Selective Outcome Reporting

Kerry Dwan[1], Jamie Kirkham[1], Carrol Gamble[1], Paula R Williamson[1], Doug G Altman[2]
SMG Training Course 2010

[1]University of Liverpool, UK, [2]University of Oxford, UK

kerry.dwan@liverpool.ac.uk

Types of selective reporting

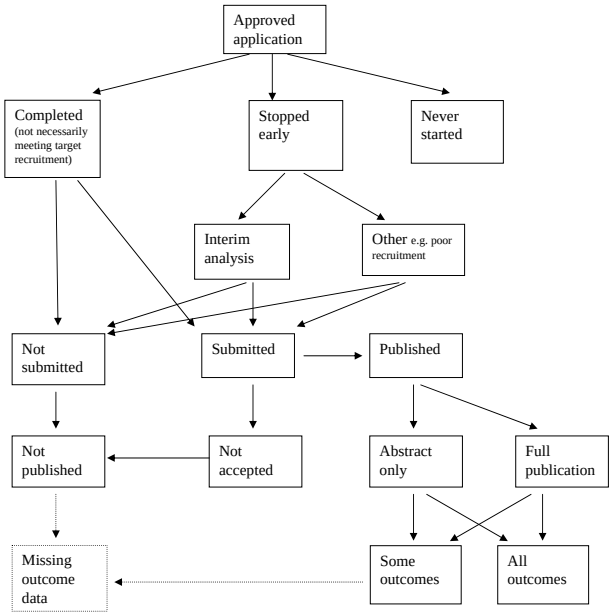
- outcomes
- subgroups
- adjusted versus unadjusted results
- prognostic or risk factors
- first period results in crossover trials
- PP rather than ITT
- complete case versus LOCF versus other methods

Outcome Reporting Bias

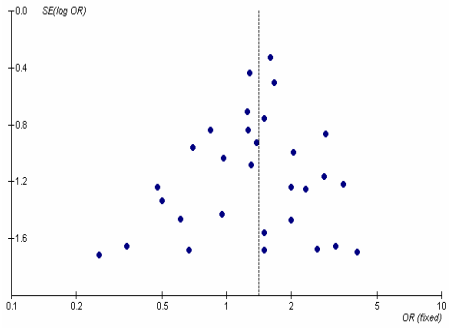
- Definition: Selection of a subset of the original recorded outcomes, on the basis of the results, for inclusion in publication
- Statistically significant outcomes more likely to be fully reported: OR 2.2 to 4.7 (Dwan et al, 2008)
- Potential threat to validity of systematic review / meta-analysis. Potentially a missing data problem if measured and analysed but not reported – similar impact to publication bias i.e. non-publication of whole studies

Types of selective outcome reporting

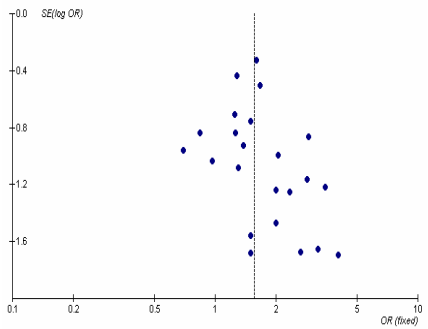
- Selective reporting of the **set** of study outcomes
  - Not all analysed outcomes are reported
- Selective reporting of a **specific** outcome
  - Hutton and Williamson (2000)
  - Selection from multiple time points
  - Subscales
  - Endpoint score versus change from baseline
  - Continuous versus binary (choice of cut-offs)
  - Different measures of same outcome, e.g. pain
- **Incomplete** reporting of a specific outcome
  - e.g. “Not significant” or “p>0.05”



Impact of ORB



OR 1.41 (1.04,1.91)



OR 1.55 (1.13,2.14)

Assessment within review

- Exclusion criteria should not include ‘did not report outcome data of interest’
- Number of eligible trials > number included in MA/ fully reported in the text

Trial ID (author, date of publication)	Review primary outcome Overall survival	Review Outcomes					Other Trial Outcomes	
		Event-free survival	Overall remission rate	Relapse rate	Toxicity and adverse events	Quality of life	Relapse site	Time to relapse
Anderson 1983	×	×	×	×	×	×	×	×
Brecher 1997	✓	✓	✓	×	✓	×	×	×
Cairo 2003a	×	O (Result to log rank test – No. events not specified	×	×	×	×	×	×
Magrath 1973	×	×	×	✓	×	×	×	✓
Magrath 1976	✓	×	×	✓	×	×	✓	✓
Neequaye 1990	✓	×	×	✓	×	×	×	✓
Olweny 1976	✓	×	✓	✓	×	×	×	×
Olweny 1977	✓	×	×	✓	✓	×	×	×
Patte 1991	✓	✓	O (some description on remission rates)	×	✓	×	×	×
Sullivan 1991	×	✓	×	×	×	×	×	×
Ziegler 1971	×	×	×	✓	×	×	×	×
Ziegler 1972a	✓	×	×	✓	×	×	O Reported for a subgroup of patients only	O K-M plot only

ORBIT classification system

- Clear that the outcome was measured and analysed

Classification	Description	Level of reporting	Risk of bias
A	States outcome analysed but only reported that result not significant (typically stating p-value >0.05)	Partial	High Risk
B	States outcome analysed but only reported that result significant (typically stating p-value <0.05).	Partial	No Risk
C	States outcome analysed but insufficient data presented to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low Risk
D	States outcome analysed but no results reported.	None	High Risk

ORBIT classification system

- Clear that the outcome was measured but not necessarily analysed

Classification	Description	Level of reporting	Risk of bias
E	Clear that outcome was measured but not necessarily analysed.	None	High Risk
F	Clear that outcome was measured but not necessarily analysed.	None	Low Risk

Examples

E : **Outcome – Overall mortality:** Trial reports on cause-specific mortality only.

F : **Ongoing study** – outcome being measured but no reason to suggest outcome analysed at current time

ORBIT classification system

- Unclear whether the outcome was measured

Classification	Description	Level of reporting	Risk of bias
G	Not mentioned but clinical judgment says likely to have been measured and analysed.	None	High Risk
H	Not mentioned but clinical judgment says unlikely to have been measured.	None	Low Risk

Examples

G : Strong belief that the PO would have been measured, e.g. Overall survival/Mortality in trials in Cancer/Aids patients

H : Follow-up appears to be too short to measure the PO, e.g. PO is live birth rate and the trial reports only on pre-birth outcomes

ORBIT classification system

- Clear the outcome was not measured

Classification	Description	Level of reporting	Risk of bias
I	Clear that outcome was not measured.	N/A	No Risk

Examples

I : **Outcome – Muscle Strength:** "No measurements of muscle strength were taken because the assessment of muscle strength with hemiparetic subjects is very difficult".

Assessment for individual study

- Review trial report
  - how likely to have been selectively not reported?
  - methods section, results section
  - incomplete reporting of outcomes
  - related outcomes reported (e.g. cause-specific and overall mortality)
  - battery of tests usually taken together (e.g. systolic and diastolic blood pressure)
  - knowledge of area suggests it is likely
- Trial protocol – search PubMed and web (www.who.int/trialsearch)
- Abstracts of presentations – mention outcomes not reported in trial report?

Example

Review: Human Albumin (2002, Issue 1)
Outcome: death for subgroup hypoalbuminaemia

- 18 (763 individuals) eligible, 16 (719 (94%)) included
- Pooled OR (95% CI): 1.51 (0.82, 2.77)
- Two trials with no data: no information in either report to indicate outcome recorded, however knowledge of clinical area suggests data would be collected routinely
- Classification (g)
- For one of the included studies, interim report (n=52) reported outcome (significant difference) whereas full report (n=94) did not.
- Original MA included preliminary data.

ORBIT: key messages

- ORB suspected in at least one trial in 34% of 283 Cochrane reviews
- 42 significant meta-analyses
  - 8 (19%) would not have remained significant
  - 11 (26%) would have overestimated the treatment effect by > 20%
- Review primary outcome less likely to be prone to ORB than other outcomes
  - under-recognition of the problem
- Interviews with trialists: 29% trials displayed ORB

The new Cochrane "Risk of Bias" tool: items to address

1. Sequence generation (randomization)
2. Allocation concealment
3. Blinding of participants, personnel and outcomes
4. Incomplete outcome data (attrition and exclusions)
5. **Selective outcome reporting**
6. Other (including topic-specific, design-specific)

Two components

- Description of what happened, possibly including ‘done’, ‘probably done’, ‘probably not done’ or ‘not done’ for some items
- Review authors’ judgement whether bias unlikely to be introduced through this item (Yes, No, Unclear)

  Yes = Low risk of bias
  No = High risk of bias

Selective outcome reporting

- Are reports of the study free of suggestion of selective outcome reporting?

Criteria for a judgment of ‘YES’ (low risk of bias)

Either:

- The study protocol is available and all of the studies’ pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way.
- The study protocol is not available but it is clear that the published reports include all of the study’s pre-specified outcomes and all expected outcomes that are of interest in the review (convincing text of this nature may be uncommon).

(rare!)

Criteria for a judgment of ‘NO’ (high risk of bias)

Any of the following:

- Not all of the study’s pre-specified primary outcomes have been reported;
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data that were not pre-specified;
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);

The protocol is need to assess the points above.

- One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis (ORBIT classifications A-D);
- The study report fails to include results for a key outcome that would be expected to have been reported for such a study (ORBIT classification G)

Criteria for a judgment of 'UNCLEAR' (uncertain risk of bias)

- Insufficient information to permit judgment of 'Yes' or 'No'.

  (It is likely that most trials will fall into this category)

'Risk of bias' assessment

Entry	Judge-ment	Description
Adequate sequence generation?	Yes	Quote: "patients were randomly allocated." Comment: Probably done, since earlier reports from the same investigators clearly describe use of random sequences (Cartwright 1980).
Allocation concealment?	No	Quote: "...using a table of random numbers." Comment: Probably not done.
… … …… … …	…	…… … …… … …
Incomplete outcome data addressed? (Short-term outcomes: 2-6 wks)	No	4 weeks: 17/110 missing from intervention group (9 due to 'lack of efficacy'); 7/113 missing from control group (2 due to 'lack of efficacy').
Incomplete outcome data addressed? (Longer-term outcomes: >6 wks)	No	12 weeks: 31/110 missing from intervention group; 18/113 missing from control group. Reasons differ across groups.
Free of selective reporting?	No	Three rating scales for cognition listed in Methods, but only one reported.
Free of other bias?	No	Trial stopped early due to apparent benefit.

Risk of bias table incorporated in RevMan 5

General approach to meta-analysis

- Undertake meta-analysis with the assumption of non-informative missing data.
- Undertake sensitivity analysis to assess robustness to assumption of informative missing data.
- Is inference robust to this? If not, consider modelling approach to assess impact under various realistic scenarios

Methods to assess ORB for RCTs

- Enumeration (Williamson and Gamble, 2005)
- Bound for maximum bias (Copas and Jackson, 2004 and Williamson and Gamble, 2007)
- Parametric selection model (Jayasekara, 2009)
- Regression (Moreno, 2009)

Copas and Jackson bound for maximum bias

- Selection model approach used to determine maximum bias when *m* trials are missing which are highly suspected of ORB. This information is used to assess the robustness to this form of bias.
- A 'worst case' sensitivity analysis
- Assumption that probability of reporting increases as standard error decreases
- The application of the Copas bias bound method initially adjusts for the known unpublished outcomes, and then the effect of various further unpublished trials can be assessed.
- C&J bound attractive due to ease of calculation

Copas and Jackson bound for maximum bias

- Calculate pooled estimated based on n-studies reporting data.
- Calculate the maximum bias bound which is a 'worst case' scenario
- Add the value of the bound to the pooled effect estimate such that the estimate moves closer to the null.
- In these situations, adjustment is conservative, hence meta-analysis found to be robust after this degree of adjustment can be considered to be robust to this form of bias
- Simulation study showed C&J adjustment works well for most cases investigated under variety of true suppression models for fixed and random effects
- In situations where treatment effect small, trial sizes are small and/or variable, number of studies with available data is small and number with missing outcome data large, C&J adjustment found to be less accurate

Moreno regression method

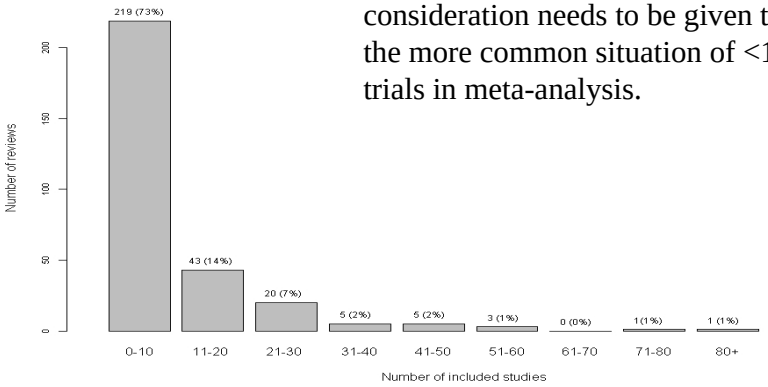
- Regression based adjustment method
  - Line of best fit
  - Predict adjusted pooled estimate for an ideal study of infinite size (se=0)
  - Quadratic version of Egger's regression
- Assumption: linear trend between effect size and variance
- Moreno approach would effectively adjust for both unpublished trials and unpublished outcomes.
- Appealing for its simplicity
- A reasonably high number of studies is needed
- Most MAs include 5-10 studies which is unlikely to be enough to give a reliable estimate using regression.

Example of the sensitivity analyses

- From Published studies only
  - 51 published studies
  - Hedge's g score SMD 0.41 (0.37, 0.45)
  - 23 (31%) including 3449 participants not published
- Sensitivity analysis results
  - Moreno: SMD 0.29 (0.23,0.35)
  - Bias bound: SMD 0.33(0.29,0.38) n=50, m=23.
  - It would take over 2000 studies to overturn the conclusion.
- Comparison to results presented to FDA
  - 74 studies
  - SMD 0.31 (0.27,0.35)

Results – number of trials per review

This is important as it means consideration needs to be given to the more common situation of <10 trials in meta-analysis.



Conclusions

- Awareness of ORB is limited but the problem must receive as much attention as between-study selection bias
- Reviewers must consider the amount of, and reasons for, data potentially missing from a meta-analysis
- To boost confidence in the review, we recommend the sensitivity of the conclusions to plausible biases should be investigated
- If robustness is lacking, present and interpret correctly both the original meta-analysis which assumes no selective reporting and the sensitivity analysis, including a description of the assumptions made regarding the nature of selection.

Solutions

Trial level

(i) Education
(ii) Core outcome sets
(iii) Better reporting - CONSORT statement, submission of protocol with manuscript (Lancet, BMJ, PLoS Med) and EQUATOR (http://www.equator-network.org/)
(iv) Reporting of legitimate outcome changes (Evans, 2007)
(v) RECs (substantial protocol amendments)
(vi) Trial and protocol registration
(vii) FDA legislation – outcome results to be made available
(viii) Need for comprehensive worldwide adoption
(ix) Funders (Guidelines)

Review level

(i) Risk of bias assessment in Cochrane reviews
(ii) Individual patient data repository (feasibility project)
(iii) Core outcome sets
(iv) Statistical methods

ORBIT: key messages

- Systematic review primary outcome data
  - missing in 25% eligible trials in Cochrane reviews
  - missing in at least one trial in 55% reviews
  - a wasted opportunity?
- Interviews with trialists about outcomes in protocol but not trial report:
  - outcomes not measured
  - outcomes measured but not analysed
  - general lack of clarity about importance and/or feasibility of data collection for outcomes chosen

Core Outcome Measures in Effectiveness Trials

Group exercise

Outcome Matrix

✓ indicates reporting in full
× indicates no reporting
o indicates partial reporting

Table 1: Bisphosphonate therapy for osteogenesis imperfect (inborn errors of metabolism)

Trial ID	Review primary outcomes		Review secondary outcomes					Trial outcomes	Review assessment of risk of bias	Our assessment of risk of bias
	Fracture reduction (as numbers and rates)	Change in bone mineral density as assessed by DEXA	Change in biochemical markers of bone and mineral metabolism (e.g., bone alkaline phosphatase measurements)	Growth (z scores; vertebral heights)	Bone pain (as assessed by self-reported questionnaires of pain and analgesic use)	Quality of life (e.g., functional changes in mobility, strength, well-being and completion of activities of daily living (ADLs))	Lung function (e.g., pulmonary function testing)			
Adami, 2003	✓	✓	o C Classification	× G classification	× H classification	× H classification	× H classification	Side effects, pQCT	None given	High risk
Chevrel, 2006	✓	✓	o	× G classification	✓	✓	× H classification	Adverse effects, calcium intake,	None given	High risk
DeMeglio, 2006	✓	✓	✓	o C classification	× H classification	× H classification	× H classification	Daily calcium intake, side effects	None given	Low risk
Gatti, 2005	✓	✓	o C classification	✓	× H classification	× H classification	× H classification	routine serum biochemistry, side effects	None given	Low risk
Glorieux, 2004 (abstract only)	o A classification	o C classification	o C classification	× G classification	o A classification	o A classification	× H classification	Side effects, cortical bone width	None given	High risk
Letocha, 2005	✓	✓	o A classification	✓	✓	✓	× H classification	None given	None given	High risk
Sakkers, 2004	✓	✓	o A classification	o A classification	× H classification	✓	× H classification	side effects	None given	High risk
Seikaly, 2005	o C classification	✓	o C classification	✓	✓	✓	× H classification	Routine serum biochemistry, side effects, stool guaiac	None given	Low risk

Feedback

Discussion points

- What should be recommended when there are concerns about ORB in a meta-analysis?
- When should reviewers consider a sensitivity analysis?
- What statistical methods should reviewers use in a sensitivity analysis?
- One-stage (publication bias generally) or two-stage (consider effect of ORB first then consider effect of unpublished studies)?
- Does the number of studies included affect this decision?

References – empirical evidence

1. Hahn S, Williamson PR and Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to an LREC. Journal of Evaluation in Clinical Practice 2002; **8.3** (August).
2. Chan A-W, Hrbjartsson A, Haahr M, Gtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to publications. JAMA 2004; 291: 2457-2465.
3. Dundar Y, Dodd S, Williamson PR, Dickson R, Walley T. (2006) Case study of the comparison of data from conference abstracts and full-text articles in health technology assessment of rapidly evolving technologies- does it make a difference? International Journal of Technology Assessment in Health Care 22(3): 288-94.
4. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171 (7): 735-740.
5. Chan AW and Altman DG. Identifying outcome reporting bias in randomised trials on Pubmed: review of publications and survey of authors. *BMJ* 2005; 330: 753-758.
6. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. (2008) Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 3(8): e3081. doi:10.1371/journal.pone.0003081

References – methodological

1. Williamson PR, Gamble C, Altman DG and JL Hutton. (2005) Outcome selection bias in meta-analysis. *Statistical Methods in Medical Research*, 14: 515-524.
2. Williamson PR and Gamble C. (2005) Identification and impact of outcome selection bias in meta-analysis. *Statistics in Medicine*, 24: 1547-1561.
3. Hutton JL and Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics* 2000*;* **49**: 359-370.
4. Hahn S, Williamson PR, Hutton JL, Garner P and Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Statistics in Medicine* 2000*;* **19**: 3325-3336.
5. Evans S. When and how can endpoints be changed after initiation of a randomized clinical trial? *PLoS Clinical Trials* 2007; e18
6. 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 Medicine* 2008;Vol. 5, No. 4, e96
7. Dwan K, Gamble C, Williamson PR, Altman DG. Reporting of clinical trials: a review of research funders' guidelines. *Trials* 2008, 9:66.
8. Kirkham JJ, Dwan K, Dodd S, Altman DG, Smyth R, Jacoby A, Gamble C, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365

References

9. Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review. *Submitted Trials 2009*
10. Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. *Biometrics* 2004; 60: 146 -153.
11. Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. *Trials* 2007; 8(9).
12. Jayasekara, L and Copas J. Modelling the outcome reporting bias in meta-analysis. *Presented at MiM meeting (June 2006).*
13. Moreno SG, Sutton AJ, Turner EH, Abrams KR, Cooper NJ, Palmer TM, Ades AE. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ, 2009;339:b2981