Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials

Comparison of Protocols to Published Articles

An-Wen Chan, MD, DPhil
Asbjørn Hróbjartsson, MD, PhD
Mette T. Haahr, BSc
Peter C. Gøtzsche, MD, DrMedSci
Douglas G. Altman, DSc

Context Selective reporting of outcomes within published studies based on the nature or direction of their results has been widely suspected, but direct evidence of such bias is currently limited to case reports.

Objective To study empirically the extent and nature of outcome reporting bias in a cohort of randomized trials.

Design Cohort study using protocols and published reports of randomized trials approved by the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark, in 1994-1995. The number and characteristics of reported and unreported trial outcomes were recorded from protocols, journal articles, and a survey of trialists. An outcome was considered incompletely reported if insufficient data were presented in the published articles for meta-analysis. Odds ratios relating the completeness of outcome reporting to statistical significance were calculated for each trial and then pooled to provide an overall estimate of bias. Protocols and published articles were also compared to identify discrepancies in primary outcomes.

Main Outcome Measures Completeness of reporting of efficacy and harm outcomes and of statistically significant vs nonsignificant outcomes; consistency between primary outcomes defined in the most recent protocols and those defined in published articles.

Results One hundred two trials with 122 published journal articles and 3736 outcomes were identified. Overall, 50% of efficacy and 65% of harm outcomes per trial were incompletely reported. Statistically significant outcomes had a higher odds of being fully reported compared with nonsignificant outcomes for both efficacy (pooled odds ratio, 2.4; 95% confidence interval [CI], 1.4-4.0) and harm (pooled odds ratio, 4.7; 95% CI, 1.8-12.0) data. In comparing published articles with protocols, 62% of trials had at least 1 primary outcome that was changed, introduced, or omitted. Eighty-six percent of survey responders (42/49) denied the existence of unreported outcomes despite clear evidence to the contrary.

Conclusions The reporting of trial outcomes is not only frequently incomplete but also biased and inconsistent with protocols. Published articles, as well as reviews that incorporate them, may therefore be unreliable and overestimate the benefits of an intervention. To ensure transparency, planned trials should be registered and protocols should be made publicly available prior to trial completion.

JAMA. 2004;291:2457-2465 www.jama.com

SELECTIVE PUBLICATION OF STUDIES with statistically significant results has received widespread recognition.[1] In contrast, selective reporting of favorable outcomes within published studies has not undergone comparable empirical investigation. The existence of outcome reporting bias has been widely suspected for years,[2-12] but direct evidence is limited to case reports that have low generalizability[13-15] and may themselves be subject to publication bias.

Our study had 3 goals: (1) to determine the prevalence of incomplete outcome reporting in published reports of randomized trials; (2) to assess the association between outcome reporting and statistical significance; and (3) to evaluate the consistency between primary outcomes specified in trial protocols and those defined in the published articles.

METHODS

In February 2003, we identified protocols and protocol amendments for randomized trials by reviewing paper files from clinical studies approved by the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark, in 1994-1995. This period was

Author Affiliations: Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, England (Drs Chan and Altman), Nordic Cochrane Centre, Copenhagen, Denmark (Drs Hróbjartsson and Gøtzsche and Ms Haahr), University Health Network, University of Toronto, Toronto, Ontario (Dr Chan).
Corresponding Author: An-Wen Chan, MD, DPhil, Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Headington, Oxford, England OX3 7LF (anwen.chan@utoronto.ca).

Table 1. Hierarchy of Levels of Outcome Reporting

Levels of Outcome Reporting	Reported Data	Data Sufficient for Inclusion in Meta-analysis
Full	No. of participants per group Effect size Precision or precise *P* value for continuous data (see Box)*	Yes
Incomplete		
Partial	Effect size or precision (± sample size and/or *P* value)†	No
Qualitative	*P* value (± sample size)†	No
Unreported	None	No

*Precise *P* value enables the calculation of the standard error if the treatment effect and sample sizes are given.
†Items in parentheses indicate "optional" data, ie, those not necessary or not sufficient on their own to meet the requirements for the particular definition.

chosen to allow sufficient time for trial completion and publication. A randomized trial was defined as a prospective study assessing the therapeutic, preventative, adverse, pharmacokinetic, or physiological effects of 1 or more health care interventions and allocating human participants to study groups using a random method. Pharmocokinetic trials measured primarily the kinetics of drug metabolism and excretion; physiological trials, with the exception of preventative trials, examined the effect of interventions on healthy volunteers rather than in the intended disease or at-risk population. Studies were included if they simply claimed to allocate participants randomly or if they described a truly random sequence of allocation. Pseudorandom methods of allocation, such as alternation or the use of date or case numbers, were deemed inadequate for inclusion.

Trials with at least 1 identified journal article were included in our study cohort. Publication in journals was identified by contacting trialists and by searching MEDLINE, EMBASE, and the Cochrane Controlled Trials Register using investigator names and keywords (final search, May 2003). For each trial, we included all published articles reporting final results. Abstracts and reports of preliminary results were excluded.

For each published trial, we reviewed the study protocol, any amendments, and all published articles to extract the trial characteristics, the number and nature of reported outcomes (including statistical significance, completeness of reporting, and specification as primary/secondary), as well as the number and specification of unreported outcomes. Data from amendments took precedence over data from earlier protocols.

An outcome was defined as a variable that was intended for comparison between randomized groups in order to assess the efficacy or harm of an intervention. We prefer the term "harm" rather than "safety" because all interventions can be potentially harmful. Unreported outcomes were those that were specified in the most recent protocol but were not reported in any of the published articles, or that were mentioned in the "Methods" but not the "Results" sections of any of the published articles. Their statistical significance and the reasons for omitting them were solicited from contact authors through a prepiloted questionnaire. We initially asked whether there were any outcomes that were intended for comparison between randomized groups but were not reported in any published articles, excluding characteristics used only for assessment of baseline comparability. We subsequently provided trialists with a list of unreported outcomes identified from our comparison of protocols with published articles. Double-checking of outcome data extraction from a random subset of 20 trials resulted in corrections to 21 of 362 outcomes (6%), 15 of which were in a single trial.

We classified the level of outcome reporting in 4 groups based on data provided across all published articles of a trial (TABLE 1). A fully reported outcome was one with sufficient data for inclusion in a meta-analysis. The nature and amount of data required to meet this criterion vary depending on the data type (BOX 1). Partially reported outcomes had some of the necessary data for meta-analysis, while qualitatively reported outcomes had no useful data except for a *P* value or a statement regarding the presence or absence of statistical significance. Unreported outcomes were those for which no data were provided in any published articles despite having been specified in the protocol or the "Methods" sections of the published articles.

We defined 2 additional terms to describe relevant composite levels of reporting (Table 1). *Reported outcomes* were defined as those with at least some data presented (full, partial, and qualitative). *Incompletely reported outcomes* were defined as those that were inadequately reported for meta-analysis (partial, qualitative, and unreported).

Analyses were conducted at the trial level and stratified by study design using Stata 7 (Stata Corp, College Station, Tex). Efficacy and harm outcomes were evaluated separately. The reasons given by trialists for not reporting outcomes were tabulated, and the proportion of unreported and incompletely reported outcomes per trial was determined.

For each trial, we tabulated all outcomes in a 2×2 table relating the level of outcome reporting (full vs incomplete) to statistical significance ($P<.05$ vs $P \geq .05$). Outcomes were ineligible if their statistical significance was unknown. An odds ratio was then calculated from the 2×2 table for every trial, except when any entire row or column total was zero. If the table included a single cell frequency of zero or 2 diagonal cell frequencies of zero, we added 0.5 to all 4 cell frequencies.[16,17] Odds ratios greater than 1.0 meant that statistically significant outcomes had a higher odds of being fully reported compared with nonsignificant outcomes. The odds ratios from each trial were pooled using a random-

effects meta-analysis to provide an overall estimate of bias. Exploratory meta-regression was used to examine the effect of funding source, sample size, and number of study centers on the magnitude of bias. Sensitivity analyses were conducted to assess the robustness of the odds ratios when (1) nonresponders to the survey were excluded; (2) pharmacokinetic and physiological trials were excluded; and (3) the level of reporting was dichotomized using a different cutoff (fully or partially reported vs qualitatively reported or unreported).

Finally, we evaluated the consistency between primary outcomes specified in the most recent trial protocols (including amendments) and those defined in the published articles. Primary outcomes consisted of those that were defined explicitly as such in the protocol or published article. If none was explicitly defined, we used the outcome stated in the power calculation. We defined major discrepancies as those in which (1) a prespecified primary outcome was reported as secondary or was not labeled as either; (2) a prespecified primary outcome was omitted from the published articles; (3) a new primary outcome was introduced in the published articles; and (4) the outcome used in the power calculation was not the same in the protocol and the published articles. A discrepancy was said to favor statistically significant results if a new statistically significant primary outcome was introduced in the published articles or if a nonsignificant primary outcome was omitted or defined as nonprimary in the published articles. Discrepancies were verified by 2 independent researchers, with disagreements resolved by consensus. Double-checking resulted in major corrections for 3 of 259 primary outcomes (1%).

RESULTS

We identified 1403 applications submitted to the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark, in 1994-1995 (FIGURE 1). We excluded 1129 studies (637 nonrandomized studies, 408 amendments to previous studies, 43 not reviewed by the committee, 30 files not found, 9 not approved, 1 diagnostic test study, and 1 duplicate file), leaving 274 protocols for randomized trials. Of these, 172 trials were not begun, not completed, or not published (96 confirmed by survey response and negative literature search results and 76 confirmed by negative literature search results alone). Thus, 102 published articles were included (54 confirmed by survey response and literature searches and 48 confirmed by literature searches alone).

Box 1. Data Required for Meta-analysis of Fully Reported Outcomes

For Unpaired Continuous Data

Sample size in each group
and
Magnitude of treatment effect (group means/medians or difference in means/medians)
and
Measure of precision or variability (confidence interval, standard deviation, or standard error for means; interquartile or other range for medians) or the precise *P* value*

For Unpaired Binary Data

Sample size in each group
and
Either the numbers (or percentages) of participants with the event for each group, or the odds ratio or relative risk with a measure of precision or variability (confidence interval, standard deviation, or standard error) or the precise *P* value*

For Paired Continuous Data

Sample size in each group
and
Either the raw data for each participant, or the mean difference between groups and a measure of its precision or variability or the precise *P* value

For Paired Binary Data

Sample size in each group
and
Paired numbers of participants with and without events

For Survival Data

Either a Kaplan-Meier curve or similar, with numbers of patients at risk over time, or a hazard ratio with a measure of precision and sample size in each group

*Sample sizes, treatment effect, and precise *P* value enable the calculation of a standard error if a measure of precision or variability is not reported.

Figure 1. Identification of Published Articles of Randomized Trials Approved by the Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark: 1994-1995

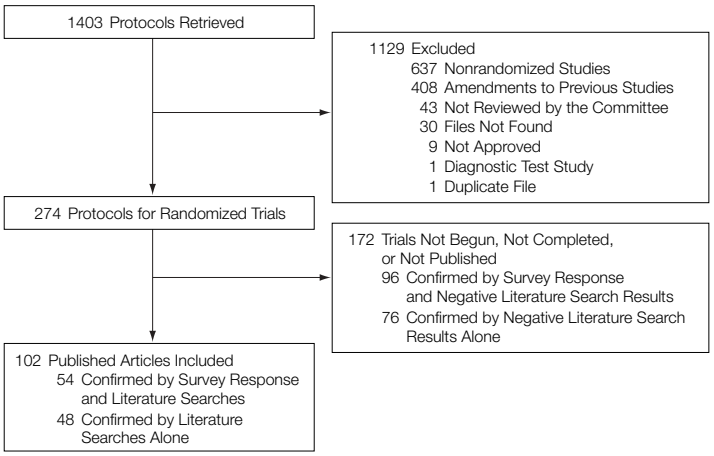
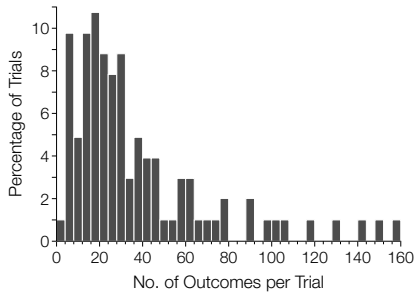


Table 2. Characteristics of the Included Trials (N = 102)

Characteristic	Trials, No. (%)
Study design	
Parallel-group	70 (69)
Crossover	30 (29)
Other*	2 (2)
Intervention	
Drug	77 (75)
Surgery/procedure	11 (11)
Counseling/lifestyle	12 (12)
Equipment	2 (2)
Study centers	
Single	53 (52)
Multiple	49 (48)
Funding	
Full industry	56 (55)
Partial industry	17 (17)
Nonindustry	22 (22)
None declared	7 (7)

*Split-body (n = 1) and factorial (n = 1) trials.

Figure 2. Total Number of Outcomes per Trial



ies, primarily because they were not randomized trials or were amendments to studies submitted before 1994. Thirty files (2%) could not be located; it is unclear whether they would have been eligible for inclusion. We found 274 randomized trial protocols, but 172 (63%) were never begun or completed, or were unpublished according to our literature searches and survey of trialists. The final cohort consisted of 102 trials with 122 published articles. Published articles for 48 of the 102 trials were identified by literature search alone, as the trialists did not respond to our request for information (Figure 1).

Trial characteristics are shown in TABLE 2. The majority were of parallel-group design, and most investigated drug interventions. One half were funded solely by industry, and one half were multicenter studies. Published articles for 39% of the trials listed contact authors located at centers outside of Denmark. The median sample size was 151 (10th-90th percentile range, 28-935) for parallel-group trials, and 16 (10th-90th percentile range, 7-43) for crossover trials.

All but 3 trials were published in specialty journals rather than in general medical journals—the latter being defined as those publishing articles from any clinical field. Fifteen trials had more than 1 published article. The publication year of the first article from each of the 102 trials ranged from 1995 to 2003. Two appeared in 1995-1996; 13 in 1997; 52 in 1998-1999; 27 in 2000-2001; 7 in 2002; and 1 in 2003.

Across the 102 trials, we identified 3736 outcomes (median, 27 per trial; 10th-90th percentile range, 7-79) from the protocols and the published articles (FIGURE 2). Ninety-nine trials measured efficacy outcomes (median, 20; 10th-90th percentile range, 5-63 per trial), and 72 trials measured harm outcomes (median, 6; 10th-90th percentile range, 1-31 per trial).

Prevalence of Unreported Outcomes

Only 48% (49/102) of trialists responded to the questionnaire regarding unreported outcomes, 86% (42/49) of whom initially denied the existence of such outcomes prior to receiving our list of unreported outcomes. However, all 42 of these trials had clear evidence of unreported outcomes in their protocols and in the published articles. None of the responders added any unreported outcomes to the list we subsequently provided.

Among trials that measured efficacy or harm outcomes, 71% (70/99) and 60% (43/72) had at least 1 unreported efficacy or harm outcome, respectively (ie, outcomes missing in "Results" sections of published articles but listed in the protocols or in the "Methods" sections of the published articles). In these trials, a median of 4 (10th-90th percentile range, 1-25; n=70 trials) efficacy outcomes and 3 (10th-90th percentile range, 1-18; n=43 trials) harm outcomes were unreported.

Among 78 trials with any unreported outcome (efficacy or harm or both), we received only 24 survey responses (31%) that provided reasons for not reporting outcomes for efficacy (23 trials) or harm (10 trials) in their published articles. The most common reasons for not reporting efficacy outcomes were lack of statistical significance (7/23 trials), journal space restrictions (7/23), and lack of clinical importance (7/23). Similar reasons were provided for harm data.

Prevalence of Incompletely Reported Outcomes

Ninety-two percent (91/99) of trials had at least 1 incompletely reported efficacy outcome, while 81% (58/72) had at least 1 incompletely reported harm outcome. Primary outcomes were specified for 63 of the published trials, but for 17 (27%) of these trials at least 1 primary outcome was incompletely reported. The median proportion of incompletely reported outcomes per trial was 50% (10th-90th percentile range, 4%-100%) for efficacy outcomes and 65% (10th-90th percentile range, 0%-100%) for harm outcomes (TABLE 3). Incomplete reporting was common even when the total number of measured trial outcomes was low, and was more common for crossover trials than for parallel-group trials (Table 3).

Association Between Completeness of Reporting and Statistical Significance

Forty-nine trials could not contribute to the analysis of reporting bias for efficacy outcomes because they had entire rows or columns that were empty in the 2×2 table (analogous to a trial assessing mortality but with no observed deaths); 54 trials were similarly noncontributory for harm outcomes. Included trials were similar to excluded trials, except the former had a lower proportion of crossover trials and a higher number of eligible outcomes per trial. Six hundred ten of 2785 efficacy outcomes (22%) and 346 of 951 harm outcomes (36%) were ineligible for analysis because their statistical sig-

nificance was unknown; only 11 trialists provided information about whether their unreported outcomes were statistically significant.

The odds ratio for outcome reporting bias in each trial is displayed in FIGURE 3. The pooled odds ratio (95% confidence interval) for trials of any design was 2.4 (1.4-4.0) for efficacy outcomes and 4.7 (1.8-12.0) for harm outcomes (TABLE 4). Thus, the odds of a particular outcome being fully reported was more than twice as high if that outcome was statistically significant. Stratifying by study design, or excluding survey nonresponders or physiologic/pharmacokinetic trials, had no important impact on the odds ratios (Table 4). Dichotomizing the level of reporting differently by combining fully reported with partially reported outcomes increased the degree of bias (Table 4). Exploratory meta-regression analysis did not reveal any significant associations between the magnitude of bias and the source of funding, sample size, or number of study centers.

Consistency Between Primary Outcomes in Protocols and Published Articles

Formal protocol amendments involving study outcomes were submitted to the ethics committee for approval for 7 trials. Most changes involved secondary outcomes, with a primary outcome being formally amended in only 2 trials. Primary outcomes were defined for 82 of the 102 trials (80%), either in the protocol or in the published articles. Among 63 trials defining primary outcomes in their published articles, 39 (62%) defined 1 primary outcome, 7 (11%) defined 2, and 17 (27%) defined more than 2.

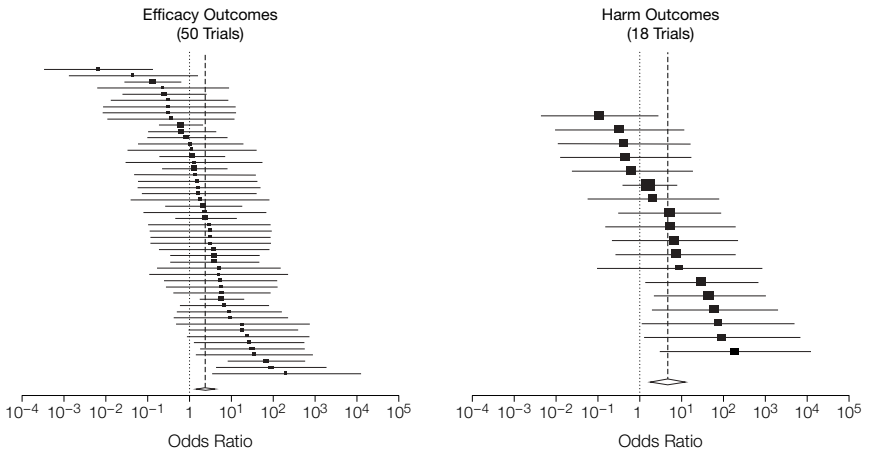
Overall, 51 of the 82 trials (62%) had major discrepancies between the primary outcomes specified in protocols and those defined in the published articles (TABLE 5). Specific examples of major discrepancies are shown in BOX 2. For 26 trials, protocol-defined primary outcomes were reported as nonprimary in the published articles, while for 20 trials primary outcomes were omitted. For 12 trials, outcomes that had been predefined as nonprimary in the protocol were called primary in the published articles. For 11 trials, new primary outcomes that were not even mentioned in the protocol appeared in the published articles. None

Table 3. Median Proportion of Incompletely Reported Efficacy and Harm Outcomes per Trial, by Study Design

	Incompletely Reported Outcomes per Trial, Median Percentage (10th-90th Percentile Range)			
Trial Design	No. of Trials	Efficacy Outcomes	No. of Trials	Harm Outcomes
All	99	50 (4-100)	72	65 (0-100)
Parallel-group	68	38 (4-78)	57	50 (0-100)
Crossover	29	98 (0-100)	14	100 (0-100)
Other	2	91 (82-100)	1	71 (NA)

Abbreviation: NA, not available because cannot be calculated.

Figure 3. Odds Ratios for Outcome Reporting Bias Involving Efficacy and Harm Outcomes



Black squares indicate odds ratios; horizontal lines, 95% confidence intervals; diamonds and dashed lines, pooled odds ratios. The size of each square reflects the statistical weight of a trial in calculating the pooled odds ratio, and the relative sizes of the squares are accurate within each plot only.

Table 4. Pooled Odds Ratio for Outcome Reporting Bias (Fully vs Incompletely Reported Outcomes), by Study Design and Sensitivity Analyses

	Efficacy Outcomes		Harm Outcomes	
Trial Population	No. of Trials*	OR (95% CI)†	No. of Trials*	OR (95% CI)†
All trials	50	2.4 (1.4-4.0)	18	4.7 (1.8-12.0)
Parallel-group trials	38	2.8 (1.5-5.4)	17	5.3 (2.0-14.0)
Crossover trials	11	1.4 (0.63-3.2)	0	NA‡
Excluding survey nonresponders	16	1.8 (0.73-4.7)	5	3.2 (0.81-13.0)
Excluding physiologic/ pharmacokinetic trials	45	2.7 (1.6-4.8)	18	4.7 (1.8-12.0)
Fully/partially reported vs qualitatively/ unreported outcomes	35	3.1 (1.7-5.9)§	15	7.6 (2.3-25.0)§

Abbreviations: CI, confidence interval; NA, not available; OR, odds ratio.
*Trials were excluded if an odds ratio could not be calculated due to entire rows or columns being empty in the 2 × 2 table.
†Except for the bottom row, an OR >1 indicates that statistically significant outcomes ($P<.05$) have higher odds of being fully reported compared with nonsignificant outcomes.
‡Cannot be calculated.
§An OR >1 indicates that statistically significant outcomes ($P<.05$) have a higher odds of being fully or partially reported compared with nonsignificant outcomes.

Table 5. Proportion of Trials With Major Discrepancies in the Specification of Primary Outcomes When Comparing Protocols and Published Articles

Discrepancy in Published Articles Relative to Protocols	Trials With Discrepancies for ≥1 Primary Outcome, No. (%)*
Primary outcomes specified in protocols (n = 76 trials)	
Any change to protocol-defined primary outcome	40 (53)
Reported as nonprimary in published articles	26 (34)
Omitted from published articles	20 (26)
Primary outcomes specified in published articles (n = 63 trials)	
Any new primary outcome defined in published articles	21 (33)
Changed from nonprimary in protocol to primary in published articles	12 (19)
Not mentioned in protocol	11 (17)
Any discrepancy in primary outcome (n = 82 trials)†	51 (62)

*Trials often defined >1 primary outcome in protocols and published articles.
†Primary outcomes defined in either protocols or published articles.

Box 2. Examples of Major Discrepancies Between Trial Protocols and Published Articles in the Specification of Primary Outcomes*

Outcome (eg, percentage of patients with severe cerebral bleeding) changed from primary to secondary

Outcome (eg, mean pain intensity) changed from primary to unspecified

Prespecified primary outcome (eg, event-free survival rate) omitted from published reports

Outcome (eg, overall symptom score) changed from secondary to primary

Outcome (eg, percentage of patients with graft occlusion) listed as a new primary outcome (ie, not mentioned in the protocol)

*Specific details of primary outcomes omitted to maintain anonymity.

of the published articles for these trials mentioned that an amendment had been made to the study protocol. Sixty-one percent of the 51 trials with major discrepancies were funded solely by industry sources, compared with 49% of the 51 trials without discrepancies.

Among the 51 trials with major discrepancies in primary outcomes, 16 had discrepancies that favored statistically significant primary outcomes in the published articles, while 14 favored nonsignificant primary outcomes (see the "Methods" section for definition of "favored"). Eleven trials had several discrepancies that favored a mixture of significant and nonsignificant results, while for 10 trials the favored direction was unclear due to a lack of survey data about statistical results for unreported primary outcomes.

When published, 38 trials reported a power calculation, but 4 calculations were based on an outcome other than the one used in the protocol. In another 6 cases, there was a power calculation presented in a published article but not in the protocol.

COMMENT

To our knowledge, this study represents the first empirical investigation of outcome reporting bias in a representative cohort of published randomized trials. The cohort was restricted only by the geographic location of the ethics committee, although many studies involved sites in other countries. A unique feature of the study was our unrestricted access to trial protocols, which provided an unbiased a priori description of study outcomes. Protocols and published reports of systematic reviews have been compared previously,[18,19] but similar assessment of primary research has been limited to case reports,[20-22] a pilot study that required permission from researchers to access their ethics protocols,[2] and a recent study of nonindustry trials conducted by a large oncology research group.[23] Other studies have compared published articles with final reports submitted to drug approval agencies.[24,25]

Inadequate Outcome Reporting

We found that incomplete outcome reporting is common. On average, more than one third of efficacy outcomes and one half of harm outcomes in parallel-group trials were inadequately reported; the proportions were much higher in crossover studies due to unreported paired data. Even primary outcomes were often incompletely reported. Furthermore, the majority of trials had unreported outcomes, which would have been difficult to identify without access to protocols. Such poor reporting not only prevents the identification and inclusion of many essential outcomes in meta-analyses but also precludes adequate interpretation of the results in individual trials.

Our findings are likely underestimates due to underreporting of omitted outcomes by trialists, with 86% of survey responders initially denying the existence of unreported outcomes despite clear evidence to the contrary. This surprisingly high percentage suggests that contacting trialists for information about unreported outcomes is unreliable, even despite our simply worded questionnaire. We also reviewed all primary and secondary published articles for a trial; if only the primary article had been reviewed, more trial outcomes would have been classified as unreported.

The adoption of evidence-based reporting guidelines such as the revised CONSORT statement[26] for parallel-group trials should help improve poor outcome reporting. The guidelines advise reporting "for each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision."[26]

Complete Outcome Reporting and Statistically Significant Results

Statistically significant outcomes had more than a 2-fold greater odds of being fully reported compared with nonsig-

nificant outcomes. As an example, an odds ratio of 2.4 corresponds to a case in which 71% of significant outcomes are fully reported, compared with only 50% of nonsignificant outcomes. The degree of bias observed was robust in sensitivity analyses and was not associated with funding source, sample size, or number of study centers. The magnitude of outcome reporting bias is similar to that of publication bias involving entire studies, which was found to be an odds ratio of 2.54 in a meta-analysis of 5 cohort studies.[27]

It should be noted that the estimated magnitude of outcome reporting bias in each trial varied widely (Figure 3); the pooled odds ratio therefore cannot be applied to reliably predict the degree of bias for a given study.

Many trials were excluded from the analysis because odds ratios could not be meaningfully calculated due to empty rows or columns in the 2×2 table. For example, if a trial did not have any fully reported outcomes, then it was not possible to compare fully reported outcomes with incompletely reported ones. Trials were therefore more likely to be included in the analysis if they had variability in the level of reporting and/or statistical significance across outcomes, such that fewer cells were empty in the 2×2 table. Accordingly, we found that included trials had a higher number of eligible outcomes and that fewer crossover trials were included, as these often did not contain any fully reported outcomes.

Unacknowledged Changes to Primary Outcomes

The purposes of prespecifying primary outcomes are to define the most clinically relevant outcomes and to protect against "data dredging" and selective reporting.[8,28] Also, the primary outcome will generally be used for the calculation of sample size. This protective mechanism is no longer functional if predefined outcomes are subsequently changed or omitted. Despite incorporating into our analyses the few protocol amendments relating to outcomes that were submitted to the ethics committee, we found that 62% of the trials had major discrepancies for primary outcomes.

Although there is little doubt that making major changes to primary outcomes after trial commencement creates the potential for bias, the rationale behind such changes is not always clear. A preference for statistically significant results is one obvious explanation, but since few trialists provided us with the statistical significance of unreported outcomes, we could rarely ascertain whether changes to primary outcomes were made in favor of statistically significant results. Evidence of such bias was observed in one third of the trials with discrepancies in our sample, but many of the remaining trials contained discrepancies that favored a combination of significant and nonsignificant results, while others contained discrepancies that favored unclear directions or nonsignificant results alone.

A second explanation for the occurrence of discrepancies favoring nonsignificant outcomes could be that our analysis did not distinguish which treatment group was favored by the significant difference, and significant results may have been omitted if they favored the control treatment. A third explanation is that the results for other outcomes in the trial may have influenced whether statistical significance was considered important for a particular outcome. For example, an outcome may have been omitted if it was inconsistent with other trial outcomes.

It is also possible that we misclassified some changes as favoring nonsignificant primary outcomes because of our rigid cutoff of $P = .05$ used to distinguish between significant and nonsignificant results, as researchers may regard a P value of .06 as sufficiently interesting to report. In addition, nonsignificance may sometimes have been the desired result, particularly for harm outcomes or equivalence trials.

Furthermore, some of the apparent changes may be attributable to deficiencies in protocols rather than to biased actions of researchers. For example, in 4 of 10 trials in which the specification of outcomes was changed from unspecified to primary, no primary outcomes were defined in the protocol. In addition, researchers may not have realized that the protocol specifies how the data will be analyzed, and that the term "primary" should refer only to prespecified primary outcomes rather than to outcomes chosen post hoc as having the most importance or interest.

Finally, it is possible that some of the discrepancies occurred for valid reasons. After trial commencement, the omission of a predefined primary outcome can be justified if a logistical obstacle impedes its measurement or if new evidence invalidates its use as a reliable measure. However, the potential for bias still exists whenever changes are made to prespecified outcomes after trial recruitment begins. The reporting of protocol amendments in published articles must therefore be routine to enable a critical evaluation of their validity, as endorsed by the revised CONSORT statement and by other individuals.[29-31] Failure to do so has been described by one journal editor as a "breach of scientific conduct."[32] Unfortunately, none of the trial reports in our cohort acknowledged that major protocol modifications were made to primary outcomes, despite the fact that an agreement between the Danish Medical Association and the Association of the Danish Pharmaceutical Industry explicitly states that "the data analyses upon which the publication is based must be in agreement with the trial protocol, which must describe the statistical methods" (our translation).[33]

Study Limitations

The survey response rate was relatively low. The number of unreported outcomes identified would therefore be underestimated. Missing data on statistical significance also necessitated the exclusion of many outcomes from our calculation of odds ratios. However, the questionnaires constituted a secondary source of data, as we relied primarily on more objective information from protocols and published articles. Fur-

thermore, we assume that trialists would have been more likely to respond if their outcome reporting was more complete and less biased. Any response bias would thus result in conservative estimates of reporting deficiencies in our cohort.

Implications for Practice and Research

Outcome reporting bias acts in addition to the selective publication of entire studies and has widespread implications. It increases the prevalence of spurious results, and reviews of the literature will therefore tend to overestimate the effects of interventions. The worst possible situation for patients, health care professionals, and policy-makers occurs when ineffective or harmful interventions are promoted, but it is also a problem when expensive therapies, which are thought to be better than cheaper alternatives, are not truly superior.

In light of our findings, major improvements remain to be made in the reporting of outcomes in randomized trials as published. First, protocols should be made publicly available—not only to enable the identification of unreported outcomes and post hoc amendments[30,31,34] but also to deter bias. Ideally, protocols should be published online after initial trial registration and prior to trial completion. Although journals constitute one obvious modality for protocol publication, academic and funding institutions should also take responsibility in providing further venues for disseminating research information.[35]

Second, deviations from trial protocols must be described in the published articles so that readers can assess the potential for bias. Third, journal editors should not only consider routinely demanding that original protocols and any amendments be submitted with the trial manuscript but that this material should also be provided to peer reviewers and preferably be made available at the journal's Web site.[20,21,36]

Finally, trialists and journal editors should bear in mind that most individual trials may well be incorporated into subsequent reviews. Outcomes that are mentioned in published articles, but are reported with insufficient data, may not always matter when interpreting a single trial report, but they can have an important impact on meta-analyses. Unreported outcomes are even more problematic for both trials and reviews. It is therefore crucial that adequate data be reported for prespecified outcomes independent of their results. The increasing use of the Internet by journals may help to provide the space needed to accommodate such data.[36]

In summary, we found that the reporting of trial outcomes in journals is frequently inadequate to provide sufficient data for interpretation and meta-analysis, is biased to favor statistical significance, and is inconsistent with primary outcomes specified in trial protocols. These deficiencies in outcome reporting pose a threat to the reliability of the randomized trial literature.

Author Contributions: Dr Chan had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analyses.
Study concept and design; analysis and interpretation of data: Chan, Hróbjartsson, Gøtzsche, Altman.
Acquisition of data: Chan, Hróbjartsson, Haahr, Gøtzsche.
Drafting of the manuscript: Chan, Gøtzsche, Altman.
Critical revision of the manuscript for important intellectual content: Chan, Hróbjartsson, Haahr, Gøtzsche, Altman.
Statistical expertise; study supervision: Altman.
Obtained funding: Chan, Gøtzsche.
Administrative, technical, or material support: Haahr, Gøtzsche, Altman.
Funding/Support: Dr Chan was supported by the Rhodes Trust. Dr Altman is supported by Cancer Research UK.
Role of the Sponsor: Neither the Rhodes Trust nor Cancer Research UK had any role in any aspect of the study or the manuscript.
Acknowledgment: We thank the Scientific-Ethical Committees for Copenhagen and Frederiksberg, who made our study possible by providing access to the trial protocols and by offering administrative support. We also thank the researchers who responded to the questionnaires.

REFERENCES

1. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess*. 2000;4:1-115.
2. Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract*. 2002;8:353-359.
3. Bunn F, Alderson P, Hawkins V. Colloid solutions for fluid resuscitation [Cochrane Review on CD-ROM]. Oxford, England: Cochrane Library, Update Software; 2002; issue 4.
4. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med*. 2000;19:3325-3336.
5. Williamson PR, Marson AG, Tudur C, Hutton JL, Chadwick D. Individual patient data meta-analysis of randomized anti-epileptic drug monotherapy trials. *J Eval Clin Pract*. 2000;6:205-214.
6. Davey Smith G, Egger M. Meta-analysis: unresolved issues and future developments. *BMJ*. 1998;316:221-225.
7. Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst*. 1996;88:206-207.
8. Mills JL. Data torturing. *N Engl J Med*. 1993;329:1196-1199.
9. Felson DT, Anderson JJ, Meenan RF. Time for changes in the design, analysis, and reporting of rheumatoid arthritis clinical trials. *Arthritis Rheum*. 1990;33:140-149.
10. Chalmers I. Underreporting research is scientific misconduct. *JAMA*. 1990;263:1405-1408.
11. Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis [published correction appears in *Control Clin Trials*. 1989;10:356]. *Control Clin Trials*. 1989;10:31-56.
12. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med*. 1987;317:426-432.
13. West RR, Jones DA. Publication bias in statistical overview of trials: example of psychological rehabilitation following myocardial infarction [abstract]. In: Proceedings of the Second International Conference on the Scientific Basis of Health Services and Fifth International Cochrane Colloquium; October 8-12, 1997; Amsterdam, the Netherlands.
14. McCormack K, Scott NW, Grant AM. Outcome reporting bias and individual patient data meta-analysis: a case study in surgery [abstract]. In: Abstracts for Workshops and Scientific Sessions, Ninth International Cochrane Colloquium; October 9-13, 2001; Lyon, France.
15. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol*. 1992;45:885-892.
16. Whitehead A. *Meta-analysis of Controlled Clinical Trials*. Chichester, England: John Wiley & Sons Inc; 2002:216.
17. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Healthcare: Meta-analysis in Context*. 2nd ed. London, England: *BMJ* Books; 2001:285-312.
18. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy*. 2002;7:51-61.
19. Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA*. 2002;287:2831-2834.
20. Goldbeck-Wood S. Changes between protocol and manuscript should be declared at submission. *BMJ*. 2001;322:1460-1461.
21. Murray GD. Research governance must focus on research training. *BMJ*. 2001;322:1461-1462.
22. Siegel JP. Editorial review of protocols for clinical trials. *N Engl J Med*. 1990;323:1355.
23. 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;328:22-24.
24. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical in-

dustry: review of studies in new drug applications. *BMJ*. 2003;326:1171-1173.
25. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ*. 1980; 280:833-836.
26. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191-1194.
27. Dickersin K. How important is publication bias? a synthesis of available data. *AIDS Educ Prev*. 1997; 9:15-21.
28. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: statistical principles for clinical trials E9. February 1998. Available at: http://www.ich.org/MediaServer.jser?@_ID=485&@_MODE=GLB. Accessed March 15, 2004.
29. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001;134:663-694.
30. Godlee F. Publishing study protocols: making them more visible will improve registration, reporting and recruitment. *BMC News Views*. 2001;2:4.
31. Lassere M, Johnson K. The power of the protocol. *Lancet*. 2002;360:1620-1622.
32. Siegel JP. Editorial review of protocols for clinical trials. *N Engl J Med*. 1990;323:1355.
33. Agreement on cooperation on clinical trials between physicians and the pharmaceutical industry [in Danish]. Available at: http://www.dadlnet.dk. Accessed September 12, 2003.
34. Hawkey CJ. Journals should see original protocols for clinical trials. *BMJ*. 2001;323:1309.
35. Lynch CA. Institutional repositories: essential infrastructure for scholarship in the digital age. *Assoc Res Libr Bimonthly Rep*. 2003;226:1-7. Available at: http://www.arl.org/newsltr/226/ir.html. Accessed March 3, 2004.
36. Chalmers I, Altman DG. How can medical journals help prevent poor medical research? some opportunities presented by electronic publishing. *Lancet*. 1999;353:490-493.

Seize the moment of excited curiosity on any subject to solve your doubts; for if you let it pass, the desire may never return, and you may remain in ignorance.
—William Wirt (1772-1834)