

Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research

An-Wen Chan, Karmela Krleža-Jerić, Isabelle Schmid, Douglas G. Altman

‡ See related article page 750

Abstract

Background: The reporting of outcomes within published randomized trials has previously been shown to be incomplete, biased and inconsistent with study protocols. We sought to determine whether outcome reporting bias would be present in a cohort of government-funded trials subjected to rigorous peer review.

Methods: We compared protocols for randomized trials approved for funding by the Canadian Institutes of Health Research (formerly the Medical Research Council of Canada) from 1990 to 1998 with subsequent reports of the trials identified in journal publications. Characteristics of reported and unreported outcomes were recorded from the protocols and publications. Incompletely reported outcomes were defined as those with insufficient data provided in publications for inclusion in meta-analyses. An overall odds ratio measuring the association between completeness of reporting and statistical significance was calculated stratified by trial. Finally, primary outcomes specified in trial protocols were compared with those reported in publications.

Results: We identified 48 trials with 68 publications and 1402 outcomes. The median number of participants per trial was 299, and 44% of the trials were published in general medical journals. A median of 31% (10th–90th percentile range 5%–67%) of outcomes measured to assess the efficacy of an intervention (efficacy outcomes) and 59% (0%–100%) of those measured to assess the harm of an intervention (harm outcomes) per trial were incompletely reported. Statistically significant efficacy outcomes had a higher odds than nonsignificant efficacy outcomes of being fully reported (odds ratio 2.7; 95% confidence interval 1.5–5.0). Primary outcomes differed between protocols and publications for 40% of the trials.

Interpretation: Selective reporting of outcomes frequently occurs in publications of high-quality government-funded trials.

CMAJ 2004;171(7):735–40

DOI:10.1503/cmaj.1041086

Selective reporting of results from randomized trials can occur either at the level of end points within published studies (outcome reporting bias)¹ or at the level of entire trials that are selectively published (study publication bias).² Outcome reporting bias has previously been demonstrated in a broad cohort of published trials approved by a regional ethics committee.¹ The Canadian Institutes of Health Research (CIHR) — the primary federal funding agency, known before 2000 as the Medical Re-

search Council of Canada (MRC) — recognized the need to address this issue and conducted an internal review process in 2002 to evaluate the reporting of results from its funded trials. The primary objectives were to determine (a) the prevalence of incomplete outcome reporting in journal publications of randomized trials; (b) the degree of association between adequate outcome reporting and statistical significance; and (c) the consistency between primary outcomes specified in trial protocols and those specified in subsequent journal publications.

Methods

In November 2002 we identified protocols for randomized trials that were approved for funding from 1990 to 1998 by CIHR or MRC through a competitive, extensively peer-reviewed application process. A randomized trial was defined as a prospective study assessing the efficacy or harm of health care interventions and randomly allocating human participants to study groups.

We identified subsequent journal publications for each trial through a survey of principal investigators and through literature searches of PubMed, EMBASE and the Cochrane Controlled Trials Register using investigator names and keywords (final search in January 2003). We included any journal article that reported final results.

We reviewed protocols and all publications to record trial characteristics as well as the number and characteristics of reported outcomes (including statistical significance, completeness of reporting and specification as primary or secondary). An outcome was defined as a variable measured at a specific time point to assess the efficacy or harm of an intervention. Completeness of outcome reporting was defined at 4 levels based on the amount of data presented in the results section of any publications (Fig. 1). Data presented in the form of text, tables or graphs were included. A *fully reported outcome* was one with sufficient data to determine both an effect size and a measure of precision, thus enabling its inclusion in a meta-analysis (see the online appendix at www.cmaj.ca/cgi/content/full/171/7/735/DC1 for the amount of data required for meta-analyses of fully reported outcomes). *Partially reported outcomes* had some data provided in publications, while *qualitatively reported outcomes* had no useful data except for a statement regarding statistical significance or a *p* value. *Unreported outcomes* were those for which no data were provided in the publication despite the outcome being defined in either the protocol or the methods section of the publication.

Two composite levels of reporting were also defined (Fig. 1). *Reported outcomes* referred to those with any amount of data pro-

vided in the publications (fully, partially and qualitatively reported). *Incompletely reported outcomes* referred to those with inadequate data presented for meta-analysis (partially, qualitatively and unreported).

Information about unreported outcomes was solicited from trial investigators through an email survey that had been previously sent to over 500 trialists. In the initial questionnaire, we asked whether there were any outcomes that were intended for comparisons between randomized groups but were not reported in any publications. In the follow-up questionnaire, which was sent within 6 weeks after the initial survey, we provided trialists with a list of unreported outcomes based on our comparison of protocols and publications and asked for details about their statistical significance and the reasons for not reporting them.

We analyzed outcomes stratified by trial. Efficacy and harm outcomes were evaluated separately. The proportion of unreported and incompletely reported outcomes per trial was determined. For each trial, outcomes were tabulated in a 2 × 2 table according to their level of reporting (fully v. incompletely reported) and their statistical significance at the $\alpha = 0.05$ level. From each table, an odds ratio was calculated to provide a measure of outcome reporting bias (the ratio of odds for a statistically significant outcome to be fully reported compared with a nonsignificant outcome). A meaningful odds ratio could not be calculated for a given trial if there were empty rows or columns in its 2 × 2 table; for example, if a trial had no statistically significant outcomes, then we could not compare the completeness of reporting between significant and nonsignificant outcomes. However, if only a single cell or 2 diagonal cells were empty, then we added 0.5 to all cells in the table, which is standard practice for meta-analyses.^{3,4} The odds ratios from individual trials were then pooled using a random-effects meta-analysis to provide an overall estimate of bias. Sensitivity analyses were conducted to evaluate the robustness of the pooled odds ratios by excluding nonrespondents to the follow-up questionnaire, and by dichotomizing the level of reporting differently (fully or partially reported v. qualitatively reported or unreported).

The availability of protocols also enabled a comparison of primary outcomes specified a priori to those defined in publications.

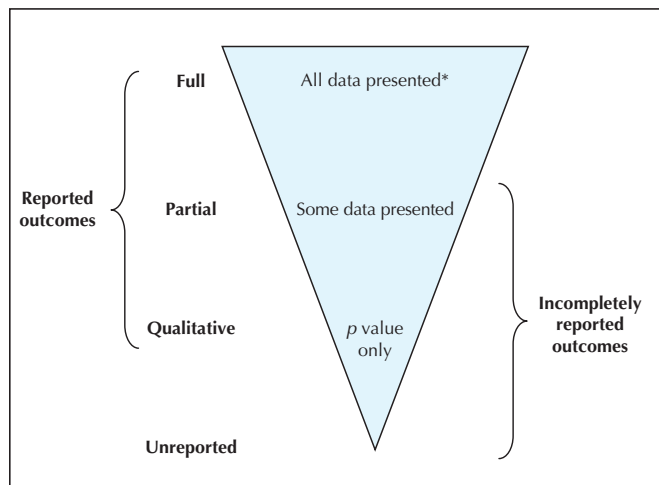


Fig. 1: Hierarchy of the levels of outcome reporting. *See the online appendix (www.cmaj.ca/cgi/content/full/171/7/735/DC1) for descriptions of the amount of data required for meta-analyses of fully reported outcomes

Primary trial outcomes were those that were specified explicitly in the text as the main or primary outcomes. If none was specified explicitly, then we considered the outcome used in the power calculation to be primary. Major discrepancies were defined to include (a) the failure to report a prespecified primary outcome; (b) reporting a prespecified primary outcome as secondary or as neither primary nor secondary in the publication; (c) the introduction of new primary outcomes in the publication; and (d) changing the outcome specified for the power calculation.

Results

We identified 141 studies approved for funding from 1990 to 1998 (Fig. 2). Files for 3 studies could not be located, and 1 trial performed no intergroup comparisons. Another 32 trials were excluded because they were not randomized trials or were approved before 1990. We asked investigators for the remaining 105 randomized trials to list any publications, and 95 (90%) responded. Fifty-seven (54%) of the 105 trials were unpublished, as confirmed by both trialists and negative literature searches ($n = 52$) or literature search alone ($n = 5$). Reasons for lack of publication were provided by 50 principal investigators: ongoing study ($n = 17$), manuscript under preparation or submitted ($n = 23$), inadequate sample size ($n = 5$), personal reasons ($n = 2$), rejected manuscript ($n = 1$), lack of statistically significant results ($n = 1$) and lack of funding ($n = 1$). The final study cohort consisted of 48 trials with 68 publications.

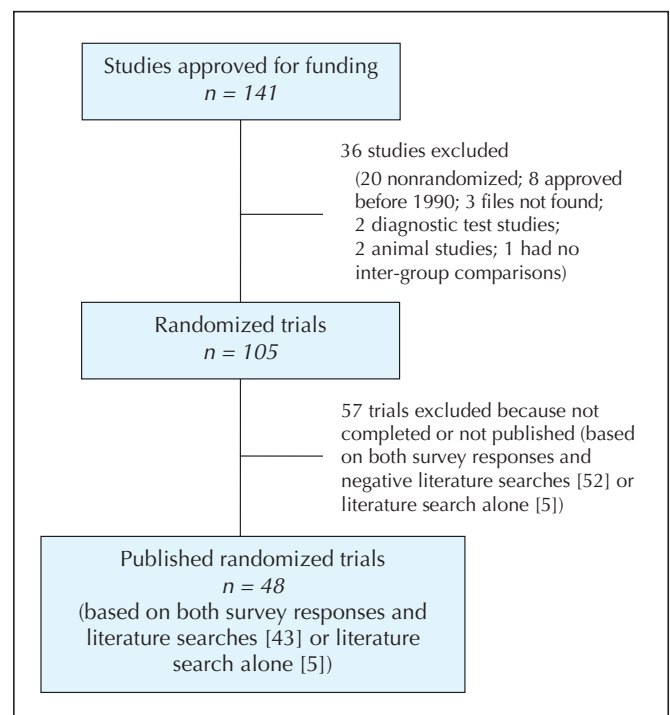


Fig. 2: Identification of published randomized trials approved for funding by the Canadian Institutes of Health Research (formerly the Medical Research Council of Canada) from 1990 to 1998.

Almost half (21 [44%]) of the trials were published in general medical journals: the *New England Journal of Medicine* (15 trials), *The Lancet* (5) and the *Journal of the American Medical Association* (1). The most common specialty fields were cardiology (10 trials), obstetrics and gynecology (8), surgery (7) and pediatrics (6). The median time from funding application to the first publication of final results was 6 years (10th–90th percentile range 4–9). The vast majority (45 [94%]) of trials were of parallel group design; the remaining 3 (6%) were of crossover design. The majority of trials (27 [56%]) examined drug interventions; the remainder examined non-drug interventions (surgical or procedural interventions, 10 [21%]; counselling or lifestyle interventions, 8 [17%]; and equipment, 3 [6%]). Most (32 [67%]) of the trials involved multiple study centres. The median sample size per trial was 299 (10th–90th percentile range 61–2568). Twenty (42%) of the trials were jointly funded by industry and CIHR/MRC; the remainder had no industry funding.

Overall, a total of 1402 outcomes were measured: 1233 efficacy outcomes in 48 trials, and 169 harm outcomes in 26 trials. A median of 26 outcomes were measured per trial (10th–90th percentile range 10–57). The median number of efficacy outcomes per trial was 20 (10th–90th percentile range 6–54); the corresponding number of harm outcomes per trial was 5 (1–11).

Prevalence of unreported outcomes

Overall, 43 (90%) of the 48 principal investigators responded to the survey with any information about unreported outcomes. Of the 35 (73%) who replied to the initial questionnaire, 28 (80%) denied the existence of unreported outcomes even though we identified such outcomes by comparing their protocols and publications. In the follow-up questionnaire, 37 (77%) of 48 investigators provided some details about the unreported outcomes. None of the respondents added any unreported outcomes to the list we provided with the questionnaire.

Forty-two (88%) of the 48 trials that measured efficacy outcomes had at least 1 unreported outcome, as compared with 16 (62%) of the 26 trials that measured harm outcomes.

For these trials, a median of 5 (10th–90th percentile range 1–16) efficacy and 2 (1–7) harm outcomes were unreported.

The most common reasons given by 29 investigators for not reporting efficacy outcomes included a lack of clinical importance (18 trials) and a lack of statistical significance (13 trials). These 2 reasons were provided by 5 of 11 survey respondents for harm outcomes.

Prevalence of incompletely reported outcomes

Incompletely reported efficacy and harm outcomes were found in 96% (46/48) and 81% (21/26) of the trials respectively. A median of 31% of efficacy outcomes per trial were incompletely reported, as compared with 59% of harm outcomes per trial (Table 1). Incompletely reported outcomes were common even when the total number of measured outcomes was low (Fig. 3). Primary outcomes were incompletely reported in 7 (16%) of 45 trials that defined such outcomes in their publications. The proportion of incompletely reported harm outcomes was lower among the trials published in general medical journals than among those published in specialty journals (Table 1).

Association between completeness of reporting and statistical significance

Eighteen trials could not contribute to the calculation of the overall odds ratio for efficacy outcomes because they had entire rows or columns that were empty in the 2×2 table. The characteristics of excluded trials were generally similar to those of included trials, except that the former had fewer outcomes with known statistical significance (median 11 [10th–90th percentile range 3–21] v. 25 [9–50]). In the analysis of harm outcomes, 22 trials were excluded for similar reasons. A total of 194 (16%) of 1233 efficacy outcomes and 24 (14%) of 169 harm outcomes were ineligible for analysis because their statistical significance was unknown; only 22 trialists provided such data in the follow-up questionnaire.

The odds ratio for outcome reporting bias in each trial is displayed in Fig. 4. The pooled odds ratio for bias across all trials was 2.7 (95% confidence interval 1.5–5.0) and 7.7 (0.5–111) for efficacy and harm outcomes respectively

Table 1: Median proportion of incompletely reported efficacy and harm outcomes* per trial by journal type

Journal type	Efficacy outcomes		Harm outcomes	
	No. of trials	Median % incompletely reported per trial (10th–90th percentile range)	No. of trials	Median % incompletely reported per trial (10th–90th percentile range)
General medical†	21	21 (4–67)	13	14 (0–71)
Specialty	27	33 (10–67)	13	67 (14–100)
All	48	31 (5–67)	26	59 (0–100)

*Outcomes measured to assess the efficacy or harm of an intervention.

†Defined as journals publishing studies from any clinical field.

(Table 2). Only 4 trials were included in the analysis of harm outcomes. We obtained similar odds ratios when we stratified the trials by journal type or excluded trials whose investigators did not provide specific details about unreported outcomes in the follow-up questionnaire (Table 2). Dichotomizing the level of reporting differently to compare fully or partially reported outcomes with qualitatively reported or unreported outcomes resulted in higher magnitudes of bias (Table 2).

Consistency between primary outcomes in protocols and publications

Nineteen (40%) of the 48 trials contained major discrepancies in the specification of primary outcomes between the protocols and the publications (Table 3). None of the publications stated that an amendment had been made to the protocol.

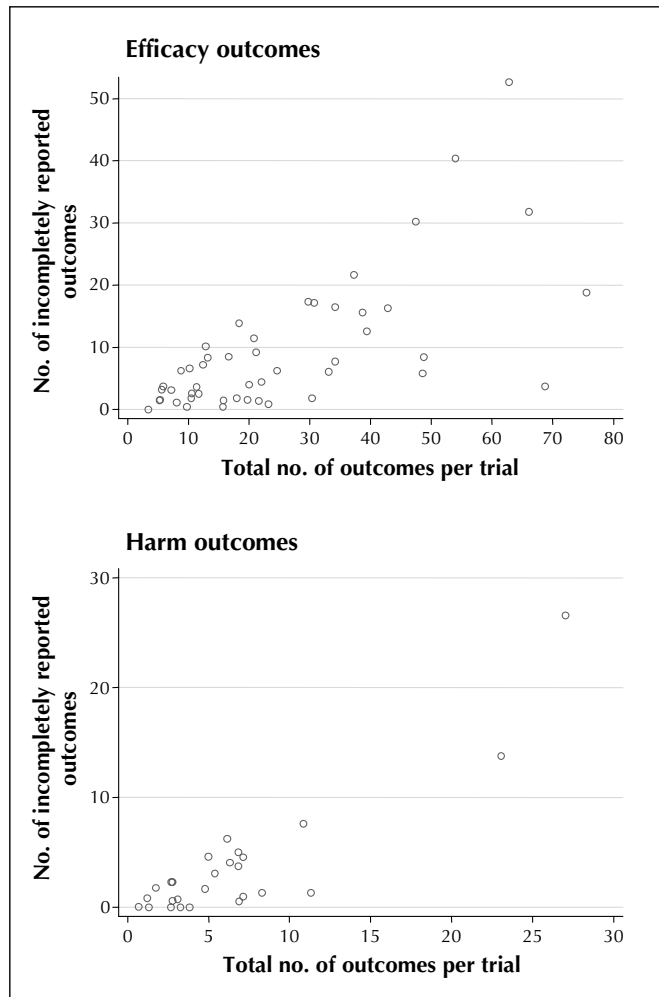


Fig. 3: Number of incompletely reported outcomes per trial, displayed according to total number of outcomes per trial, for outcomes measured to assess the efficacy of an intervention (top panel) and those measured to assess the harm of an intervention (bottom panel).

All 48 trial protocols defined primary outcomes; however, for 33% of the trials, at least 1 of these outcomes was reported as non-primary (11 trials) or was not reported in any publication (6 trials). Investigators for 3 of the 6 studies with unreported primary outcomes provided reasons for the omission: to be submitted for future publication (2 trials) and not relevant for the published article (1 trial).

For 45 trials, the primary outcomes were defined in the publications: 35 (78%) defined one, 6 (13%) defined two, and 4 (9%) defined more than two primary outcomes. For 11 trials, at least one publication-defined primary outcome had been specified as non-primary (4 trials) or had not been mentioned in the protocol (8 trials).

A discrepancy was said to favour statistically significant primary outcomes if it resulted in the reporting of signifi-

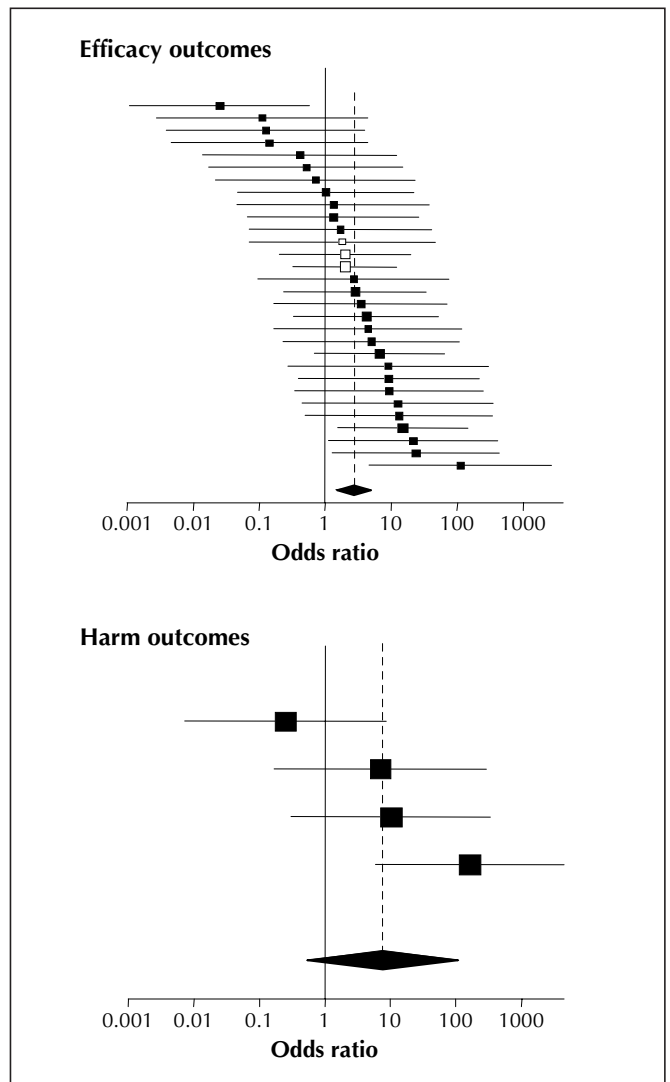


Fig. 4: Odds ratios (squares) and 95% confidence intervals for outcome reporting bias involving efficacy outcomes (30 trials, top panel) and harm outcomes (4 trials, bottom panel). The size of the square reflects the statistical weight of the trial in calculating the pooled odds ratio (represented by the diamond and dotted line).

cant primary outcomes or the omission of nonsignificant primary outcomes. Of the 19 inconsistent trials, discrepancies favoured statistically significant outcomes alone in 9 trials, nonsignificant outcomes alone in 4, a mixture of both in 5, and an unclear direction owing to a lack of information about statistical significance in 1.

Of 36 trials that reported a power calculation based on a particular outcome in their publications, 2 used an outcome that differed from the one used in the protocol, and another introduced a power calculation that had not been mentioned in the protocol.

Interpretation

Compared with recent descriptions of an “average” population of published trials (A.W.C. and D.G.A.: unpublished data), our cohort consisted of relatively large, government-funded trials whose protocols were subjected to rigorous peer review and whose publications often appeared in general medical journals. Even among trials of this quality and prominence, we identified major deficiencies in outcome reporting that were similar in magnitude to those previously observed in a broad cohort of trials approved by a regional ethics committee.¹

Other literature on outcome reporting bias is limited to case reports⁵⁻⁷ and a small pilot study that required permission from researchers to access their ethics protocols.⁸ Comparisons between journal publications and final reports submitted to drug-approval agencies have also revealed discrepancies in data reporting.^{9,10}

In our study we found that, on average, almost one-third of the efficacy outcomes and more than half of the harm outcomes per trial were inadequately reported. Even primary outcomes were incompletely reported in 16% of trials.

Our results may underestimate the degree of incomplete reporting, since trialists probably did not disclose all unreported outcomes in their survey responses. In fact, more than three-quarters of respondents initially denied the exist-

tence of unreported outcomes in trials where we had identified at least one based on our comparison of protocols and publications. We thus relied primarily on protocols and publications, rather than on survey responses, as objective data sources.

Statistically significant efficacy outcomes had more than a 2-fold greater odds of being fully reported compared with nonsignificant efficacy outcomes. An odds ratio of 2.7 corresponds to a trial in which 73% of significant outcomes are fully reported, as compared with 50% of nonsignificant outcomes. The estimate was robust or conservative in various subgroup and sensitivity analyses. Few trials were included in the analysis of harm outcomes, which precluded a precise estimate of bias. The magnitude of outcome report-

Table 3: Proportion of trials with major discrepancies in the specification of primary outcomes between protocols and publications

Discrepancy in trial publications relative to protocol	No. (%) of trials with discrepancies for ≥ 1 primary outcome*
Primary outcomes specified in protocol ($n = 48$ trials)	
Any change to protocol-defined primary outcome	16 (33)
Reported as non-primary outcome in publications	11 (23)
Omitted from publications	6 (13)
Primary outcomes specified in publications ($n = 45$ trials)	
Any new publication-defined primary outcome	11 (24)
Changed from non-primary in protocol to primary in publications	4 (9)
Not mentioned in protocol	8 (18)
Any change to primary outcomes ($n = 48$ trials)	19 (40)

*Trials often defined more than one primary outcome.

Table 2: Pooled odds ratios for outcome reporting bias by journal type and sensitivity analyses

Trial population	Efficacy outcomes		Harm outcomes	
	No. of trials*	OR† (95% CI)	No. of trials*	OR† (95% CI)
Published in general medical journal‡	13	3.2 (1.4–7.1)	2	6.6 (0.01–3818)
Published in specialty journal	17	2.3 (0.9–5.8)	2	8.6 (0.7–110)
Investigators responded to follow-up survey	27	2.8 (1.5–5.4)	3	7.8 (0.2–314)
Fully or partially reported outcomes v. qualitatively or unreported outcomes	20	5.1 (2.5–10)§	4	12.3 (1.5–99)§
All	30	2.7 (1.5–5.0)	4	7.7 (0.5–111)

Note: OR = odds ratio, CI = confidence interval.

*Trials were excluded if an OR could not be calculated owing to entire rows or columns being empty in the 2×2 table.

†Except for the row above “All,” an OR greater than 1 means that statistically significant outcomes ($p < 0.05$) have a higher odds than nonsignificant outcomes of being fully reported.

‡Defined as journals publishing studies from any clinical field.

§An OR greater than 1 means that statistically significant outcomes ($p < 0.05$) have a higher odds than nonsignificant outcomes of being fully or partially reported.

ing bias in our cohort is similar to the odds ratio of 2.54 for publication bias involving entire studies.¹¹

Because only 22 trialists provided information about the statistical significance of their unreported outcomes, many outcomes with unknown significance could not be included in the calculation of odds ratios. However, we assume that any response bias would act in a conservative direction such that we may have underestimated the impact of the deficiencies identified.

The specification of primary outcomes and analysis plans in protocols before trial initiation is intended to prevent post hoc revisions that may be data-driven. A previous review found that 4 of 98 trial publications described sample size end points that differed from reported end points.¹² Major discrepancies in primary outcomes were identified between protocols and publications in 40% of the trials in our cohort. Possible explanations for the observed discrepancies include the following: a preference for primary outcomes that demonstrate particular results; logistical barriers to measuring the original primary outcome; low event rates for binary primary outcomes; new evidence that invalidated the original primary outcome, or supported the use of a more appropriate outcome; formal amendments made to the original protocol before trial initiation that were not submitted to CIHR/MRC; and researchers' lack of awareness that retrospective revisions to prespecified outcomes and analyses can be methodologically unsound.

In the best case, the amendments were made independently of the data. In the worst case, retrospective modifications were conducted to highlight the most desirable or interesting results, while suppressing less favourable data. Unfortunately, none of the publications in our cohort stated or explained why primary outcomes had been amended.

The omission of nonsignificant findings leads to an overrepresentation of statistically significant results within published trials. We found this bias to be present even in government-funded studies, which are generally free of commercial influences and are often viewed as being more reliable than studies funded solely by industry. As a result, there would be a tendency to overestimate the effects of interventions based on data provided in the published literature. Outcome reporting bias exerts an effect that is distinct from, and is in addition to, the effect of the bias that arises from selective publication of entire studies. Further studies are needed to evaluate the impact of outcome reporting bias on overall conclusions regarding the effectiveness of interventions.

Our findings support the need for major improvements in the reporting of randomized trials. Deviations from trial protocols should be described in the published reports so that readers can assess the potential for bias.^{1,13} At a minimum, they should be declared at the time of submission to journals.^{14,15} Most importantly, protocols should be made

publicly available to deter, and enable the identification of, outcome reporting bias and unacknowledged post hoc amendments to prespecified outcomes.^{1,13,15-17}

This article has been peer reviewed.

From the Randomized Controlled Trials Unit, Canadian Institutes of Health Research, Ottawa, Ont. (Chan, Krleža-Jerić, Schmid), the University Health Network, University of Toronto, Toronto, Ont. (Chan), and Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford, UK (Altman)

Competing interests: None declared.

Contributors: An-Wen Chan is the guarantor of the study, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. An-Wen Chan also contributed to the study conception and design, the acquisition, analysis and interpretation of the data, and drafting of the article. Karmela Krleža-Jerić contributed to the study design, interpretation of the data and revising of the article. Isabelle Schmid contributed to the study design and revising of the article. Douglas Altman contributed to the study conception and design, the analysis and interpretation of the data, and drafting of the article.

References

1. Chan AW, Haahr MT, Hróbjartsson A, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
2. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases [review]. *Health Technol Assess* 2000;4(10):1-115.
3. Whitehead A. *Meta-analysis of controlled clinical trials*. Chichester (UK): Wiley; 2002. p. 216.
4. 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, editors. *Systematic reviews in healthcare. Meta-analysis in context*. 2nd ed. London (UK): BMJ Books; 2001. p. 285-312.
5. 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*. 9th International Cochrane Colloquium; Lyon; 2001. p. 34-5.
6. West RR, Jones DA. Publication bias in statistical overview of trials: example of psychological rehabilitation following myocardial infarction [abstract]. In: Proceedings of the 2nd International Conference on the Scientific Basis of Health Services and 5th Annual Cochrane Colloquium; Amsterdam; 1997. p. 82.
7. Felson DT. Bias in meta-analytic research. *J Clin Epidemiol* 1992;45:885-92.
8. 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;8:353-9.
9. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
10. Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;280:833-6.
11. Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9:15-21.
12. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol* 2003;56:717-20.
13. Godlee F. Publishing study protocols: making them more visible will improve registration, reporting and recruitment. *BMC News Views* 2001;2:4.
14. Goldbeck-Wood S. Changes between protocol and manuscript should be declared at submission. *BMJ* 2001;322:1460-1.
15. Siegel JP. Editorial review of protocols for clinical trials. *N Engl J Med* 1990; 323:1355.
16. Lasserre M, Johnson K. The power of the protocol. *Lancet* 2002;360:1620-2.
17. Hawkey CJ. Journals should see original protocols for clinical trials. *BMJ* 2001;323:1309.

Correspondence to: Dr. An-Wen Chan, Randomized Controlled Trials Unit, Canadian Institutes of Health Research, Rm. 97, 160 Elgin St., Address Locator 4809A, Ottawa ON K1A 0W9; fax 613 964-1800; anwen.chan@utoronto.ca