Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors
An-Wen Chan, Douglas G Altman

Abstract
Objective To examine the extent and nature of outcome reporting bias in a broad cohort of published randomised trials.
Design Retrospective review of publications and follow up survey of authors.
Cohort All journal articles of randomised trials indexed in PubMed whose primary publication appeared in December 2000.
Main outcome measures Prevalence of incompletely reported outcomes per trial; reasons for not reporting outcomes; association between completeness of reporting and statistical significance.
Results 519 trials with 553 publications and 10 557 outcomes were identified. Survey responders (response rate 69%) provided information on unreported outcomes but were often unreliable—for 32% of those who denied the existence of such outcomes there was evidence to the contrary in their publications. On average, over 20% of the outcomes measured in a parallel group trial were incompletely reported. Within a trial, such outcomes had a higher odds of being statistically non-significant compared with fully reported outcomes (odds ratio 2.0 (95% confidence interval 1.6 to 2.7) for efficacy outcomes; 1.9 (1.1 to 3.5) for harm outcomes). The most commonly reported reasons for omitting efficacy outcomes included space constraints, lack of clinical importance, and lack of statistical significance.
Conclusions Incomplete reporting of outcomes within published articles of randomised trials is common and is associated with statistical non-significance. The medical literature therefore represents a selective and biased subset of study outcomes, and trial protocols should be made publicly available.

Introduction
Researchers have often commented on the existence of unreported study outcomes. Even when outcomes are presented in journal publications, they may be reported with inadequate detail. Outcome reporting bias refers to the selective reporting of some results but not others in trial publications. Direct evidence of such bias has recently been shown in two cohort studies that compared trial publications with the original protocols. However, it is unknown whether selective outcome reporting can be identified when protocols are unavailable.

We used a large representative sample of publications of randomised trials indexed on PubMed to determine the prevalence of incomplete outcome reporting; the reasons for omitting outcomes; and the degree of association between completeness of reporting and statistical significance for trial outcomes.

Methods
Study selection
Using an extended version of Phase 1 from the Cochrane search strategy (see box 1), we identified primary publications of randomised trials published in December 2000 and included in PubMed by August 2002. A primary publication was the first report of final trial results. We excluded studies of cost effectiveness and diagnostic test properties, and those published in languages other than English and French.

We recorded trial characteristics from the primary publications published in December 2000. We identified extra journal publications for these trials through a survey of contact authors, as well as literature searches of PubMed, Embase, the Cochrane Controlled Trials Register, and PsychINFO using investigator names and keywords (final search January 2003). For each trial, we reviewed the primary and any subsequent publications to extract the number and characteristics of reported outcomes (including statistical significance, level of reporting, and specification as primary, secondary, or unspecified according to the text).

Reporting of outcomes
We identified unreported outcomes if they were described in the methods section but not the results section of any publication. Furthermore, using a pre-piloted questionnaire, we asked the contact authors to list any outcomes that were not reported in the published papers and to indicate for each unreported outcome whether it reached statistical significance (P<0.05) in any inter-group comparison; whether it was a primary, secondary, or unspecified outcome; and the degree of association between completeness of reporting and statistical significance for trial outcomes.

Box 1: Modified Cochrane search strategy for identifying randomised trials on PubMed that were published in December 2000

1) randomized controlled trial [pt]
2) controlled clinical trial [pt]
3) randomized controlled trials [mh]
4) random allocation [mh]
5) double blind method [mh]
6) single blind method [mh]
7) cross-over studies [mh]
8) multicenter study [pt]
9) #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10) #9 NOT (animal [mh] NOT human [mh])
11) limit 00/12/01 – 00/12/31
secondary, or unspecified outcome according to the protocol; whether it was of little, moderate, or high clinical importance; and the reasons for not reporting it. We sent two reminders via email or post to non-responders.

For each identified outcome, we recorded the level of reporting as one of four levels based on the amount of data presented in any of the journal publications (fig 1). If sufficient data were provided for inclusion in a meta-analysis, we recorded the outcome as fully reported (see box 2). We recorded an outcome as partially reported if the publications provided only some of the data necessary for meta-analysis, and as qualitatively reported if the publications presented only a P value or some indication of the presence or absence of statistical significance. Finally, unreported outcomes were those for which no data were provided in any of the publications despite being identified in the methods section or the authors’ responses to our survey.

We used two further terms to describe composite levels of reporting (fig 1). “Reported outcomes” referred to those with some data presented in any of the publications (full, partial, and qualitative). “Incompletely reported outcomes” referred to those with inadequate data for meta-analysis (partial, qualitative, and unreported).

**Statistical analyses**

Using Stata 7 (Stata Corporation, College Station, TX, USA), we conducted analyses at the trial level stratified by efficacy and harm outcomes. Primary variables of interest included the proportion of incompletely reported outcomes per trial and the reasons given by authors for not reporting outcomes. We also examined the association between the level of outcome reporting and statistical significance.

For each trial, we created a 2×2 table for the outcomes, relating the level of reporting (full or incomplete) to statistical significance at the P<0.05 level. We calculated odds ratios for each trial and pooled these using a random effects meta-analysis to provide an overall estimate of outcome reporting bias. We excluded trials if entire rows or columns were empty in the 2×2 table, as meaningful odds ratios could not be calculated. If one cell or two diagonal cells were empty we added 0.5 to all four cell frequencies in the table. We conducted sensitivity analyses by excluding trials without survey responses as well as excluding physiological and pharmacokinetic trials. We also assessed the impact of using a different cut-off point for dichotomising the level of reporting (fully or partially reported vs qualitatively reported or unreported).

We used exploratory meta-regression to evaluate the effect of 13 factors on the size of bias: trial characteristics (study design, intervention type, sample size, blinding, source of funding, number of centres; journal characteristics (specialty, general medical journal, short vs full length publication); and the reporting of important methodological details (power calculation, specification of primary outcomes, description of random sequence generation, allocation concealment, and handling of attrition). Definitions for these characteristics are described in detail elsewhere.1 We used a restricted maximum likelihood method to estimate residual heterogeneity for the univariate and multivariate backward stepwise regression analyses.

**Results**

We identified 519 trials with 553 publications and 10,557 outcomes (8325 efficacy and 2232 harm outcomes) (fig 2). Table 1 shows the main trial characteristics. Median sample sizes were 80 (10-90th centile range 25-369) and 15 (8-38) for parallel group and crossover trials respectively. A detailed report of the cohort characteristics has been published separately.7

For our survey of contact authors, 69% (356/519) of authors responded to the questionnaire with information. Among the 466 trials with identified funding sources, we obtained lower response rates for those funded solely by industry (65% (108/167)) compared with those with partial industry funding (75% (46/61)), non-industry funding (80% (147/184)), or no funding (100% (54/54)).

**Prevalence of incompletely reported outcomes**

From publications and survey responses, we identified a median of 11 (10-90th centile range 3-36) efficacy outcomes per trial (n=505) and 4 (1-17) harm outcomes per trial (n=508). Of these trials, 75% (380/505) and 64% (196/308) respectively did not fully report all their efficacy and harm outcomes in any journal publications. The median proportion of incompletely reported outcomes per trial was 42% (table 2). For harm outcomes, the median proportion per trial was 50%. Of the 232 trials (45%) that defined primary outcomes in their publications, 83 (36%) presented at least one that was incompletely reported.

---

**Box 2: Amount of data required in trial publications for outcomes to be classified as fully reported**

**Unpaired continuous data**
- Group numbers and
- Size of treatment effect (group means or medians and difference in means or medians)
- Measure of precision or variability (confidence interval, standard deviation, or standard error for means; range for medians) or the precise P value

**Unpaired binary data**
- Group numbers and
- Numbers of events or event rates in each group

**Paired continuous data**
- Mean difference between groups and a measure of its precision or exact P value or
- Raw data for each participant

**Paired binary data**
- Paired numbers of participants with and without events

**Survival data**
- Kaplan-Meier curve with numbers of patients at risk over time or
- Hazard ratio with a measure of precision
When we stratified trials by study design, we found that parallel group trials contained much lower percentages of incompletely reported efficacy and harm outcomes than crossover trials (table 2). We found little difference between specialty and general medical journals, but greater deficiencies for reporting of harm outcomes among trials that were solely funded by industry (median 56% per trial) compared with those that were not (27%) (table 2).

Prevalence of unreported outcomes
Among 356 survey responders, 281 stated that there were no unreported outcomes. However, for 32% (90/281) of these responses, we found evidence of outcomes that were mentioned in the methods section but not the results section of individual publications.

From our survey responses alone, we identified 343 unreported outcomes in 71 trials. We identified another 545 unreported outcomes based solely on discrepancies between the methods and results sections of publications for 174 trials. Finally, we identified 27 unreported outcomes based on both survey responses and publications for eight trials.

Using combined data from survey responses and publications, we identified at least one unreported efficacy outcome in 33% (169/505) of trials that measured efficacy data, and 28% (85/308) of trials with unreported harms data. A median of 2 (10-90th centile range 1-7) efficacy and 2 (1-6) harm outcomes were unreported for each of these trials.

Characteristics of unreported outcomes based on survey responses
Fifty three survey responders provided data on the clinical importance of 238 unreported efficacy outcomes (table 3). Of these trials, 26% (14/53) had unreported outcomes that were categorised as having high clinical importance. According to survey responses, the important efficacy outcomes for three of these trials were to be reported in future publications. Sixteen survey responders indicated the clinical importance of 38 unreported harm outcomes, all of which were classified as having low or moderate clinical importance (table 3); 13 authors provided the statistical significance of their unreported harm outcomes, all of which were non-significant.

Fifty four survey responders indicated the specification of their unreported efficacy outcomes as primary, secondary, or neither (table 3). These responders included the authors who provided clinical importance ratings. Primary efficacy outcomes were unreported for 13 trials; according to authors, the primary outcomes for six of these trials were to be reported in future manuscript submissions. Eighteen responders indicated the specification of their unreported harm outcomes: three trials

---

### Table 1 Overview of characteristics of 519 randomised trials published in December 2000 and cited in PubMed

<table>
<thead>
<tr>
<th>Study design</th>
<th>No (%) of trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel group</td>
<td>383 (74)</td>
</tr>
<tr>
<td>Crossover</td>
<td>116 (22)</td>
</tr>
<tr>
<td>Other*</td>
<td>20 (4)</td>
</tr>
<tr>
<td>Intervention:</td>
<td></td>
</tr>
<tr>
<td>Drug</td>
<td>993 (76)</td>
</tr>
<tr>
<td>Surgery or procedure</td>
<td>51 (10)</td>
</tr>
<tr>
<td>Counselling or lifestyle</td>
<td>55 (11)</td>
</tr>
<tr>
<td>Equipment</td>
<td>20 (4)</td>
</tr>
<tr>
<td>Study centres:</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>376 (72)</td>
</tr>
<tr>
<td>Multiple</td>
<td>134 (26)</td>
</tr>
<tr>
<td>Unclear</td>
<td>9 (2)</td>
</tr>
<tr>
<td>Funding:</td>
<td></td>
</tr>
<tr>
<td>Full industry</td>
<td>167 (32)</td>
</tr>
<tr>
<td>Partial industry</td>
<td>61 (12)</td>
</tr>
<tr>
<td>Non-industry</td>
<td>184 (35)</td>
</tr>
<tr>
<td>None</td>
<td>54 (10)</td>
</tr>
<tr>
<td>Unknown</td>
<td>53 (10)</td>
</tr>
<tr>
<td>Journal type:</td>
<td></td>
</tr>
<tr>
<td>General†</td>
<td>37 (7)</td>
</tr>
<tr>
<td>Specialty</td>
<td>482 (93)</td>
</tr>
</tbody>
</table>

*Split body (n=9), cluster (n=6), factorial (n=4), and “n of 1” (n=1) trials. †Journals publishing studies from any specialty field.

---

### Table 2 Median proportion of incompletely reported efficacy and harm outcomes per trial, stratified by trial characteristics, among 519 randomised trials published in December 2000 and cited in PubMed

<table>
<thead>
<tr>
<th>Trial characteristic</th>
<th>Efficacy outcomes</th>
<th></th>
<th></th>
<th>Harm outcomes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of trials</td>
<td>Median % of incompletely reported outcomes per trial*</td>
<td></td>
<td>No of trials</td>
<td>Median % of incompletely reported outcomes per trial*</td>
</tr>
<tr>
<td>All trials</td>
<td>505</td>
<td>42</td>
<td>22</td>
<td>508</td>
<td>50</td>
</tr>
<tr>
<td>Parallel group trials</td>
<td>375</td>
<td>22</td>
<td>25</td>
<td>237</td>
<td>25</td>
</tr>
<tr>
<td>Crossover trials</td>
<td>110</td>
<td>100</td>
<td>82</td>
<td>62</td>
<td>82</td>
</tr>
<tr>
<td>Other study designs</td>
<td>20</td>
<td>59</td>
<td>9</td>
<td>9</td>
<td>80</td>
</tr>
<tr>
<td>General medical journal</td>
<td>37</td>
<td>40</td>
<td>26</td>
<td>26</td>
<td>58</td>
</tr>
<tr>
<td>Specialty journal</td>
<td>468</td>
<td>43</td>
<td>47</td>
<td>282</td>
<td>47</td>
</tr>
<tr>
<td>Full industry funding</td>
<td>183</td>
<td>46</td>
<td>58</td>
<td>133</td>
<td>58</td>
</tr>
<tr>
<td>Partial or non-industry funding*</td>
<td>290</td>
<td>42</td>
<td>27</td>
<td>142</td>
<td>27</td>
</tr>
</tbody>
</table>

*10-90th centile ranges were 0-100% for all median percentages except for efficacy outcomes in crossover trials (48-100%). †Trials with unknown funding sources were excluded.

---

![Fig 2 Identification of randomised trials from PubMed citations of studies published in December 2000](image-url)
(17%) had at least one unreported primary harm outcome listed in the survey responses.

Table 4 shows the reasons given by authors for not reporting outcomes. For efficacy outcomes, the most common reasons were journal space restrictions (47%), lack of clinical importance (37%), and lack of statistical significance (24%). For harm outcomes, the commonest reasons were lack of clinical importance (75%) or of statistical significance (50%).

**Association between completeness of reporting and statistical significance**

Fig 3 shows the odds ratios for outcome reporting bias in each trial. Statistically significant outcomes had a higher odds of being fully reported than those that were non-significant. The pooled odds ratio for outcome reporting bias in all trials was 2.0 (95% confidence interval 1.6 to 2.7) for efficacy outcomes and 1.9 (1.1 to 3.5) for harms (table 5). Across study designs, the size of bias was similar for efficacy outcomes. We found greater variation between study designs with harm outcomes, although the subset of crossover trials contained few studies.

Sensitivity analyses showed that the overall odds ratios were not greatly affected when we excluded non-responders to our survey or physiological or pharmacokinetic trials (table 5). Dichotomising the level of outcome reporting differently (fully or partially reported v qualitatively reported or unreported) produced greater bias (table 5).

In our exploratory multivariate analysis, sample size and reporting of a power calculation were excluded post hoc based on collinearity with other variables. The final exploratory model revealed that multicentre trials were associated with significantly less bias than were single centre trials (odds ratio 0.44 (95% confidence interval 0.24 to 0.80)), while those that defined primary outcomes in their publications were associated with greater bias than those which did not specify any (1.8 (1.0 to 3.2)).

**Discussion**

We identified deficiencies in outcome reporting in a large sample of randomised trials that was unrestricted by study location or funding source. Two recent cohort studies compared study protocols with publications to show similar sizes of outcome reporting bias in 102 trials approved by a Danish ethics committee7 and 48 trials funded by a government research agency.8 Other evidence of selective outcome reporting is limited to case reports.9 10

**Contacting authors for information about outcomes**

Our survey results indicate that a response rate of almost 70% is achievable when asking authors for unreported outcomes. Overall, a fifth of responders provided us with a list of unreported trial outcomes, many of which were primary outcomes that would not have been identified based on the publications alone. Thus there is potential benefit in reviewers contacting authors for information about study outcomes.

However, we were writing to researchers whose studies were published recently, which is often not the case for systematic reviews. In addition, many authors provided responses that contradicted evidence within their publications. Response rates were
lower for industry funded trials, and it is unclear whether authors would have been willing to provide actual data.

How common are incompletely reported outcomes?
We have shown that trial outcomes are often reported inadequately for inclusion in meta-analysis. Even among primary outcomes, over a third of trials had at least one that was incompletely reported. Publications for crossover trials were particularly deficient, as they often did not provide the necessary paired outcome data. The revised CONSORT statement and its extension for harms recommend full reporting of data for all primary and secondary outcomes. Why are outcomes unreported?
Unreported outcomes were common, with over a third of trials omitting an average of two or more outcomes each. The decision to omit outcomes from publications seems to be made by investigators based on a combination of journal space restrictions, the importance of the outcome, and the statistical results. However, the omission of outcomes because of space constraints and a lack of clinical importance may well be directly associated with a lack of statistical significance. Outcomes could be deemed post hoc to have little clinical relevance if they fail to show significant findings and may thus be omitted when accommodating space limitations.

Is outcome reporting associated with statistical significance?
On average, the completeness of outcome reporting was biased to favour statistically significant outcomes. The sizes of bias for efficacy and harm outcomes (2.0 and 1.9 respectively) were robust or conservative in various sensitivity analyses, and are similar to those reported in other studies (efficacy odds ratios 2.4 and 2.7). However, we identified fewer unreported outcomes because we did not review trial protocols. Accordingly, we found 44% of trials to have one or more unreported outcomes, compared with 76% and 98% of trials in the earlier studies. We may therefore have underestimated the size of reporting deficiencies in our cohort.

Limitations of study
Response bias was expected in our study, as we relied on self-reported data from questionnaires and publications. We observed that for 32% of authors who denied the existence of unreported outcomes there was evidence to the contrary in their publications. In addition, we found lower response rates for trials funded solely by industry. This finding is consistent with previous observations that industry funded researchers may be less willing or able to offer data from their studies. Regardless of the funding source, a non-response or inaccurate response may arise from a reluctance to reveal biased practices, and we may therefore have underestimated the deficiencies in outcome reporting.

Implications for health care and research
Inadequate reporting of outcomes can have a detrimental effect on the critical interpretation of individual trial publications and reviews by limiting the availability of all existing data. Outcome reporting bias acts in addition to and in the same direction as publication bias of entire studies to produce inflated estimates of treatment effect. At its worst, the suppression of non-significant findings could lead to the use of harmful interventions. Perhaps more commonly, a treatment may be considered to be of more value than it merits, depriving patients of more effective or cheaper alternatives.

To limit outcome reporting bias, researchers and journal editors should ensure that complete data are provided for all pre-specified trial outcomes, independent of their results. Discrepancies between outcomes in the methods and results sections of publications should also be addressed during peer review. The increasing use of journal internet sites should help to alleviate concerns over space restrictions.

Given the difficulties in identifying unreported outcomes and in contacting investigators for further information, trials should be registered and protocols should be made available in the public domain before trial completion. At the least, they should be submitted with manuscripts and reviewed when being considered for publication by journals. The choice of outcomes and analysis plans would then be transparent, serving as a deterrent to selective reporting.

Contributors: AWC 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. AWC contributed to the study conception and design, acquisition of data, analysis and interpretation of data, and drafting the article. DGA contributed to the study conception and design, analysis and interpretation of data, and drafting the article. Funding: AWC was funded by the Rhodes Trust; DGA is funded by Cancer Research UK. The funding sources had no role in any aspect of the study. Competing interests: None declared. Ethical approval: None required.

Table 5
<table>
<thead>
<tr>
<th>Trial population</th>
<th>Efficacy outcomes</th>
<th>Harm outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of trials*</td>
<td>Odds ratio (95% CI)†</td>
</tr>
<tr>
<td>All trials</td>
<td>161</td>
<td>2.0 (1.6 to 2.7)</td>
</tr>
<tr>
<td>Parallel group trials</td>
<td>135</td>
<td>1.9 (1.4 to 2.6)</td>
</tr>
<tr>
<td>Crossover trials</td>
<td>22</td>
<td>2.7 (1.4 to 5.3)</td>
</tr>
<tr>
<td>Excluding survey non-responders</td>
<td>114</td>
<td>2.1 (1.6 to 2.9)</td>
</tr>
<tr>
<td>Excluding physiological or pharmaco kinetic trials</td>
<td>144</td>
<td>1.9 (1.4 to 2.5)</td>
</tr>
<tr>
<td>Fully or partially reported outcomes v qualitatively or unreported outcomes</td>
<td>134</td>
<td>3.2 (2.4 to 4.2)</td>
</tr>
</tbody>
</table>

†Except for the bottom row, odds ratio >1 signifies that fully reported outcomes have higher odds of being significant (P<0.05) than incompletely reported outcomes. In bottom row, odds ratio >1 means that fully or partially reported outcomes have higher odds of being significant (P<0.05) than qualitatively or unreported outcomes.

Papers
What is already known on this topic

Selective reporting of some measured outcomes but not others within published trials has been shown in cohorts restricted by geography and funding source.

Outcome reporting bias limits the critical interpretation of individual trials as well as the conclusions of literature reviews.

What this study adds

Outcome reporting bias exists in published trials indexed on PubMed.

Contacting authors for a list of unreported outcomes has the potential to identify important omissions from publications, although responses are often unreliable.

Clinically important trial outcomes are often inadequately reported.

Trials should be registered, and protocols should be made publicly available.

13 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommenda-
23 Godlee F, Wolfe S. Changes between protocol and manuscript should be declared at submission. *BMJ* 2003;326:1567.
(Accepted 20 December 2004)
doi 10.1136/bmj.38356.42406.8F

University Health Network, Department of Medicine, Suite RFE 3-805, 190 Elizabeth Street, Toronto, ON M5G 2C4, Canada

An-Wen Chan, resident physician

Cancer Research UK/NHS Centre for Statistics in Medicine, Oxford

Douglas G Altman, director

Correspondence to: A-W Chan anwen.chan@utoronto.ca