

## Reported Methodologic Quality and Discrepancies between Large and Small Randomized Trials in Meta-Analyses

Lise L. Kjaergard, MD; John Villumsen, MSc; and Christian Gluud, MD, DrMSc

**Purpose:** To explore whether reported methodologic quality affects estimated intervention effects in randomized trials and contributes to discrepancies between the results of large randomized trials and small randomized trials in meta-analyses.

**Data Sources:** Meta-analyses of randomized trials that included at least one large trial ( $\geq 1000$  participants) were included, regardless of the therapeutic area. Eligible meta-analyses were identified through electronic searches and bibliographies of relevant articles.

**Study Selection:** Full-length randomized trials.

**Data Extraction:** Methodologic quality was assessed according to reported randomization, double blinding, and follow-up as separate components and by using the Jadad composite scale.

**Data Synthesis:** Fourteen meta-analyses involving 190 randomized trials from eight therapeutic areas were included. Compared with large trials, intervention effects were exaggerated in small trials with inadequate allocation sequence generation (ratio of

odds ratios, 0.46 [95% CI, 0.25 to 0.83];  $P = 0.011$ ), inadequate allocation concealment (ratio of odds ratios, 0.49 [CI, 0.27 to 0.86];  $P = 0.014$ ), and no double blinding (ratio of odds ratios, 0.52 [CI, 0.28 to 0.96];  $P = 0.01$ ). Large trials did not differ significantly from small trials with adequate generation of the allocation sequence, adequate allocation concealment, or adequate double blinding. No association was seen between reported follow-up and intervention effects. The Jadad scale provided no additional information because the scale and the quality components overlapped substantially.

**Conclusions:** Inadequate generation of the allocation sequence, allocation concealment, and double blinding lead to exaggerated estimates of intervention benefit and may contribute to discrepancies between the results of large randomized trials and small randomized trials in meta-analyses.

*Ann Intern Med.* 2001;135:982-989.

[www.annals.org](http://www.annals.org)

For author affiliations and current addresses, see end of text.

Discrepancies may occur between the results of large randomized trials and the pooled results of several small trials in meta-analyses (1–4). Previous studies have suggested that discrepancies may be due to publication bias, that is, the fact that small trials are more likely to be published if they show a statistically significant intervention effect (5–8).

Previous empirical studies of the association between methodologic quality and intervention effects have had inconsistent conclusions (9–12). In theory, adequate randomization requires adequate generation of the allocation sequence and adequate allocation concealment. The assumption is partly supported by studies from Schulz and colleagues (10) and Moher and associates (11, 12), who found that trials with inadequate allocation concealment exaggerate intervention effects significantly compared with trials reporting adequate allocation concealment. However, Emerson and coworkers (9) found no association between reported allocation concealment and intervention effects. Furthermore, none of the studies (9–12) found a significant association between generation of allocation sequence and intervention effects, although Schulz and colleagues found a nonsignificant trend (10).

Schulz and colleagues (10), who analyzed trials in pregnancy and childbirth, found that trials without double blinding exaggerate intervention effects significantly compared with double-blind trials. However, Emerson and coworkers (9) and Moher and associates (11), who included trials from several therapeutic areas, found no association between double blinding and intervention effects.

Methodologic quality can be assessed by using separate components, as in the studies discussed above, or by using one of several quality scales (13). One popular scale, developed by Jadad and colleagues in 1996 (14), has thus far received 233 registered citations (Institute for Scientific Information, Philadelphia, Pennsylvania). The scale includes assessment of the reported generation of allocation sequence, double blinding, and follow-up. Moher and associates (11, 12) found that trials with a low score on this scale exaggerate intervention effects significantly compared with trials that have high quality scores. However, the use of this and other quality scales has been disputed by Juni and coworkers (15), who showed that several quality scales produce inconsistent conclusions.

We studied the potential association between re-

ported methodologic quality and intervention effects to assess whether methodologic quality may explain discrepancies between the results of large and small randomized trials in meta-analyses.

## METHODS

### Identification and Selection of Meta-Analyses and Trials

According to suggestions in other studies (2–4), we arbitrarily defined trials with 1000 or more participants as “large.” We searched the Cochrane Library, MEDLINE on PubMed (using *meta-analysis*, *review*, *randomi\*ed*, and *controlled clinical trial* as free text search words), and reference lists of relevant articles to identify potentially eligible meta-analyses that included at least one large trial.

We identified 23 eligible meta-analyses. Nine were excluded because they included trials that were also included in larger eligible meta-analyses ( $n = 5$ ), lacked references to the primary trials ( $n = 3$ ), or excluded low-quality trials ( $n = 1$ ). Accordingly, 14 meta-analyses (16–26) were included. Three of the included Cochrane Reviews included two meta-analyses each (22, 25, 26). The meta-analyses included 248 trials, of which 58 were excluded because they were unpublished ( $n = 33$ ), were quasi-randomized ( $n = 15$ ), or were published as abstracts ( $n = 8$ ). We were unable to translate 2 Spanish-language articles. The remaining 190 trials, published as English-language ( $n = 188$ ) or German-language ( $n = 2$ ) full-length articles, were included.

Our analyses included 23 large and 167 small randomized trials and a total of 136 164 participants. On the basis of the study by Schulz and colleagues (10), who analyzed 250 controlled trials with 62 091 participants, we estimated that our sample size would be large enough to show significant differences between intervention benefits in high-quality and low-quality trials.

### Assessment of Methodologic Quality

Methodologic quality was defined as the confidence that the trial’s design, conduct, analysis, and presentation minimized or avoided biases in the trial’s intervention comparisons (12). The reported methodologic quality was assessed in an unmasked manner by using four separate components and a composite quality scale.

The four components were generation of the allocation sequence (adequate [computer-generated random numbers or similar] or inadequate [not described]),

location concealment (adequate [central independent unit, sealed envelopes, or similar] or inadequate [not described or open table of random numbers or similar]), double blinding (adequate [identical placebo tablets or similar] or inadequate [not performed or tablets versus injections or similar]), and follow-up (adequate [number and reasons for dropouts and withdrawals described] or inadequate [number or reasons for dropouts and withdrawals not described]).

The five-point quality scale included generation of the allocation sequence (2 points, computer-generated random numbers or similar; 1 point, not described; 0 points, quasi-randomized trial [which we excluded]), double blinding (2 points, identical placebo tablets or similar; 1 point, not described; 0 points, no blinding or inadequate method, such as tablets versus injections or similar), and follow-up (1 point, number and reasons for dropouts and withdrawals described; 0 points, number or reasons for dropouts and withdrawals not described). The quality score was ranked as low ( $\leq 2$  points) or high ( $\geq 3$  points), as suggested elsewhere (11).

Two reviewers assessed the effect of masking and the interobserver reliability of the quality assessments. The reported methodologic quality of 30 trials was assessed with and without masking the names of the authors and journal, the year of publication, acknowledgments, institutional affiliations, and funding. The unmasked assessments were performed first. The masked assessments were performed 3 months later, ensuring that the assessors did not remember the trials. The difference between masked and unmasked quality assessments was not significant (mean [ $\pm$ SE] quality score,  $3.70 \pm 0.19$  vs.  $3.63 \pm 0.21$ ;  $P > 0.2$ ). Interobserver reliability was assessed by using 30 randomized trials randomly selected from the Cochrane Hepato-Biliary Group Controlled Trials Register and was found to be high (intraclass correlation coefficient, 0.96 [95% CI, 0.92 to 0.98]) (27). After assessing the quality of 100 trials, we reassessed the methodologic quality of the trials from the Cochrane Hepato-Biliary Group Controlled Trials Register and found that the “test–retest” reliability was high (0.98 [CI, 0.97 to 0.99]) (27).

### Data Extraction

Data were extracted independently by two reviewers. First, the primary binary outcome measure described by the largest number of trials in each meta-

**Table 1. Included Meta-Analyses and Primary Outcome Measures**

| Author, Year (Reference)         | Outcome Measure        |
|----------------------------------|------------------------|
| Honan et al., 1990 (16)          | Mortality              |
| Garg and Yusuf, 1995 (17)        | Mortality              |
| Heidenreich et al., 1997 (18)    | Mortality              |
| Palmer et al., 1997 (19)         | Deep venous thrombosis |
| Tan and Hannah, 1998 (20)        | Cesarean section       |
| Kennedy et al., 1998 (21)        | Dropouts               |
| Martin-Hirsch et al., 1998 (22)  | Endocervical cells     |
| Mulrow et al., 1998 (23)         | Mortality              |
| Neilson and Alfirevic, 1998 (24) | Neonatal mortality     |
| Silagy et al., 1998 (25)         | Resumed smoking        |
| Thacker and Stroup, 1998 (26)    | Cesarean section       |

analysis was identified. We then extracted the number of events in the intervention and control groups and the number of participants randomly assigned to the intervention and control groups. All disagreements were due to inaccurate data extraction and were resolved through further consulting of the original articles and meta-analyses. Consensus was achieved before analyses were done in all cases.

### Statistical Analysis

Analyses were performed by using SAS for Windows, version 6.12 (SAS Institute, Inc., Cary, North Carolina) or SPSS for Windows, version 10.0 (SPSS, Inc., Chicago, Illinois). Differences between the masked and unmasked quality assessments were estimated by using the Kruskal–Wallis test. The number of participants and the year of publication in trials with adequate versus inadequate generation of allocation sequence, allocation concealment, double blinding, and follow-up and high versus low quality scores were compared by using the Pearson chi-square test.

To estimate evidence of publication bias and other biases, we used linear regression to analyze funnel-plot asymmetry (7). The standard normal deviate, defined as the log odds ratio divided by its standard error, was regressed against the precision (the inverse of the standard error). If funnel-plot asymmetry is present, the regression line will not run through the origin, and the intercept will provide a measure of asymmetry.

Intervention effects were estimated by using the number of events and participants in the treatment group and the number of events and participants in the control group (Appendix). Accordingly, two observa-

tions were needed per trial, one for the intervention group and one for the control group. When necessary, positive outcomes were re-expressed as unwanted end points, for example, mortality instead of survival. Discrepancies between intervention effects in large and small trials were estimated by the ratio of odds ratios (10), which is the summary odds ratio of large trials divided by the summary odds ratio of small trials. In this modeling convention, a ratio of odds ratios less than 1.0 indicates that a group of trials (for example, small trials with inadequate allocation concealment) exaggerates the intervention effect compared with the referent group (for example, large trials). Variance and confidence intervals were increased to adjust for overdispersion (Appendix). The Pearson chi-square test was used to estimate the potential overlap between quality components.

## RESULTS

### Characteristics of Included Trials

We were able to locate and retrieve all 190 included trials and to reproduce the results reported in the meta-analyses (16–26) (Table 1). Of the 190 trials, 81 (43%) reported adequate generation of the allocation sequence, 68 (36%) reported adequate allocation concealment, 103 (54%) reported adequate double blinding, and 133 (70%) reported adequate follow-up. Three large trials (11%) and 58 small trials (36%) had low quality scores.

The median number of participants was 1741 in the large trials (range, 1004 to 17 187 participants) and 165 in the small trials (range, 19 to 982 participants). The median number of participants in small trials did not significantly differ in trials with adequate and those with inadequate generation of allocation sequence ( $P > 0.2$ ), allocation concealment ( $P > 0.2$ ), or double blinding ( $P > 0.2$ ) or low versus high quality scores ( $P > 0.2$ ). The included trials were published from 1960 to 1998 (interquartile range, 1986 to 1992). Only 5% were published after the Consolidated Standards of Reporting Trials statement (28). Overall, the year of publication did not significantly differ in trials with adequate versus inadequate generation of the allocation sequence ( $P > 0.2$ ), allocation concealment ( $P > 0.2$ ) or double blinding ( $P > 0.2$ ), or low versus high quality scores ( $P > 0.2$ ). In two meta-analyses (25, 26), trials with low quality scores were published before trials with high quality scores (median year, 1977 versus 1986 [ $P = 0.027$ ] and

**Table 2. Discrepancies between Large Trials and Small Trials, by Methodologic Quality**

| Comparison  | Ratio of Odds Ratios (95% CI) | P Value |
|---|-------------------------------|---------|
| Large trials vs. small trials with inadequate generation of the allocation sequence | 0.46 (0.25–0.83)              | 0.011   |
| Large trials vs. small trials with adequate generation of the allocation sequence   | 0.90 (0.47–1.76)              | >0.2    |
| Large trials vs. small trials with inadequate allocation concealment                | 0.49 (0.27–0.86)              | 0.014   |
| Large trials vs. small trials with adequate allocation concealment                  | 1.01 (0.48–2.11)              | >0.2    |
| Large trials vs. small trials with inadequate or no double blinding                 | 0.52 (0.28–0.96)              | 0.01    |
| Large trials vs. small trials with adequate double blinding                         | 0.84 (0.43–1.66)              | >0.2    |
| Large trials vs. small trials with inadequate follow-up                             | 0.72 (0.30–1.71)              | >0.2    |
| Large trials vs. small trials with adequate follow-up                               | 0.58 (0.32–1.02)              | 0.06    |

1985 versus 1993 [ $P = 0.028$ ], respectively). Conversely, in two other meta-analyses (21, 25), trials with low quality scores were published later than trials with high quality scores (median year, 1989 versus 1986 [ $P = 0.0025$ ] and 1996 versus 1993 [ $P = 0.022$ ], respectively). In the remaining meta-analyses, the publication year of trials with low versus high quality scores did not significantly differ. Statistically significant funnel-plot asymmetry suggesting publication bias was present in two meta-analyses (25, 26). The remaining meta-analyses did not show significant funnel-plot asymmetry.

#### Quality Components and Intervention Effects

The ratio of odds ratios between large trials and small trials with inadequate generation of allocation sequence was 0.46 (CI, 0.25 to 0.83) (Table 2). Compared with large trials, odds ratios were exaggerated by 54% in small trials with inadequate generation of the allocation sequence ( $P = 0.011$ ), by 51% in small trials with inadequate allocation concealment ( $P = 0.014$ ), and by 48% in small trials without double blinding ( $P = 0.01$ ). Odds ratios generated by small trials with adequate generation of the allocation sequence, adequate allocation concealment, or adequate double blinding did not significantly differ from those generated by large trials. The analyses showed no significant association between reported follow-up and estimated intervention effects.

The odds ratios generated by all trials (large and small) with inadequate generation of the allocation sequence were on average significantly exaggerated by 51% compared with all trials reporting adequate generation of allocation sequence (Table 3). All trials with inadequate allocation concealment exaggerated intervention benefits by 40% compared with all trials reporting adequate allocation concealment ( $P = 0.12$ ). The odds ratios generated by all trials without double blinding were significantly exaggerated by 44% compared with all double-blind trials ( $P = 0.041$ ). The analyses showed no significant association between reported follow-up and estimated intervention effects.

The odds ratios were significantly exaggerated by 48% in small trials with inadequate versus adequate generation of the allocation sequence (Table 4). Odds ratios were significantly exaggerated by 52% in small trials with inadequate versus adequate allocation concealment ( $P = 0.027$ ). The odds ratios generated by small trials without versus with double blinding did not differ significantly. The analyses showed no significant association between reported follow-up and estimated intervention effects in small trials.

Some components overlapped significantly. Of 68 trials with adequate allocation concealment, 64 (94%) also reported adequate generation of allocation sequence ( $P = 0.001$ ) and 33 (49%) were double blind ( $P = 0.022$ ). Of the 103 double-blind trials, 86 (83%) reported

**Table 3. Discrepancies between Large and Small Trials with Adequate versus Inadequate Methodologic Quality**

| Comparison  | Ratio of Odds Ratios (95% CI) | P Value |
|---|-------------------------------|---------|
| Large and small trials with adequate vs. inadequate generation of the allocation sequence | 0.49 (0.30–0.81)              | <0.001  |
| Large and small trials with adequate vs. inadequate allocation concealment                | 0.60 (0.31–1.15)              | 0.12    |
| Large and small trials with adequate vs. inadequate (no) double blinding                  | 0.56 (0.33–0.98)              | 0.041   |
| Large and small trials with adequate vs. inadequate follow-up                             | 1.50 (0.80–2.78)              | 0.2     |

**Table 4. Discrepancies between Small Trials, by Methodologic Quality**

| Comparison  | Ratio of Odds Ratios (95% CI) | P Value |
|---|-------------------------------|---------|
| Small trials with adequate vs. inadequate generation of the allocation sequence | 0.52 (0.28–0.93)              | 0.029   |
| Small trials with adequate vs. inadequate allocation concealment                | 0.48 (0.25–0.92)              | 0.027   |
| Small trials with adequate vs. inadequate or no double blinding                 | 0.62 (0.33–1.15)              | 0.13    |
| Small trials with adequate vs. inadequate follow-up                             | 1.24 (0.56–2.74)              | >0.2    |

follow-up adequately ( $P = 0.026$ ). The remaining quality components did not overlap significantly.

We performed an analysis limited to trials with inadequate allocation concealment to assess whether allocation concealment explained the association between generation of the allocation sequence and intervention effects. In this analysis, trials with inadequate generation of the allocation sequence exaggerated intervention benefits significantly compared to trials with adequate generation of the allocation sequence (ratio of the odds ratios, 0.51 [CI, 0.31 to 0.87];  $P = 0.013$ ). Thus, the reported allocation concealment did not explain the association between generation of the allocation sequence and intervention effects. When we limited our analyses to trials with inadequate allocation concealment, we found that intervention benefits were significantly exaggerated in trials without double blinding compared with double-blind trials (ratio of the odds ratios, 0.52 [CI, 0.29 to 0.92];  $P = 0.024$ ).

#### Quality Score and Intervention Effects

Compared with large trials, odds ratios were significantly exaggerated in small trials with low quality scores (ratio of the odds ratios, 0.52 [CI, 0.28 to 0.96];  $P = 0.037$ ) but not in small trials with high quality scores (ratio of the odds ratios, 0.84 [CI, 0.43 to 1.64];  $P = 0.61$ ). Odds ratios were significantly exaggerated in all trials with low versus high quality scores (ratio of the odds ratios, 0.56 [CI, 0.33 to 0.98];  $P = 0.042$ ), but not in small trials with low versus high quality scores (ratio of the odds ratios, 0.62 [CI, 0.33 to 1.14];  $P = 0.12$ ).

We performed an analysis limited to all trials with inadequate allocation concealment to assess whether allocation concealment explained the association between quality score and intervention effects. In this analysis, trials with low quality scores exaggerated intervention effects significantly compared with trials with high qual-

ity scores (ratio of the odds ratios, 0.52 [CI, 0.29 to 0.92];  $P = 0.025$ ). Thus, the association between the quality score and intervention effects was not explained by the reported allocation concealment. However, the reported double blinding and the quality score overlapped considerably: Only 15 of 129 trials (12%) with high quality scores lacked double blinding, and only 2 of 58 double-blind trials (3%) had low quality scores. When we separately analyzed double-blind trials and trials without double blinding, we found no significant association between the quality score and the estimated intervention effect. Thus, the association between the quality score and intervention effects was largely explained by the reported double blinding.

#### DISCUSSION

Small trials with inadequate generation of allocation sequence, inadequate allocation concealment, or inadequate or no double blinding may significantly exaggerate intervention effects. Intervention effects in small trials with adequate generation of allocation sequence, adequate allocation concealment, or adequate double blinding did not significantly differ from those in large trials. Therefore, our findings suggest that trials with inadequate randomization and double blinding may contribute to discrepancies between the results of large randomized trials and meta-analyses.

Our study clarifies the evidence concerning the association between reported methodologic quality and intervention effects. First, adequate generation of the allocation sequence and adequate allocation concealment should be required for adequate randomization. Unlike previous investigators (9–12), we found that trials with inadequate generation of allocation sequence exaggerate intervention effects significantly. In accordance with previous evidence (10–12), we found that trials with inadequate allocation concealment also gen-

erate exaggerated results. Despite the considerable overlap between generation of allocation sequence and allocation concealment, our results suggest that both factors may independently affect the estimated intervention effect. Second, Schulz and colleagues (10), who found a significant association between intervention effects and double blinding, included only trials on pregnancy and childbirth. Our study supports their findings and extends the evidence by including trials from several therapeutic areas. We also performed sensitivity analyses limited to interventions, which could be blinded with reasonable effort; these analyses did not change our conclusions (data not shown). Finally, previous studies have questioned the reliability of reported losses to follow-up (10, 29). In accordance with Schulz and colleagues' results (10), we found no association between intervention effects and reported follow-up.

Previous studies of the association between methodologic quality and intervention effects did not adjust for sample size or assess the potential influence of publication bias (10–12), even though these factors may be important. Small trials are more susceptible to publication bias than are larger trials (5–8). Furthermore, sample size has been associated with methodologic quality (30). This may lead to speculation that the association between methodologic quality and intervention effects simply reflects publication bias. In our study, the average sample size of small trials with adequate and inadequate methodologic quality did not significantly differ, indicating an equivalent risk for publication bias. Furthermore, only two of the included meta-analyses showed evidence of publication bias.

In agreement with the findings of Moher and associates (11, 12) and Jüni and coworkers (15), we found that trials with a low quality score on the scale developed by Jadad and colleagues (14) significantly exaggerate intervention benefits. Jüni and coworkers (15), who assessed 25 quality scales, concluded that the outcomes of different scales vary considerably (15). These findings are not surprising, considering the lack of agreement on the definition of quality among the scales (13, 15), but they do not necessarily mean that all scales are invalid. Reliable estimates of abstract constructs, such as quality of life, can be obtained by using psychometrically developed and validated scales (27), and quality scales have potential benefits.

We found a significant association between a psy-

chometrically developed quality scale (14) and intervention effects. However, we also found that the scale has several shortcomings. First, allocation concealment was not included because of the low frequency of endorsement (14). In accordance with previous studies (10–12), we found a substantial risk for bias in trials with inadequate allocation concealment. Allocation concealment should therefore be considered when methodologic quality is assessed. Second, the scale consists of only three items. Double blinding carries large weight in the scale, and overlap between this component and the scale was considerable. Furthermore, some interventions are difficult or impossible to blind. Finally, the scale includes the reported follow-up that was not associated with intervention effects and perhaps primarily concerns the quality of the reporting. Accordingly, assessment of methodologic quality should focus on generation of allocation sequence, allocation concealment, and double blinding.

Our study has limitations. First, the quality assessments were unmasked. However, in agreement with previous findings (31), our unmasked and masked assessments did not differ significantly, and masking would not have greatly influenced our results. Second, methodologic quality was assessed by using the published reports, which may not reflect the actual design or execution of the trials. However, the association between the reported methodologic quality and intervention effects is very important because clinicians and researchers usually must use published reports. Few authors respond to requests for further information (32), and authors of low-quality trials may be less likely to reveal flaws.

Small trials are often conducted before larger trials. Control treatments may have improved or changed in a manner that could reduce the efficacy of the experimental treatment. This may be an alternative explanation for discrepancies between the results of small randomized trials with inadequate methodologic quality and large trials. However, in our study, most small trials with adequate and inadequate methodologic quality were published at the same time. Furthermore, 50% of the included trials were published from 1986 to 1992, and 80% were published from 1981 to 1992. This time frame makes confounding by date less likely.

Another possible explanation for our findings is that interventions may be implemented less thoroughly in larger pragmatic trials. However, results of large trials

and small trials with adequate methodologic quality were in agreement.

In conclusion, the reported generation of allocation sequence, allocation concealment, and double blinding should be appraised before making causal inferences about the effect of medical interventions. Furthermore, methodologic quality should be incorporated into meta-analyses to guard against bias. If the methodologic quality of the evidence is not considered, new interventions may be implemented on the basis of misleading evidence.

## APPENDIX

Intervention effects (log odds ratio) in large trials and groups of trials with adequate or inadequate methodologic quality were estimated by using unconditional logistic regression. The regression models included indicator variables for intervention group, trial, meta-analysis, and interactions between intervention group and meta-analysis and trial nested within meta-analysis. The effect of methodologic quality was estimated by calculating the ratio of odds ratios of intervention effects in 1) large trials versus small trials with inadequate methodologic quality, 2) large trials versus small trials with adequate methodologic quality, 3) large and small trials with adequate versus inadequate methodologic quality, and 4) small trials with adequate versus inadequate methodologic quality.

The likelihood deviance was estimated to obtain a measure of the residual variation in each model. The deviance with reference to degrees of freedom was used as a measure of overdispersion, which may lead to underestimation of the true standard error if not considered. Accordingly, an overdispersion parameter was used to correct the covariance matrix by choosing the “dscale” option in the “Genmod” procedure in SAS to estimate more appropriate confidence intervals of the ratio of odds ratios.

“Methodologic quality” consisted of generation of the allocation sequence, allocation concealment, double blinding, follow-up, or quality score.

From Cochrane Hepato-Biliary Group, the Copenhagen Trial Unit, Center for Clinical Intervention Research, Copenhagen University Hospital, H:S Rigshospitalet, Copenhagen, Denmark.

**Acknowledgments:** The authors thank Doug Altman and Peter Gøtzsche for valuable comments, Nader Salassahri for computer assistance, and Dimitrinka Nikolova for secretarial assistance.

**Grant Support:** By grants from the Danish Medical Research Council and the 1991 Pharmacy Foundation in Denmark.

**Requests for Single Reprints:** Lise L. Kjaergard, MD, Cochrane Hepato-Biliary Group, the Copenhagen Trial Unit, Center for Clinical Intervention Research, Copenhagen University Hospital, H:S Rigshospitalet, DK-2100 Copenhagen, Denmark; e-mail, kjaergard@ctu.rh.dk.

**Current Author Addresses:** Drs. Kjaergard and Gluud and Mr. Villumsen: Cochrane Hepato-Biliary Group, the Copenhagen Trial Unit, Center for Clinical Intervention Research, Copenhagen University Hospital, H:S Rigshospitalet, DK-2100 Copenhagen, Denmark.

## References

- Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. *JAMA*. 1998;279:1089-93. [PMID: 9546568]
- Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA*. 1996;276:1332-8. [PMID: 8861993]
- Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet*. 1995;345:772-6. [PMID: 7891492]
- LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337:536-42. [PMID: 9262498]
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-72. [PMID: 1672966]
- Egger M, Smith GD. Bias in location and selection of studies. *BMJ*. 1998;316:61-6. [PMID: 9451274]
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629-34. [PMID: 9310563]
- Gluud C. “Negative trials” are positive! [Editorial] *J Hepatol*. 1998;28:731-3. [PMID: 9566846]
- Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials*. 1990;11:339-52. [PMID: 1963128]
- 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. [PMID: 7823387]
- 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;352:609-13. [PMID: 9746022]
- Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess*. 1999;3:i-iv, 1-98. [PMID: 10374081]
- 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. [PMID: 7743790]
- 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. [PMID: 8721797]
- Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054-60. [PMID: 10493204]
- Honan MB, Harrell FE, Reimer KA, Califf RM, Mark DB, Pryor DB, et al. Cardiac rupture, mortality and the timing of thrombolytic therapy: a meta-analysis. *J Am Coll Cardiol*. 1990;16:359-67. [PMID: 2142705]
- Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group on ACE Inhibitor Trials. *JAMA*. 1995;273:1450-6.

[PMID: 7654275]

18. **Heidenreich PA, Lee TT, Massie BM.** Effect of beta-blockade on mortality in patients with heart failure: a meta-analysis of randomized clinical trials. *J Am Coll Cardiol.* 1997;30:27-34. [PMID: 9207617]
19. **Palmer AJ, Schramm W, Kirchhof B, Bergemann R.** Low molecular weight heparin and unfractionated heparin for prevention of thrombo-embolism in general surgery: a meta-analysis of randomised clinical trials. *Haemostasis.* 1997;27:65-74. [PMID: 9212354]
20. **Tan BP, Hannah ME.** Prostaglandins for prelabour rupture of membranes at or near term. *Cochrane Database Syst Rev.* 2000;CD000178. [PMID: 10796171]
21. **Kennedy E, Song F, Hunter R, Clarke A, Gilbody S.** Risperidone versus typical antipsychotic medication for schizophrenia. *Cochrane Database Syst Rev.* 2000;CD000440. [PMID: 10796543]
22. **Martin-Hirsch P, Jarvis G, Kitchener H, Lilford R.** Collection devices for obtaining cervical cytology samples (Cochrane review). *Cochrane Database Syst Rev.* 2000;CD001036. [PMID: 10908482]
23. **Mulrow C, Lau J, Cornell J, Brand M.** Pharmacotherapy for hypertension in the elderly. *Cochrane Database Syst Rev.* 2000;CD000028. [PMID: 10796688]
24. **Neilson JP, Alfirevic Z.** Doppler ultrasound for fetal assessment in high risk pregnancies. *Cochrane Database Syst Rev.* 2000;CD000073. [PMID: 10796113]
25. **Silagy C, Mant D, Fowler G, Lancaster T.** Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev.* 2000;CD000146. [PMID: 10908462]
26. **Thacker SB, Stroup DF.** Continuous electronic heart rate monitoring for fetal assessment during labor. *Cochrane Database Syst Rev.* 2000;CD000063. [PMID: 10796109]
27. **Streiner DL, Norman GR.** *Health Measurement Scales: A Practical Guide to Their Development and Use.* New York: Oxford Univ Pr; 1995.
28. **Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al.** Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276:637-9. [PMID: 8773637]
29. **Gotzsche PC.** Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials.* 1989;10:31-56. [PMID: 2702836]
30. **Kjaergard LL, Nikolova D, Gluud C.** Randomized clinical trials in HEPATOLOGY: predictors of quality. *Hepatology.* 1999;30:1134-8. [PMID: 10534332]
31. **Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Knipschild PG.** Balneotherapy and quality assessment: interobserver reliability of the Maasricht criteria list and the need for blinded quality assessment. *J Clin Epidemiol.* 1998;51:335-41. [PMID: 9539890]
32. **Liberati A, Himel HN, Chalmers TC.** A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol.* 1986;4:942-51. [PMID: 3711962]