The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews

Jamie J Kirkham,[1] Kerry M Dwan,[1] Douglas G Altman,[2] Carrol Gamble,[1] Susanna Dodd,[1] Rebecca Smyth,[3] Paula R Williamson[1]

[1]Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool L69 3GS
[2]Centre for Statistics in Medicine, University of Oxford, Oxford OX2 6UD
[3]Population, Community and Behavioural Sciences, University ofLiverpool, Liverpool L69 3GB
Correspondence to: P R Williamson prw@liv.ac.uk
Accepted: 26 October 2009

Cite this as: *BMJ* 2010;340:c365
doi: 10.1136/bmj.c365

Outcome reporting bias in randomised controlled trials can affect the results of systematic reviews. **Jamie J Kirkham and colleagues** describe and apply a new classification system to assess the impact of selective outcome reporting bias on systematic reviews

The systematic review process has been developed to minimise biases and random errors in the evaluation of healthcare interventions.[1] Little is known about the impact of outcome reporting bias on systematic reviews. This type of bias can arise if the outcome of interest in the review had been measured and analysed in a study but not reported on the basis of the results.

In this paper we report the findings of the Outcome Reporting Bias In Trials (ORBIT) study, in which we applied a new classification system for the assessment of selective outcome reporting and evaluated the validity of the tool. We used the classification system to estimate the prevalence of outcome reporting bias and its impact on an unselected cohort of Cochrane reviews.

Methods

We examined an unselected cohort of new reviews from 50 of the 51 Cochrane collaboration review groups published in three issues of the *Cochrane Library* (Issue 4, 2006, Issue 1, 2007, and Issue 2, 2007). For each review, two investigators determined the single primary outcome, contacting the lead reviewer if this was not clear. Where the reviewer could not be contacted, two investigators independently agreed upon a single primary outcome from those listed.

SUMMARY POINTS

Empirical research indicates that statistically significant outcomes are more likely to be fully reported than non-significant results in published reports of randomised controlled trials

Little is known about the impact of outcome reporting bias in source trial reports on the conclusions of systematic reviews

Few review authors mentioned the potential problem of outcome reporting bias

Outcome reporting bias was suspected in at least one trial in more than a third of reviews

In a sensitivity analysis, nearly a fifth of statistically significant meta-analyses of the review primary outcome were affected by outcome reporting bias and a quarter would have overestimated the treatment effect by 20% or more

Assessment of systematic reviews

We first undertook an internal pilot study of 33 reviews from Issue 4, 2006 to determine the level of agreement between two investigators on the need for further assessment of outcome reporting bias. Agreement was reached on all but two reviews.

Each remaining review was then read by only one of the investigators to check whether all included trials fully reported the review primary outcome. Any uncertainties regarding the excluded studies were referred to a third reviewer.

Classification of randomised controlled trials in systematic reviews

For each review, an outcome matrix was constructed showing the reporting of the primary outcome and other outcomes in each trial included, distinguishing full, partial, or no reporting. The matrix was completed using the information in the review and revised in light of any extra information obtained from the trial reports or through contact with the trialists. Outcomes for which the data could be included in a meta-analysis were considered to be fully reported.

A classification system was developed to assess the risk of bias when a trial was excluded from a meta-analysis (table 1).

On the basis of all identified publications for a trial, one investigator and the corresponding review author independently classified any trial that did not report or partially reported results for the review primary outcome (table 1). All trials excluded from the review but selected for assessment were also classified. Any discrepancies were discussed until a final overall classification was agreed for each trial and the justification for the classification documented in full.

To assess how many reviewers had considered the possibility of outcome reporting bias, we searched the text of included reviews for the words "selective" and "reporting."

Accuracy of classification

For trials for which it was uncertain whether the review primary outcome had actually been measured and/or analysed (E, F, G, or H classification; table 1), the trialists were contacted and asked to confirm whether the review primary outcome was measured and analysed. If so, the reason for not reporting the results was requested.

Table 1 | The Outcome Reporting Bias In Trials (ORBIT) study classification system for missing or incomplete outcome reporting in reports of randomised trials

	Description	Level of reporting	Risk of bias*
Clear that the outcome was measured and analysed			
A	Trial report states that outcome was analysed but only reports that result was not significant (typically stating P>0.05)	Partial	High risk
B	Trial report states that outcome was analysed but only reports that result was significant (typically stating P<0.05)	Partial	No risk
C	Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated	Partial	Low risk
D	Trial report states that outcome was analysed but no results reported	None	High risk
Clear that the outcome was measured			
E	Clear that outcome was measured but not necessarily analysed. Judgment sa≠≠≠ys likely to have been analysed but not reported because of non-significant results	None	High risk
F	Clear that outcome was measured but not necessarily analysed. Judgment says unlikely to have been analysed but not reported because of non-significant results	None	Low risk
Unclear whether the outcome was measured			
G	Not mentioned but clinical judgment says likely to have been measured and analysed but not reported on the basis of non-significant results	None	High risk
H	Not mentioned but clinical judgment says unlikely to have been measured at all	None	Low risk
Clear that the outcome was not measured			
I	Clear that outcome was not measured	NA	No risk

*Risk of bias arising from the lack of inclusion of non-significant results when a trial was excluded from a meta-analysis or not fully reported in a review because the data were unavailable.

Two separate sensitivity and specificity analyses were performed. The first analysis considered only G and H classifications and aimed to determine how good our classification system was at judging whether the primary outcome of interest in the review had been measured when it was not mentioned in the trial report. For this analysis only, we incorporated an extra category of G classification for trials with binary outcomes where we predicted that the outcome was measured but it was not reported because there were no events.

The second analysis compared our classifications with information from the trialists to establish whether we could predict if biased reporting had occurred. Implicitly, E and G classifications suggested that bias was likely because it was either clear or assumed that the outcome had been measured and possible that non-reporting could have been influenced by the non-significance of the result. These classifications were taken to imply bias on the basis of the lack of inclusion of non-significant results. The specificity was calculated taking F and H classifications to indicate no bias.

Amount and impact of missing trial data

The amount of missing data per review was calculated, firstly based on trials that omitted data for any reason and secondly based only on those trials where data omission was suspected on the basis of the results (that is, outcome reporting bias was suspected). The maximum bias bound approach was used in a sensitivity analysis[2 3] to estimate the impact of outcome reporting bias on the review meta-analysis. See bmj.com. This method was applied only to reviews that had a single meta-analysis of the review primary outcome. The impact was assessed in terms of both the percentage change in the treatment effect estimate and the change in the statistical significance of the treatment effect estimate after adjustment.

Results

Assessments of systematic reviews

The *Cochrane Library* published 309 new reviews in Issue 4, 2006, Issue 1, 2007, and Issue 2, 2007. After exclusions, 157 reviews requiring further assessment were left; that is, 55% (157/283) of reviews did not include full data on the primary outcome of interest from all eligible trials. See bmj.com. By text searching for the words "selective" and "reporting," 20 (7%) of the 283 reviews assessed were found to have mentioned outcome reporting bias.

Full reporting of review primary outcomes in trials

Further exclusions left 2486 assessable trials, and 712 trial reports requiring a classification (545 included in reviews and 167 excluded from reviews; table 2). See bmj.com.

For 155 (6%) of the 2486 assessable trials, it was clear that the review primary outcome was measured and analysed (A, B, C, or D classification), but partial reporting meant the data could not be included in a meta-analysis. Trials classified as C were grouped according to the nature of the missing data.

A total of 359 (50%) of the 712 trials with missing data were under high suspicion for outcome reporting bias (A, D, E, or G classification; table 2). The prevalence of reviews containing at least one trial with high outcome reporting bias suspicion was 34% (96/283).

Accuracy of classification

Information on whether the outcome of interest was measured and analysed was lacking in 538 trial reports (E, F, G, or H classification). We found the email addresses of 167 (31%) authors and contacted these individuals. Responses were received from 65 authors (39%).

To determine whether the outcome of interest was measured or not, we compared our assessments against the trialists' information for 55 trials for which the outcome had not been mentioned in the trial report (G or H classification). The sensitivity for predicting that the outcome had been measured was 92% (23/25, 95% CI 81% to 100%), whereas the specificity for predicting that the outcome had not been measured was 77% (23/30, 95% CI 62% to 92%.).

To measure our judgment on whether outcome reporting bias occurred or not, we compared our assessments against the trialists' information for 62 trials for which

Table 2 | Trials assessed for outcome reporting bias (n=712)

Classification	Number of fully published trials	Number of abstracts	Total number of trials (%)
A	23	7	30 (4)
B	2	6	8 (1)
C	113	4	117 (16)
D	0	0	0 (0)
E	113	9	122 (17)
F	24	9	33 (5)
G	192	15	207 (29)
H	148	28	176 (25)
I	15	4	19 (3)
Total	**630**	**82**	**712**

the outcome was either clearly measured but not necessarily analysed (E and F classification) or had not been mentioned in the trial report (G or H classification). The sensitivity of our classification system for detecting bias was calculated to be 88% (7/8, 95% CI 65% to 100%), whereas the specificity was 80% (43/54, 95% CI 69% to 90%).

Amount and impact of missing trial data

The median amount of review primary outcome data missing from trials for any reason was 10%. For the 96 reviews that included at least one trial with a high suspicion of outcome reporting bias, the median amount of missing data was 43%.

Of the 283 reviews in our study cohort, 81 included a single meta-analysis of the review primary outcome and were included in the assessment of the impact of outcome reporting bias on the review meta-analysis.

A total of 52 of the 81 reviews included at least one trial that had a high suspicion of outcome reporting bias. In 27 of these 52 reviews, no sensitivity analysis was undertaken because classifications for all trials with missing data suggested that the review primary outcome seemed not to have been measured or it was suspected that there were no events (H and some G classifications, respectively; 17 reviews), or the reviewer or review text suggested that the missing studies would not have been combined with the other trials in the meta-analysis for reasons not related to outcome reporting bias (10 reviews). For the other 25 reviews that could be assessed, the maximum bias bound sensitivity analysis indicated that the statistically significant conclusions of eight of these reviews were not robust to outcome reporting bias; that is, the treatment effect estimate changed from a significant result favouring treatment (95% confidence interval excludes the null value) to a non-significant result.

In a further eight analyses, the result was robust to outcome reporting bias; that is, the result for the adjusted pooled estimate was also statistically significant ($P<0.05$). The remaining nine analyses had non-significant treatment effect estimates for which the application of the sensitivity analysis produced no substantial change in three analyses and a change from favouring one group to moving the effect estimate closer to the null value of no difference in treatment effect in six analyses. For all the 25 reviews assessed, the median percentage change in the treatment effect estimates after the adjustment based on the maximum bias bound was 39% (IQR 18% to 67%).

Our sensitivity analysis indicates that of the 81 reviews where there was a single meta-analysis of the review primary outcome, the significance of the results was not robust to outcome reporting bias in eight (10%) cases and the treatment effect estimate was reduced by more than 20% in 19 (23%) reviews. If only the 42 meta-analyses with a statistically significant result are considered, however, then eight (19%) become non-significant after adjustment for outcome reporting bias and 11 (26%) overestimated the treatment effect estimate by 20% or more.

Discussion

Outcome reporting bias was suspected in at least one randomised controlled trial in more than a third of the systematic reviews we examined (35%), which is substantially higher than the number of reviews in which a reference to the potential for outcome reporting bias was found (7%), thus demonstrating under-recognition of the problem. We have also shown through sensitivity analysis that outcome reporting bias affects the treatment effect estimate in a substantial proportion of Cochrane reviews.

Strengths and limitations of the study

We evaluated a large, unselected cohort of reviews, review authors were involved in the assessment of outcome reporting bias, and the authors of the trials included in the reviews were contacted for information. The textual justification for each trial classification was checked by a senior investigator.

We undertook an internal pilot study to determine the level of agreement between two researchers on the need for further assessment of a review. Given that agreement was high, a single reviewer assessed the remainder of the reviews, provided a second reviewer checked where there was uncertainty.

For the majority of trials that were missing outcome data, judgment was needed regarding the potential for outcome reporting bias. We believe we have shown that sufficiently accurate assessments are possible. This conclusion rests on the assumption that the trialists we contacted provided accurate information. A previous study suggested that trialists may be reluctant to admit selective reporting.[4] In our study, the response rate for those trialists for whom an email address was obtained was similar in trials with a high risk classification and those with a low risk classification. If response bias was operating, we would expect the sensitivity of our classifications to be underestimated (as a result of trialists with high risk classifications being less likely to respond if they have selectively reported outcomes) and the specificity overestimated (as a result of trialists with low risk classifications being more likely to respond if they have not selectively reported outcomes). With such response bias, the number of selectively reported trials in a review would be underestimated; thus the impact of outcome reporting bias on the conclusions of the reviews studied here may have been underestimated.

Our classifications of trials for outcome reporting bias facilitated an assessment of the robustness of review conclusions to such bias.[2 3] The maximum bias bound approach was the method chosen to examine this source of bias because it can be applied to any outcome type. Although only 81 (29%) of the 283 reviews studied comprised a single meta-analysis of the primary outcome of interest, there is no reason to believe the results of this assessment would not be generalisable to those reviews containing multiple meta-analyses. A limitation of our study is that it has not examined how the impact of outcome reporting bias should be assessed in reviews that do not include a meta-analysis.

bmj.com: recent Research Methods & Reporting articles

- Economic impact of disease and injury: counting what matters (2010;340:c924)
- The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews (2010;340:c365)
- Meta-analysis of individual participant data: rationale, conduct, and reporting (2010;340:c221)
- Preparing raw clinical data for publication: guidance for journal editors (2010;340:c181)
- The routine use of patient reported outcome measures in healthcare settings (2010;340:c186)

Comparison with other studies

Only one previous study used similar methods to examine the prevalence of outcome reporting bias and its impact on systematic reviews.[5] The findings were similar to the current study.

A second study of meta-analyses in Cochrane reviews demonstrated a weak positive association between the amount of outcome data missing from the source trial reports and the treatment effect estimate.[6] Our study goes further by reviewing excluded studies and classifying the likelihood of outcome reporting bias in a review on the basis of the individual trial reports.

Implications for systematic reviews

Trials should not be excluded because there is "no relevant outcome data" as the outcome data may be missing as a direct result of selective outcome reporting. Increasing the accuracy of data extraction, possibly by involving a second reviewer, could reduce the amount of missing data. Reviewers should be encouraged to contact the trialists to confirm whether the outcome was measured and analysed and obtain the results. Some review authors did not declare when a trial report stated that no events were observed in any group. We believe that reviewers should report all such data in their review.

Review authors will need to use their judgment regarding the potential for outcome reporting bias. Unfortunately, we believe there are few practical alternatives to this approach, since to do nothing is unacceptable and to contact trialists for the information or data is recommended but is not always feasible or successful. To support their judgment, reviewers should justify fully in the text of their report the classification assigned and should include verbatim quotes from the trial publication whenever possible.

The classification system that we used has been presented and applied by participants during workshops. The feedback has so far not indicated any major shortcomings of this classification system or that any additional categories are required. Adoption of the new Cochrane risk of bias tool,[7] which includes a judgment of the risk of selective outcome reporting, should also help to raise awareness of outcome reporting bias.

If a sensitivity analysis used to assess the impact of outcome reporting bias on an individual review shows that the results are not robust to outcome reporting, the review conclusions may need to be amended. Even if the results appear robust, the reviewer should still consider the potential for bias caused by unpublished studies.

Implications for trials

Recent long term initiatives could reduce the problem of outcome reporting bias in trials. For example, registration of randomised controlled trials before initiation[8] and advance publication of detailed protocols document that the trials exist and ensure their planned outcomes are specified. Reviewers can search registries to locate unpublished trials eligible to be included in a systematic review. Trialists should be encouraged to describe all changes to the outcomes stated in the protocol.

The standardisation of outcome measures in specific clinical areas, if implemented, will reduce the potential for bias.[9 10] Current recommendations state that all prespecified primary and secondary outcomes should be fully reported; any changes to the prespecified outcomes from the protocol should be explained in the final report; and the choice of outcomes included in the final report should not be based on the results.[11] International organisations also support better reporting of all trials.[12]

We hope that such strategies and raised awareness will reduce the prevalence of outcome reporting bias in clinical research.

The authors are grateful to the many Cochrane reviewers who collaborated in some way to make this research possible. Their input includes defining the review primary outcome of interest, forwarding on trial reports from their reviews, and establishing the outcome reporting bias classification for particular trials within their reviews. We thank the Cochrane Steering Group, and Cochrane statisticians and clinicians from the University of Liverpool who have provided expert knowledge when further assistance was required. We also acknowledge the trialists who answered specific queries about the reporting of outcomes in their trials. Finally, we thank Steve Taylor for undertaking some of the assessments and Chris Braithwaite for designing the study database.

Contributors: PRW, DGA, and CG designed the study protocol and developed the classification system for assessing outcome reporting bias in trials. The study case report form was designed by JJK, SD, and PRW. JJK contacted all review authors to confirm the chosen review primary outcome for use in the study. PRW and SD identified and agreed on the review primary outcome when no contact with the review authors was achieved. Assessments of outcome reporting bias (including obtaining trial reports and completing the outcome reporting bias classifications and justifications with the guidance from the systematic reviewers) were completed by JJK, SD, and KMD. All assessment justifications were checked by PRW. All the data were entered into the database by JJK. JJK and RS contacted trialists to verify if the review primary outcome was measured and/or analysed when outcome reporting bias was suspected. JJK and KMD undertook the data analysis under the supervision of PRW. JJK prepared the initial manuscript. JJK, PRW, DGA, and KMD were all involved in the substantial revision of this manuscript. All authors commented on the final manuscript before submission. PRW is the guarantor for the project.

Funding: The Outcome Reporting Bias In Trials (ORBIT) project was funded by the Medical Research Council (grant number G0500952). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Competing interests: PRW, CG, DGA, KMD, and RS are members of the Cochrane Collaboration. JJK and SD declare no competing interests.

Ethical approval: No ethics committee opinion was required for this study.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Egger M, Davey Smith G, Altman DG. *Systematic reviews in healthcare: meta-analysis in context.* 2nd ed. BMJ Books, 2001.
2 Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007;8:9.
3 Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics* 2004;60:146-53.
4 Chan A, Altman D. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
5 Williamson PR and Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547-61.
6 Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007;297:468-70.
7 Alderson P, Bunn F, Lefebvre C, Li Wan Po A, Li L, Roberts I, et al. Human albumin solution for resuscitation and volume expansion in critically ill patients. *Cochrane Database Syst Rev* 2004;4:CD001208.
8 World Health Organization. *World Health Organization international clinical trials registry platform: unique ID assignment.* World Health Organization, 2005.
9 Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med* 2008;5:96.
10 Clarke M. Standardising outcomes in paediatric clinical trials. *PLoS Med* 2008;5:e120.
11 Dwan K, Gamble C, Williamson PR, Altman DG. Reporting of clinical trials: a review of research funders' guidelines. *Trials* 2008;9:66.
12 Ghersi D, Clarke M, Berlin J, Gulmezoglu M, Kush R, Lumbiganon P, et al. Reporting the findings of clinical trials: a discussion. *Bull World Health Organ* 2008;86:492-3.