Review

Empirical Evidence of Bias

Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials

Kenneth F. Schulz, PhD, MBA; Iain Chalmers, MBBS, MSc; Richard J. Hayes, MSc; Douglas G. Altman

Objective.—To determine if inadequate approaches to randomized controlled trial design and execution are associated with evidence of bias in estimating treatment effects.

Design.—An observational study in which we assessed the methodological quality of 250 controlled trials from 33 meta-analyses and then analyzed, using multiple logistic regression models, the associations between those assessments and estimated treatment effects.

Data Sources.—Meta-analyses from the Cochrane Pregnancy and Childbirth Database.

Main Outcome Measures.—The associations between estimates of treatment effects and inadequate allocation concealment, exclusions after randomization, and lack of double-blinding.

Results.—Compared with trials in which authors reported adequately concealed treatment allocation, trials in which concealment was either inadequate or unclear (did not report or incompletely reported a concealment approach) yielded larger estimates of treatment effects ($P<.001$). Odds ratios were exaggerated by 41% for inadequately concealed trials and by 30% for unclearly concealed trials (adjusted for other aspects of quality). Trials in which participants had been excluded after randomization did not yield larger estimates of effects, but that lack of association may be due to incomplete reporting. Trials that were not double-blind also yielded larger estimates of effects ($P=.01$), with odds ratios being exaggerated by 17%.

Conclusions.—This study provides empirical evidence that inadequate methodological approaches in controlled trials, particularly those representing poor allocation concealment, are associated with bias. Readers of trial reports should be wary of these pitfalls, and investigators must improve their design, execution, and reporting of trials.

(*JAMA*. 1995;273:408-412)

IN 1977, CHALMERS and colleagues[1] showed that nonrandomized studies yielded larger estimates of treatment effects than studies using random allocation. Others have found similar associations.[2-8] All those findings likely reflect the effects of inadequately controlled biases in nonrandomized studies.

Bias even appears to arise in trials labeled as "randomized" if investigators fail to prevent foreknowledge of treatment allocation. In 1983, Chalmers and colleagues[4] found that trials in which the allocation schedule had been inadequately concealed yielded larger estimates of treatment effects than trials in which allocation had been adequately concealed. Again, these results likely reflect bias but could have been due partly to confounding.[9]

A later study[10] addressed that confounding issue by restricting analysis to trials from meta-analyses with comparable treatments. Rather than concentrating on the adequacy of allocation concealment, however, Emerson et al[10] sought a relationship between "quality scores" intended to characterize the overall methodological quality of each trial[11] and estimates of treatment effects. Additionally, they suspected that methodologically inferior trials might produce bias in both directions, thereby causing greater variability in estimates of treatment effects. In neither analysis, however, did they detect a relationship.

Using a database of systematic reviews of controlled trials in pregnancy and childbirth,[12] we sought evidence of bias related to use of inadequate methodological approaches to trial design and execution. Rather than using quality scores, we investigated specific aspects that we believed might be influential.[13] We hypothesized that estimates of treatment effects would be larger in trials in which (1) adequate measures had not been taken to conceal treatment allocation; (2) adequate measures had not been taken to generate the allocation schedule; (3) some allocated participants had been excluded from the analysis; and (4) measures had not been taken to implement double-blinding. Furthermore, we examined whether treatment effects varied more in trials in which allocation schedules had not been adequately concealed.

MATERIALS AND METHODS

Derivation of Study Material

The systematic reviews of controlled trials used in this analysis have been published by the Pregnancy and Childbirth Group of the Cochrane Collaboration.[12,14,15] Published and unpublished primary trials potentially relevant for inclusion in the reviews were entered in a register.[16] Trials were eligible if some attempt to create unbiased comparison groups had been reported, either by randomization or by using a method such as alternation in a consecutive series, case record number, or date of birth. The register formed the basis for preparing systematic reviews.[17]

The database contained more than 500 systematic reviews,[12] but almost 60% of the reviews included just one or two trials. We derived a defined universe from all the reviews in three steps. First, we identified an initial subset of 82 meta-analyses, each of which included at least

From the Division of STD/HIV Prevention, National Center for Prevention Services, Centers for Disease Control and Prevention, Atlanta, Ga (Dr Schulz); the UK Cochrane Centre, Oxford, England (Drs Schulz and Chalmers); Department of Epidemiology and Population Sciences, London (England) School of Hygiene and Tropical Medicine (Dr Schulz and Mr Hayes); and Medical Statistics Laboratory, Imperial Cancer Research Fund, London (Mr Altman).

Presented in part at the 15th annual meeting of the Society for Clinical Trials, Houston, Tex, May 10, 1994, and at the Second International Cochrane Colloquium, Hamilton, Ontario, October 3, 1994.

Reprint requests to Division of STD/HIV Prevention, NCPS, Mailstop E-02, Centers for Disease Control and Prevention, Atlanta, GA 30333 (Dr Schulz).

five trials with a total of at least 25 outcome events among the control groups. Second, to ensure that data from each trial would contribute only once to the main analysis, we identified all meta-analyses to which component trials had contributed and retained only the meta-analysis with the most homogeneous grouping of interventions for inclusion in the analysis. Thus, we strove for similarity among trials within a meta-analysis to isolate more effectively those differences in treatment effects due to methodological quality. For example, a meta-analysis incorporating a specific class of antibiotics given for prophylaxis with cesarean delivery was included in preference to a meta-analysis that had included trials of any antibiotic used in that way. With only minor levels of overlap between two meta-analyses, we deleted an overlapping trial from one of them. That happened in only six instances; a random-number table determined the deletions. Third, the meta-analyses we included in our analysis had to comprise at least one component trial with adequate concealment (described herein) of the treatment allocation schedule and at least one trial without. We dropped from analysis 21 unpublished and 13 non–English-language trials due to difficulty in evaluating methodological quality.

Of the remaining 33 meta-analyses, two related to care during pregnancy, four to preterm labor and delivery, seven to induction of labor, six to labor and delivery, seven to prophylactic antibiotics for cesarean delivery, three to the puerperium, and four to the very early neonatal period. Each investigated similar comparison groups with the same binary outcome measure. The 33 meta-analyses included data from 250 primary trials involving a total of 62 091 participants and 12 030 outcome events. Of the 250 trials, 10% were published from 1955 to 1969, 18% from 1970 to 1979, 61% from 1980 to 1989, and 11% from 1990 to 1992.

Assessment of Trial Quality

Randomization, avoidance of exclusions after trial entry, and blinding have been proposed as the most important methodological components of controlled trials.[17] Randomization in a trial should involve both generating an unpredictable assignment sequence and concealing that sequence until allocation occurs. Allocation concealment appears the more important of those two aspects. It seeks to prevent selection bias, protects the assignment sequence before and until allocation, and can always be implemented.[18] A double-blind trial shields participants, care givers, and outcome evaluators from knowledge of treatment assignments. In contrast to allocation concealment, double-blinding seeks to prevent ascertainment bias, protects the sequence after allocation, and cannot always be implemented.[18]

One of us (K.F.S.) assessed the methodological quality of the 250 trial reports on the following four dimensions:

1. Concealment of Treatment Allocation Schedule.—Trials were divided into three groups: (1) Adequately concealed trials, the referent group, that were deemed to have taken adequate measures to conceal allocation (ie, central randomization; numbered or coded bottles or containers; drugs prepared by the pharmacy; serially numbered, opaque, sealed envelopes; or other description that contained elements convincing of concealment). (2) Inadequately concealed trials, in which concealment was inadequate (such as alternation or reference to case record numbers or to dates of birth). (3) Unclearly concealed trials, in which the authors either did not report an allocation concealment approach at all or reported an approach that did not fall into one of the categories just named. This group undoubtedly contained a mixture of inadequately and adequately concealed trials, but with the latter probably in the minority.[19]

2. Generation of Allocation Sequences.—Trials were divided into two groups: a referent group of trials with adequate sequence generation (reported using random-number table, computer random-number generator, coin tossing, or shuffling) and a second group that did not report one of the adequate approaches, those with inadequate sequence generation. We also analyzed this dimension in the subset of 79 trials that had reported adequate allocation concealment, because having a randomized (unpredictable) sequence should make little difference without adequate concealment. In trials with adequate concealment, particularly those in which group assignment becomes known after allocation (such as in unblinded trials), a randomized sequence may become important.

3. Inclusion in the Analysis of All Randomized Participants.—Trials were divided into two groups: a referent group of trials that reported, or gave the impression, that no exclusions had taken place (vast majority were not explicit) and a second group of trials that reported having made exclusions. The reasons for exclusions (when given) included protocol deviations, withdrawals, dropouts, and losses to follow-up.

4. Double-blinding.—Trials were divided into two groups: a referent group of trials that reported having been double-blinded and a second group that did not report as such, deemed not double-blinded. Authors provided meager information on the approaches used for double-blinding, so our classification was based necessarily on whether the reports purported to be accounts of double-blind trials.

We made special efforts to ensure consistency of quality assessments by using a single assessor with a detailed classification scheme and data abstraction instrument and by preventing the assessor from being aware of the results of the trials. That was achieved in assessing the descriptions of randomization and double-blinding in the articles, because authors embedded those descriptions in the methods sections. Preventing awareness was not always possible in assessing exclusions after trial entry, however, because authors frequently addressed those descriptions in the results sections.

The data were entered interactively into an *Epi Info* questionnaire.[20] To examine the reproducibility of items on the questionnaire, one of us (D.G.A.) reassessed a sample (computer random-number generator) of 10 trials unaware of the assessments made by the principal investigator (K.F.S.). Information on sequence generation, allocation concealment, blinding, and numbers of participants randomized and analyzed revealed no notable inconsistencies with the initial assessments. Information on whether a trial used intent-to-treat principles, however, revealed inconsistencies on two trials, largely because the reports forced that assessment to be made subjectively by not providing explicit information. Therefore, we did not analyze our assessments of intent to treat.

Statistical Methods

We used the GLIM 4 statistical system for modeling.[21] Simply, our modeling compared treatment effects (odds ratios [ORs]) yielded by subgroups of trials within each meta-analysis, and then those results were aggregated over all 33 sets of trials. Our multiple logistic regression models used data on the binary outcomes from the 250 trials and accounted for the effects of treatment, trials, and the different treatment effects in the 33 meta-analyses. In our first model, we added interaction terms[22] to address the question: "On average, do trials with inadequate or unclear allocation concealment yield different ORs relative to the referent group of adequately concealed trials?" In our next model, we added four interaction terms,[22] each addressing the question (controlling for the others): "On average, do trials judged to have been methodologically inferior yield different ORs from trials assigned to the referent category?" Interpretation of these interaction terms was simplified because in each of the

Table 1.—Reporting of Exclusions, Double-blinding, and Allocation Schedule Generation Related to the Level of Allocation Concealment for 250 Controlled Trials

	Adequately Concealed, % (n=79)	Unclearly Concealed, % (n=150)	Inadequately Concealed, % (n=21)	All Trials (n=250)
Authors reported				
No apparent exclusions	53	67	76	63
Double-blinding	73	39	14	48
Adequate generation of allocation schedule	29	15	0	18

individual 33 meta-analyses the treatment intervention was more effective than the control in preventing an adverse outcome (a summary OR in each of less than 1.0). A ratio of ORs (ROR) of less than 1.0 for an interaction term indicated that trials that were methodologically inferior had yielded larger (exaggerated) estimates of treatment effects, on average, compared with the referent group. Conversely, an ROR greater than 1.0 indicated association with smaller treatment effects.

To analyze the variability (heterogeneity) of the three allocation concealment groups, we fitted a separate model to each. The models accounted for the effects of treatment, trials, meta-analyses, and the different summary ORs in the meta-analyses. The deviance divided by its degrees of freedom was regarded as an approximate measure of overdispersion, reflecting the degree of heterogeneity between trials. We used approximate F ratio tests to compare the heterogeneity of unclearly and inadequately concealed trials with the adequately concealed trials.[23] Although approximate, these tests appear appropriate for these data.[23]

In separate models not shown, we used an adjustment to the scale parameter to take rough account of overdispersion (extrabinomial variation) in estimating SEs[24] for the effects of inadequately and unclearly concealed trials. While yielding wider confidence intervals (CIs),[23] our basic conclusions remained unchanged.

RESULTS

Steps taken to conceal treatment allocation schedules were adequate in 79 trials, unclear in 150, and inadequate in 21 (Table 1). The trials reporting adequately concealed allocation were the most likely to have reported excluding participants after allocation whereas the trials reporting inadequately concealed allocation were the least likely. The unclearly concealed trials reported at an intermediate level. Double-blinding and having used adequate sequence generation, however, were much more common in the adequately concealed trials than in the other groups.

Trials with inadequate or unclear allocation concealment yielded exaggerated estimates of treatment effects, on average, compared with trials that had taken adequate measures to conceal allocation (Table 2). Furthermore, ORs derived from trials that had reported inadequately concealed allocation were more variable (heterogeneous) than those from adequately concealed trials (P=.001; F=4.27; df_1=7, df_2=46), whereas ORs derived from unclearly concealed trials were not (P=.19; F=1.26; df_1=117, df_2=46).

We excluded inadequately concealed trials from further analyses for three main reasons. First, they had more heterogeneous estimates of treatment effects. Second, including them in further analyses makes little theoretical sense. For example, analyzing the effect of double-blinding in inadequately concealed trials would be unjustified because double-blinding would likely be impossible under such circumstances. Third, they are not randomized trials.

Trials with unclear allocation concealment still yielded exaggerated estimates of treatment effects after accounting for sequence generation, postallocation exclusions, and double-blinding (Table 3). The ROR was 0.70, which means that the ORs in the unclearly concealed trials were, on average, 30% lower than in the adequately concealed trials, ie, estimating larger treatment effects.

That 30% lower result for unclearly concealed trials represents a weighting and combining of the results from the 33 meta-analyses. To put that overall result in context, we present two individual examples. The first, a meta-analysis of polyglycolic acid vs catgut for perineal repair on short-term pain, had five trials. One adequately concealed trial yielded an OR of 0.89, a protective effect of polyglycolic acid, whereas the four unclearly concealed trials yielded an aggregated OR of 0.44. That OR was 51% lower than in the adequately concealed trial (ROR, 0.49; 95% CI, 0.35 to 0.69), apparently exaggerating the protective effect of polyglycolic acid. The second meta-analysis compared the effect of corticosteroids with no corticosteroids, after preterm rupture of membranes, on respiratory distress syndrome. Three adequately concealed trials yielded an aggregated OR of 0.72, a protective effect of corticosteroids, whereas the four unclearly concealed trials yielded an aggregated OR of 0.53. That OR was 27% lower than in the adequately concealed trials (ROR, 0.73; 95% CI, 0.35 to 1.5). That exaggeration of the protective effect closely resembles the overall average effect for unclearly concealed trials of 30%.

Examining 33 separate logistic regression analyses, the estimates of treatment effects for the unclearly concealed trials were larger (exaggerated) in 27 meta-analyses and smaller in six than the effects derived from the referent group of concealed trials. The effect of unclear allocation concealment varied among the 33 sets of trials by more than would be expected by chance (P=.01).

Trials with inadequate sequence generation yielded estimates of treatment effects that were similar to those derived from trials with adequate sequence generation, after adjusting for the other three methodological dimensions (Table 3). However, limiting analysis to just those 79 trials that had reported adequately concealed allocation, trials with inadequate sequence generation yielded larger estimates of effects, on average, than trials with adequate sequence generation (ROR, 0.75; 95% CI, 0.55 to 1.02; P=.07).

Adjusting for allocation concealment, sequence generation, and double-blinding (Table 3), trials known to have excluded participants yielded estimates of treatment effects that were similar, on average, to those derived from trials that apparently had not excluded any participants. However, trials that had not been double-blinded yielded estimates of treatment effects that were larger, on average, than those derived from trials that had been double-blinded, adjusting for the other three methodological elements (Table 3). Odds ratios were exaggerated by 17%, on average, in those trials that were not double-blinded.

COMMENT

Comparing interventions can be misleading unless investigators take precautions to ensure that their studies contain unbiased comparison groups. Random allocation remains the only way to eliminate selection biases. That unique strength is of crucial importance in the common circumstances in which the treatment effects may be of comparable magnitude to the biases that occur in most nonrandomized comparisons of health care alternatives.

Surprisingly, in view of the central importance of randomization, authors often provide inadequate details of the steps taken to assign participants to comparison groups,[25-27] including authors in the medical specialty represented in this analysis.[18] Because randomization prevents selection bias, trials that have failed

Table 2.—Odds Ratios in the Unclearly and Inadequately Concealed Trials Compared With Those in Adequately Concealed Trials*

Level of Allocation Concealment	Ratio of Odds Ratios (95% Confidence Interval)	χ^2 (*df*)	*P*
Adequate	1.00 (referent)		
Unclear	0.67 (0.60-0.75)	57.9 (2)	<.001
Inadequate	0.59 (0.48-0.73)		

*Multiple logistic regression model with the dependent variable being binary outcome measures from each meta-analysis. The independent variables included a binary variable for treatment group (experimental vs control); indicator variables to control for the effects of each of the 250 trials; terms for the "meta-analysis by treatment group" interaction to control for the different summary odds ratios for the treatment effects in the 33 meta-analyses; and the "allocation concealment by treatment" interaction terms displayed in this table to analyze their associations with estimates of treatment effects. Model deviance=434.2; *df*=215.

Table 3.—Association Between Four Dimensions of Methodological Quality and Estimates of Treatment Effects in the 229 Adequately and Unclearly Concealed Trials*

Measure of Methodological Quality	Ratio of Odds Ratios (95% Confidence Interval)	χ^2 (*df*)	*P*
Allocation concealment			
Adequate	1.00 (referent)	32.9 (1)	<.001
Unclear	0.70 (0.62-0.79)		
Sequence generation			
Adequate	1.00 (referent)	0.31 (1)	.58
Inadequate	0.95 (0.81-1.12)		
Exclusions			
No	1.00 (referent)	0.99 (1)	.32
Yes	1.07 (0.94-1.21)		
Double-blinded			
Yes	1.00 (referent)	6.16 (1)	.01
No	0.83 (0.71-0.96)		

*Multiple logistic regression model with the dependent variable being binary outcome measures from each meta-analysis. The independent variables included a binary variable for treatment group (experimental vs control); indicator variables to control for the effects of each of the 229 trials; terms for the "meta-analysis by treatment group" interaction to control for the different summary odds ratios for the treatment effects in the 33 meta-analyses; and the four "quality measure by treatment" interaction terms displayed in this table to analyze their associations with estimates of treatment effects. Model deviance=325.3; *df*=192.

to ensure proper randomization should yield systematically different estimates of treatment effects than those estimates derived from trials that have used adequate approaches. Our analyses, using the available information in reports of controlled trials, support that hypothesis. We also found that estimates of treatment effects were larger in trials that had not reported double-blinding.

Peto[28] has warned of potential biases of 30% with nonrandomized studies. His concern also pertains to poorly executed randomized trials and, based on our study, appears appropriate. Trials that reported either inadequate or unclear concealment methods yielded estimates of ORs that were exaggerated by an average of 41% or 30%, respectively, compared with estimates of ORs derived from trials that apparently had taken adequate steps to conceal treatment allocation. These associations probably reflect selection biases. Further evidence of bias comes from the greater heterogeneity of treatment effects displayed by the inadequately concealed trials compared with the adequately concealed trials. This variability reflects biases that inconsistently bounce in both directions, which is particularly pernicious.

Inadequate concealment can lead to introduction of bias in many ways, sometimes as the result of deliberate subversions (usually well intentioned), sometimes as the net result of subconscious actions. For example, if those responsible for admitting participants have foreknowledge of treatment allocations, they may channel participants with a better prognosis to the experimental group and those with a poorer prognosis to the control group, or vice versa. That could easily be accomplished by delaying a participant's entry into the trial until the next desired allocation appears or by excluding eligible participants from the trial or encouraging them to refuse entry. Without allocation concealment, biases in either direction become possible, although we found a clear tendency toward exaggerated estimates of treatment effects.

The effects yielded by unclearly concealed trials varied by more than chance among the meta-analyses. Therefore, while an exaggeration of 30% in ORs for the unclearly concealed trials appropriately estimates the average association, it should not be interpreted as representing each meta-analysis. Furthermore, distortions in ORs, in either direction, should not be interpreted to mean that distortions in relative risks would be of a similar magnitude. The proportion of outcome events in the control groups of the trials we studied was almost 20% overall and ranged from less than 1% to more than 90%. An exaggeration of 30% in an OR could translate to a much lower level of exaggeration in a relative risk.

Inadequate allocation concealment also may be a surrogate measure for the quality of other aspects of trial design and execution. Thus, the magnitude of the associations we have observed may reflect biases other than selection biases. In any case, our results support the policy decision taken by one journal not to publish reports of trials with inadequate allocation concealment (not truly randomized).[29] Our findings emphasize the importance of securing adequate allocation concealment and of ensuring that authors and journal editors publish reports that make those aspects of trial design explicit.

That inadequate sequence generation had little effect on estimation of treatment effects augments the case for the greater importance of allocation concealment. However, in the 79 trials that used adequate allocation concealment, trials that did not report an adequate approach to sequence generation yielded larger estimates of treatment effects. That finding implies that an unpredictable (random) sequence helps to protect against bias if steps are taken to conceal the allocation sequence. We advise caution in interpreting this result, however, because of the wide CI that includes the possibility of no effect.

Contrary to our prior hypothesis, trials that reported having excluded participants after randomization did not yield exaggerated estimates of treatment effects compared with trials in which the reports gave the impression of no exclusions. That unexpected finding could be because some authors inappropriately reported that they had randomized the same number of participants as they had analyzed, even though some randomized participants actually had been excluded. That interpretation receives support from the paradoxical finding that trials using adequately concealed allocation were the most likely to report excluding participants and that trials using inadequately concealed allocation were the least likely to report excluding participants (Table 1). Of those trial reports that gave the impression that no exclusions had occurred, few explicitly stated that no exclusions had occurred. This all suggests that published information on exclusions may currently have little value in assessing trial quality, and users of randomized controlled trials should be wary of the potentially misleading information currently provided.

Indeed, Gøtzsche[30] warned of that misleading information when he detected nonreporting of exclusions in two separate published trials. Investigators must report accurately the number of exclusions and, if none took place, must state so explicitly. They also must explicitly

state their approach to the analysis, primarily whether they performed an intent-to-treat analysis.

Trials for which no double-blinding was reported yielded estimates of ORs that were exaggerated by 17%, on average, compared with trials that reported having used double-blinding. Trials that reported double-blinding usually provided little, if any, information on the methods used. Consequently, some trials claiming to be double-blind may not have been, and so misclassification error could have caused an underestimate of the independent effect of not double-blinding. In any case, blinding should be of greater importance to minimizing bias for some outcomes than for others.

The selection criteria for our subset of meta-analyses could have affected our results. For example, we selected only published trials. We found those trials with certain indications of inadequate methodological quality tended to report more extreme beneficial effects. Those results potentially could have materialized, in part, because of publication bias if trials of lower methodological quality tended to be published more often if their results were more extreme.

We sought primarily internal consistency in our assessments of quality and achieved that by using one assessor. We examined interassessor reproducibility in a small sample to identify any major potential problems, not to precisely estimate reproducibility. To obtain precise estimates would require a larger sample size. Our measures appeared reasonable except for the one on intent to treat. Moreover, our measures seemed reliable in other studies.[18,27]

Thus, our study could not be completely shielded from biases. We doubt, however, that those potential biases could account fully for our findings. Moreover, some of the trials with unclear concealment may have used an acceptable method of concealment but simply did not report it. Thus, the true bias associated with lack of concealment could be higher than we observed. Furthermore, we feel our results can be cautiously generalized beyond the perinatal field. In medical specialties where generalizing appears questionable, we encourage replication of this study.

Our results for allocation concealment are consistent with a previous finding[4] but, unlike those,[9] cannot be attributable to confounding by type of treatment. Our analysis accounted for the effects of the 33 meta-analyses, each of which was composed of trials that had investigated similar treatments. Our findings do not necessarily conflict with the findings of Emerson et al[10] because those investigators used a quality score[11] that quantifies many aspects of trial design and analysis, some of which were unlikely to be related to bias.[13] Also, their inclusion of both study size and quality score as explanatory variables in their model could have obscured the effects of quality relating to bias because study size correlated with quality score.

CONCLUSIONS

Our findings not only highlight the importance of adequate methodological quality in controlled trials but also the importance of complete and reliable reporting. Without adequate reporting, assessing quality becomes impossible. Authors should, as a minimum, explicitly describe their approaches to sequence generation, allocation concealment, blinding, and handling of exclusions after allocation.[18,31]

Preventing foreknowledge of treatment allocation by effective concealment of allocation schedules in controlled trials emerges from our analyses as crucially important to reducing bias. Without proper application of measures to achieve concealment, the whole point of randomization vanishes and bias is likely to distort results. Our results support the comment of Mosteller and colleagues[25]: "When the randomization leaks, the trial's guarantee of lack of bias runs down the drain." Investigators, editors, and readers need to be made aware of the importance of allocation concealment. With greater methodological vigilance, more randomized trials will actually fulfill their promise of minimizing bias.

This work was supported by the Centers for Disease Control and Prevention. Dr Chalmers was supported by the National Health Service Research and Development Programme. We thank Mark Starr, PhD, of Update Software for his assistance in identifying the meta-analyses for analysis. We gratefully acknowledge suggestions on an earlier draft provided by Simon Thompson, MA, DipStat, of the Medical Statistics Unit, London (England) School of Hygiene and Tropical Medicine, and the invaluable assistance provided by Jini Hetherington and Sally Hunt of the UK Cochrane Centre in identifying the data from the meta-analyses.

References

1. Chalmers TC, Matta RJ, Smith H Jr, Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med.* 1977;297:1091-1096.
2. Pocock SJ. Allocation of patients to treatment in clinical trials. *Biometrics.* 1979;35:183-197.
3. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233-240.
4. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1983;309:1358-1361.
5. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy, II: surgical. *Stat Med.* 1989;8:455-466.
6. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy, I: medical. *Stat Med.* 1989;8:441-454.
7. Devine EC, Cook TD. A meta-analytic analysis of effects of psychoeducational interventions on length of post-surgical hospital stay. *Nurs Res.* 1983; 32:267-274.
8. Wortman PM, Yeaton WH. Synthesis of results in controlled trials of coronary artery bypass graft surgery. In: Light RJ, ed. *Evaluation Studies Review Annual.* Beverly Hills, Calif: Sage; 1983;8:536-551.
9. Gillman MW, Runyan DK. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1984;310:1610-1611.
10. 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;11:339-352.
11. Chalmers TC, Smith H Jr, Blackburn B, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials.* 1981;2:31-49.
12. Chalmers I, ed. *Oxford Database of Perinatal Trials: Version 1.2, Disk Issue 8.* New York, NY: Oxford University Press Inc; 1992.
13. Greenland S. A critical look at some popular meta-analytic methods. *Am J Epidemiol.* 1994;140: 290-296.
14. Chalmers I, Enkin M, Keirse MJNC, eds. *Effective Care in Pregnancy and Childbirth.* New York, NY: Oxford University Press Inc; 1989.
15. Enkin MW, Keirse MJNC, Renfrew MJ, Neilson JP. *Cochrane Database of Systematic Reviews.* Oxford, England: Update Software; 1993.
16. Chalmers I, Hetherington J, Newdick M, et al. The Oxford database of perinatal trials: developing a register of published reports of controlled trials. *Control Clin Trials.* 1986;7:306-325.
17. Chalmers I, Hetherington J, Elbourne D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJN, eds. *Effective Care in Pregnancy and Childbirth, Volume 1: Pregnancy.* New York, NY: Oxford University Press Inc; 1989:39-65.
18. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA.* 1994;272:125-128.
19. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol.* 1986;4: 942-951.
20. Dean AG, Dean JA, Burton AH, Dicker RC. *Epi Info, Version 5: A Word Processing, Database, and Statistics Program for Epidemiology on Microcomputers.* Atlanta, Ga: Centers for Disease Control and Prevention; 1990.
21. Francis B, Green M, Payne C, et al. *GLIM 4: The Statistical System for Generalized Linear Interactive Modelling.* New York, NY: Oxford University Press Inc; 1993.
22. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol.* 1992;45:255-265.
23. Schulz KF. *Methodological Quality and Bias in Randomised Controlled Trials.* London, England: University of London; 1994. Thesis.
24. Aitkin M, Anderson D, Francis B, Hinde J. *Statistical Modeling in GLIM.* New York, NY: Oxford University Press Inc; 1989.
25. Mosteller F, Gilbert JP, McPeek B. Reporting standards and research strategies for controlled trials. *Control Clin Trials.* 1980;1:37-58.
26. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med.* 1982;306:1332-1337.
27. Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet.* 1990;335: 149-153.
28. Peto R. Why do we need systematic overviews of randomized trials? *Stat Med.* 1987;6:233-244.
29. Altman DG. Randomisation: essential for reducing bias. *BMJ.* 1991;302:1481-1482.
30. Gøtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol.* 1989;36:429-432.
31. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA.* 1994;272:1926-1931.