The Impact of Outcome Reporting Bias on Systematic Reviews

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

[1]Centre for Medical Statistics and Health Evaluation, University of Liverpool [2]Centre for Statistics in Medicine, University of Oxford
[3]Institute of Child Health, University of Liverpool

Email: prw@liv.ac.uk

BACKGROUND

Outcome reporting bias (ORB) has been defined as the selection for publication of a subset of the original recorded outcome variables based on the results [1]. Empirical research provides strong evidence of an association between significant results and publication in randomised controlled trials (RCTs); studies that report positive or significant results (P<0.05) are more likely to be published and outcomes that are statistically significant have higher odds of being fully reported [2]. The prevalence and impact of the problem in Cochrane reviews is unknown. Some of the reasons provided by trialists for the non-reporting of non-significant outcomes help demonstrate the existence of ORB.

"It was just uninteresting and we thought it confusing so we left it out. It didn't change [i.e. was non-significant] so it was a result that was not particularly informative let's say and was to us distracting and uninteresting."

"In the end there was nothing really in it, I mean there were no differences. If the result was statistically significant - I guess we would have had to put it in."

"Yes because what happened is, I am left with a study where everything is [non-significant], even though we walked in believing that we would see a difference, and even though we had some preliminary information you know anecdotal that there should be a difference, there was no difference. So, it really turned out to be a very negative study. So we did collect that information, and again it's a non-result, but there are only so many negative results you can put into a paper."

OBJECTIVES

- Describe and apply a classification system for the assessment of selective outcome reporting and assess its validity.
- Estimate the prevalence and impact of ORB in an unselected cohort of Cochrane reviews.
- Discuss potential solutions for reducing the prevalence of ORB in future trials.

MATERIALS AND METHODS

We examined an unselected cohort of new reviews from three issues of the Cochrane Library published between Issue 4, 2006 and Issue 2, 2007, from 50 of the 51 Cochrane Collaboration Review Groups. For each review a primary outcome was selected by the lead reviewers.

Assessment of systematic reviews

For trials not fully reporting on the review primary outcome or trials that were excluded for a reason suggestive of ORB, (e.g. "no relevant outcome data "(NROD)) were scrutinised as the relevant outcome may have been measured but not reported. Trials were classified according to whether:

- *Clear that the outcome was measured and analysed*
- *Clear that the outcome was measured but not necessarily analysed*
- *Unclear whether the outcome was measured*
- *Clear that the outcome was not measured*

For each category assignment an assessment of the risk of outcome reporting bias, in terms of high, low or no risk, was made. Lead reviewers assisted with all assessments because of the specialist clinical areas involved.

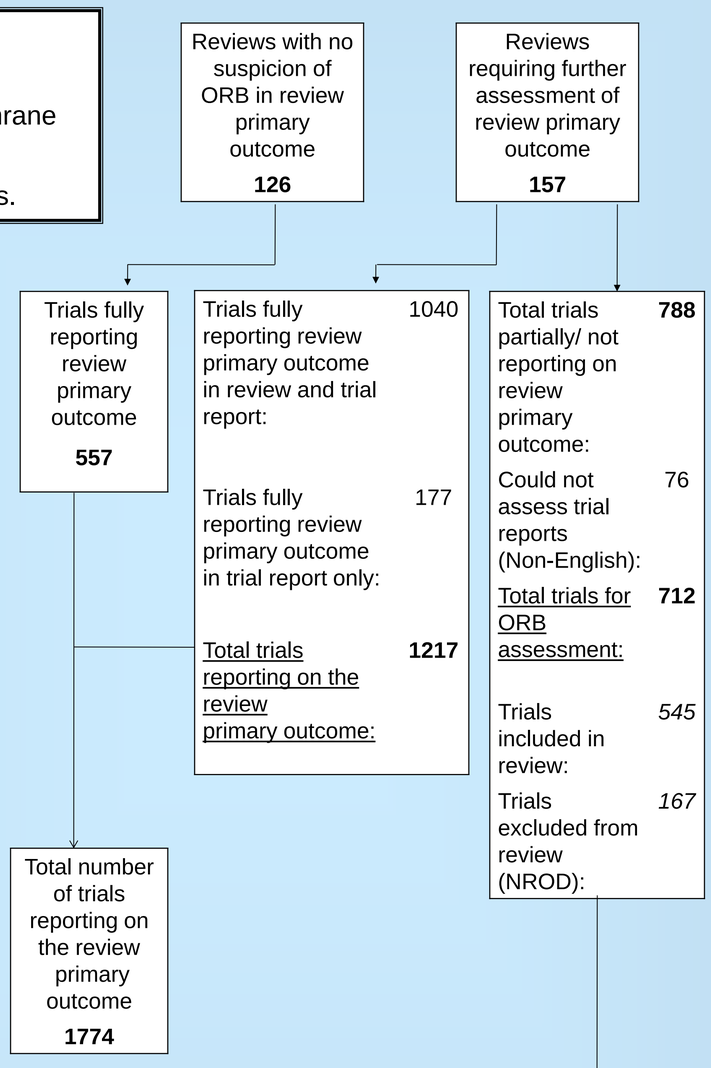
Accuracy of classification

Trialists were contacted and asked to confirm whether the review primary outcome was measured and analysed. A sensitivity/specificity was performed to firstly determine how good our classification system was at predicting whether the review primary outcome had been measured when it was not mentioned in the trial report. A second analysis considered whether we could predict whether bias had occurred.

Amount and impact of missing trial data

The amount of missing data per review was calculated when the primary outcome was not reported. The maximum bias bound approach was used in a sensitivity analysis [3,4] to estimate the impact of ORB. The method was applied to reviews where there was a single meta-analysis of the review primary outcome. The impact was assessed both in terms of the percentage change in the treatment effect estimate and the statistical significance after adjustment.

FIGURE 1: Trial assessment within reviews



Classification	Total n (%)
Clear that the outcome was measured and analysed	
A : Analysed p>0.05 (High Risk)	**30 (4)**
B: Analysed p<0.05 (Low Risk)	**8 (1)**
C: Analysed but insufficient for MA (Low Risk)	**117 (16)**
D: Analysed but no results reported (High Risk)	**0 (0)**
Clear that the outcome was measured but not necessarily analysed	
E: Measured but not necessarily analysed (High Risk)	**122 (17)**
F: Measured but not necessarily analysed (Low Risk)	**33 (5)**
Unclear whether the outcome was measured	
G: Not mentioned – LIKELY measured (High Risk)	**207 (29)**
H: Not mentioned – UNLIKELY measured (Low Risk)	**176 (25)**
Clear that the outcome was not measured	
I: Outcome NOT measured (No Risk)	**19 (3)**
TOTAL	**712**

RESULTS

Over half (157/283, 55%) the reviews did not include data for the review primary outcome from all eligible trials (Figure 1). The median amount of review primary outcome data missing for any reason was 10%, whilst 50% or more of the potential data was missing in 70 (25%) of reviews.

For 788/2562 (31%) trials, the review primary outcome was either partially reported or not reported in the review. Seventy-six trial reports could not be assessed as the articles were non-English language, leaving 2486 assessable trials and a total of 712 trial reports requiring a classification (545 included in reviews and 167 excluded from reviews due to no relevant outcome data).

In 155 (6%) of 2486 assessable trials, it was clear from the trial report that the review primary outcome was measured and analysed but results were not reported or only partially reported (Figure 1). The sensitivity for predicting that the outcome had been measured was 92% (23/25, 95% CI 81%, 100%) while the specificity for predicting that the outcome had not been measured was 77% (23/30, 95% CI 62%, 92%). For trials with no mention of the outcome in the report, judgment regarding the presence of bias was shown to have a sensitivity of 88% (95% CI 65%, 100%) and specificity of 80% (95% CI 69%, 90%). A third of Cochrane reviews (96/283, 34%) contained at least one trial with high ORB suspicion for the review primary outcome.

IMPACT

A sensitivity analysis suggested that, of 81 cases where there was a single meta-analysis of the review primary outcome, the significance of the results was not robust to ORB in 8 (10%) and reduction in the treatment effect estimate was 20% or more in 19 (23%). Considering meta-analyses with a statistically significant result only, 19% (8/42) were not robust to ORB and 26% (11/42) would have overestimated the treatment effect estimate by 20% or more.

CONCLUSIONS

ORB is an under-recognised problem in systematic reviews. To be considered a reliable source of evidence, systematic reviewers need to address explicitly the issue of missing outcome data. Extra care is required during data extraction, reviewers should identify when a trial reports that an outcome was measured or observed no events, and contact with trialists should be encouraged. Many reviewers will need to use judgment regarding the potential for outcome reporting bias and we have shown that accurate assessment is possible.

REFERENCES

[1] Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics* 2000; 49: 359-370.
[2] Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS ONE* 2008; 3(8).
[3] Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007; 8(9).
[4] Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics* 2004; 60: 146 -153.

UNIVERSITY OF LIVERPOOL

(Supported by MRC Research Grant G0500952)