. The combined evidence suggests that publication bias, time-lag bias, and outcome-reporting bias may lead to false-positive conclusions about treatment effects.

DISCUSSION

In theory, clinical interventions should only be used if they have been shown to be safe and effective in well-designed trials. The present review suggests that clinicians using evidence-based methods often have to base treatment decisions on randomized trials with small sample sizes and unclear control of bias. In such situations, a systematic review of the evidence is an important tool. Systematic reviews are often necessary to identify limitations in the existing evidence, such as bias or inadequate statistical power. Systematic reviews may also help identify interventions that are supported by convincing evidence. Still, it is the clinicians who must decide whether or not they believe the intervention should be used in clinical practice.

Research on sources of bias in randomized trials and systematic reviews is important to evidence-based clinical practice. Future methodological studies may need to consider the influence of blinding on trial validity. Additional evidence is necessary to determine the effects of blinded outcome assessment, sample size calculations, intention-to-treat analyses, and losses to follow-up on the control of bias in clinical trials. The actual influence of potential sources of bias may differ significantly from the theoretical influence. Large trial cohorts are often necessary to identify patterns of bias. The present review suggests that adequate randomization, blinding, and follow-up may be important in bias control, but the influence of the different components in individual trials cannot be predicted. Critical appraisal of bias control in individual trials is an essential tool for clinicians as well as researchers.

ACKNOWLEDGMENTS

This work was funded by the Danish Medical Research Council, the 1991 Pharmacy Foundation, the Copenhagen Hospital Corporation Medical Research Council, the Danish Institute of Health Technology Assessments, and the Copenhagen Trial Unit, Centre for Clinical Intervention Research.

The author thanks B. Als-Nielsen, B. Djulbegovic, S. Klingenberg, K. Krogsgaard, D. Nikolova, C. Gluud, P. Gotzsche, and Nader Salas for their invaluable help.

The author is an editor of the Cochrane Hepato-Biliary Group.

REFERENCES

- Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991;44:91–8.
- Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- 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.
- Jüni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
- Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42–6.
- Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982–9.
- Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973–82.
- Als-Nielsen B, Chen W, Gluud L, et al. Are trial size and reported methodological quality associated with treatment effects? Observational study of 523 randomised trials. Ottawa, Ontario, Canada: Canadian Cochrane Network and Centre, University of Ottawa, 2004. (<http://www.cochrane.org>).
- 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.
- Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. (Editorial). *BMJ* 1996;312:71–2.
- Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA* 2000;284:1290–6.
- Guyatt G, Sinclair J, Cook D, et al. Grading recommendations—a qualitative approach. In: Hayward R, elec. ed. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago, IL: JAMA Publishing Group, 2002. (<http://www.usersguides.org>). (Accessed November 27, 2002).
- MERIT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). *Lancet* 1999;353:2001–6.
- Stern L. In vivo assessment of the teratogenic potential of drugs in humans. *Obstet Gynecol* 1981;58(suppl):3S–8S.
- Nuesch R, Schroeder K, Dieterle T, et al. Relation between insufficient response to antihypertensive treatment and poor compliance with treatment: a prospective case-control study. *BMJ* 2001;323:142–6.

16. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215–18.
17. Owen CG, Whincup PH, Gilg JA, et al. Effect of breast feeding in infancy on blood pressure in later life: systematic review and meta-analysis. *BMJ* 2003;327:1189–95.
18. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233–40.
19. White E, Hunt JR, Casso D. Exposure measurement in cohort studies: the challenges of prospective data collection. *Epidemiol Rev* 1998;20:43–56.
20. Wacholder S. Design issues in case-control studies. *Stat Methods Med Res* 1995;4:293–309.
21. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1–173.
22. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. (Cochrane methodology review). In: *The Cochrane Library*, issue 4, 2002. Chichester, United Kingdom: John Wiley and Sons Ltd, 2002. (<http://www.cochrane.org/reviews/clibintro.htm>).
23. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and non-randomized studies. *JAMA* 2001;286:821–30.
24. Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995;345:772–6.
25. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA* 1996;276:1332–8.
26. LeLorier J, Gregoire G, Benhaddad A, et al. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536–42.
27. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984;3:409–22.
28. Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998;317:1170–1.
29. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA* 1998;279:1089–93.
30. Furukawa TA, Streiner DL, Hori S. Discrepancies among megatrials. *J Clin Epidemiol* 2000;53:1193–9.
31. McDonald S, Westby M, Clarke M, et al. Number and size of randomized trials reported in general health care journals from 1948 to 1997. *Int J Epidemiol* 2002;31:125–7.
32. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in *Hepatology*: predictors of quality. *Hepatology* 1999;30:1134–8.
33. Adetugbo K, Williams K. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000;136:381–5.
34. Kyriakidi M, Ioannidis JP. Design and quality considerations for randomized controlled trials in systemic sclerosis. *Arthritis Rheum* 2002;47:73–81.
35. Latronico N, Botteri M, Minelli C, et al. Quality of reporting of randomised controlled trials in the intensive care literature. A systematic analysis of papers published in *Intensive Care Medicine* over 26 years. *Int Care Med* 2002;28:1316–23.
36. Kjaergard LL, Frederiksen SL, Gluud C. Validity of randomized clinical trials in *Gastroenterology* from 1964 to 2000. *Gastroenterology* 2002;122:1157–60.
37. Huang W, LaBerge JM, Lu Y, et al. Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods. *J Vasc Interv Radiol* 2002;13:247–55.
38. Guyatt GH, Sinclair J, Cook DJ, et al. Users' Guides to the Medical Literature: XVI. How to use a treatment recommendation. Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. *JAMA* 1999;281:1836–43.
39. Alderson P, Green S, Higgins JPT, eds. *Cochrane reviewers' handbook 4.2.1* (updated December 2003). In: *The Cochrane Library*, issue 1, 2004. Chichester, United Kingdom: John Wiley and Sons Ltd, 2004. (<http://www.cochrane.org/reviews/clibintro.htm>).
40. Detsky AS, Naylor CD, O'Rourke K, et al. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255–65.
41. Pocock SJ. *Clinical trials—a practical approach*. Chichester, United Kingdom: John Wiley and Sons Ltd, 1996.
42. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209.
43. Altman DG, Bland JM. How to randomise. *BMJ* 1999;319:703–4.
44. Klein MC, Kaczorowski J, Robbins JM, et al. Physicians' beliefs and behaviour during a randomized controlled trial of episiotomy: consequences for women in their care. *CMAJ* 1995;153:769–79.
45. Swingler GH, Zwarenstein M. An effectiveness trial of a diagnostic test in a busy outpatients department in a developing country: issues around allocation concealment and envelope randomization. *J Clin Epidemiol* 2000;53:702–6.
46. Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995;274:1456–8.
47. Gluud C, Nikolova D. Quality assessment of reports on clinical trials in the *Journal of Hepatology*. *J Hepatol* 1998;29:321–7.
48. Schulz KF, Chalmers I, Altman DG, et al. The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. *Online J Curr Clin Trials* 1995(doc. no. 197).
49. Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. *J Orthod* 2003;30:309–15.
50. Kjaergard LL, Gluud C. Funding, disease area, and internal validity of hepatobiliary randomized clinical trials. *Am J Gastroenterol* 2002;97:2708–13.
51. Devereaux PJ, Choi PT, 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* 2005;57:1232–6.
52. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359:696–700.
53. Soares HP, Daniels S, Kumar A, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22–4.
54. Pildal J, Chan AW, Hróbjartsson A, et al. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. *BMJ* 2005;330:1049–52.
55. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet* 1991;337:867–72.
56. Als-Nielsen B, Gluud L, Gluud C. Methodological quality and treatment effects in randomised trials—a review of six empirical studies. Ottawa, Ontario, Canada: Canadian

- Cochrane Network and Centre, University of Ottawa, 2004. (<http://www.cochrane.org>).
57. Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000–3.
 58. Campbell IA, Lyons E, Prescott RJ. Stopping smoking. Do nicotine chewing-gum and postal encouragement add to doctors' advice? *Practitioner* 1987;231:114–17.
 59. Karlowski TR, Chalmers TC, Frenkel LD, et al. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. *JAMA* 1975;231:1038–42.
 60. Chalmers TC. Effects of ascorbic acid on the common cold. An evaluation of the evidence. *Am J Med* 1975;58:532–6.
 61. Montori VM, Bhandari M, Devereaux PJ, et al. In the dark: the reporting of blinding status in randomized controlled trials. *J Clin Epidemiol* 2002;55:787–90.
 62. Schulz KF, Grimes DA, Altman DG, et al. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;312:742–4.
 63. Corrigan JD, Harrison-Felix C, Bogner J, et al. Systematic bias in traumatic brain injury outcome studies because of loss to follow-up. *Arch Phys Med Rehabil* 2003;84:153–60.
 64. DiFranceisco W, Kelly JA, Sikkema KJ, et al. Differences between completers and early dropouts from 2 HIV intervention trials: a health belief approach to understanding prevention program attrition. *Am J Public Health* 1998;88:1068–73.
 65. Kemeny MM, Adak S, Gray B, et al. Combined-modality treatment for resectable metastatic colorectal carcinoma to the liver: surgical resection of hepatic metastases in combination with continuous infusion of chemotherapy—an intergroup study. *J Clin Oncol* 2002;20:1499–505.
 66. Egger M, Jüni P, Bartlett C, et al. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996–9.
 67. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002;359:781–5.
 68. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670–4.
 69. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001;165:1339–41.
 70. Kjaergard LL, Krogsgaard K, Gluud C. Interferon alfa with or without ribavirin for chronic hepatitis C: systematic review of randomised trials. *BMJ* 2001;323:1151–5.
 71. Smith R. Beyond conflict of interest. Transparency is the key. (Editorial). *BMJ* 1998;317:291–2.
 72. Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the *BMJ*. *BMJ* 2002;325:249.
 73. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454–65.
 74. Lexchin J, Bero LA, Djulbegovic B, et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167–70.
 75. Als-Nielsen B, Chen W, Gluud C, et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921–8.
 76. Djulbegovic B, Lacevic M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635–8.
 77. Joffe S, Harrington DP, George SL, et al. Satisfaction of the uncertainty principle in cancer clinical trials: retrospective cohort analysis. *BMJ* 2004;328:1463.
 78. Krimsky S, Rothenberg LS. Conflict of interest policies in science and medical journals: editorial practices and author disclosures. *Sci Eng Ethics* 2001;7:205–18.
 79. Weller IVD, Ashby D, Brook R, et al. Interim report of the Committee on Safety of Medicines' Expert Working Group on Selective Serotonin Reuptake Inhibitors. London, United Kingdom: Medicines and Healthcare Products Regulatory Agency, United Kingdom Department of Health, 2003. (Available from the Seroxat Users' Group at www.seroxatusergroup.org.uk). (Accessed January 5, 2004).
 80. Khan A, Khan S, Kolts R, et al. Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. *Am J Psychiatry* 2003;160:790–2.
 81. Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. *Arch Gen Psychiatry* 2000;57:311–17.
 82. Abbott A. British panel bans use of antidepressant to treat children. *Nature* 2003;423:792.
 83. Mitka M. FDA alert on antidepressants for youth. *JAMA* 2003;290:2534.
 84. Hirschfeld RMA. Suicide and antidepressant treatment. *Arch Gen Psychiatry* 2000;57:325–6.
 85. Oranski I. FDA questions antidepressant safety for children. *Lancet* 2003;362:1558.
 86. Bombadier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. *N Engl J Med* 2003;343:1520–8.
 87. Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA* 2001;286:954–9.
 88. Ray WA, Stein CM, Daugherty JR, et al. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet* 2002;360:1071–3.
 89. McCormack J, Rangno R. Digging for data from the COX-2 trials. *CMAJ* 2002;166:1649–50.
 90. Food and Drug Administration, US Department of Health and Human Services. FDA Public Health Advisory—October 27, 2003. Subject: reports of suicidality in pediatric patients being treated with antidepressant medications for major depressive disorder (MDD). Washington, DC: Food and Drug Administration, 2003. (www.fda.gov/cder/drug/advisory/mdd.htm). (Accessed December 12, 2003).
 91. Center for Drug Evaluation and Research, Food and Drug Administration, US Department of Health and Human Services. Vioxx (rofecoxib) information. Washington, DC: Food and Drug Administration, 2005. (<http://www.fda.gov/cder/drug/infopage/COX2/default.htm>). (Date created: April 6, 2005).
 92. Williams HC. Evening primrose oil for atopic dermatitis. (Editorial). *BMJ* 2003;327:1358–9.
 93. Dickersin K, Chan S, Chalmers TC, et al. Publication bias and clinical trials. *Control Clin Trials* 1987;8:343–53.
 94. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374–8.

95. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993(doc. no. 50).
96. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640–5.
97. Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–65.
98. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281–6.
99. Olson CM, Rennie D, Cook D, et al. Publication bias in editorial decision making. *JAMA* 2002;287:2825–8.
100. Song F, Eastwood S, Gilbody S, et al. Publication and related biases. *Health Technol Assess* 2000;4:1–115.
101. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. *J Clin Epidemiol* 2002;55:407–10.
102. Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987;295:645–6.
103. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992;305:15–9.
104. Hutchison BG, Lloyd S. Comprehensiveness and bias in reporting clinical trials. Study of reviews of pneumococcal vaccine effectiveness. *Can Fam Physician* 1995;41:1356–60.
105. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753–8.