2014

Coping with Publication and Reporting Biases in Research Reviews

Julia H. Littell
Bryn Mawr College, jlittell@brynmawr.edu

David L. Albright

Coping with Publication and Reporting Biases in Research Reviews

Julia H. Littell
Bryn Mawr College
David L. Albright
University of Missouri-Columbia

Society for Social Work and Research
January 18, 2014

Why publication and reporting biases matter

If the literature is more likely to contain trials showing benefits of therapy while **equally valid trials showing no or negative effects remain unpublished or inaccessible**, how can reviews of the literature serve as objective guides to decision-making in clinical practice and health policy?

More technically, failure to include all valid studies results in less information, biased information, and less powerful tests.

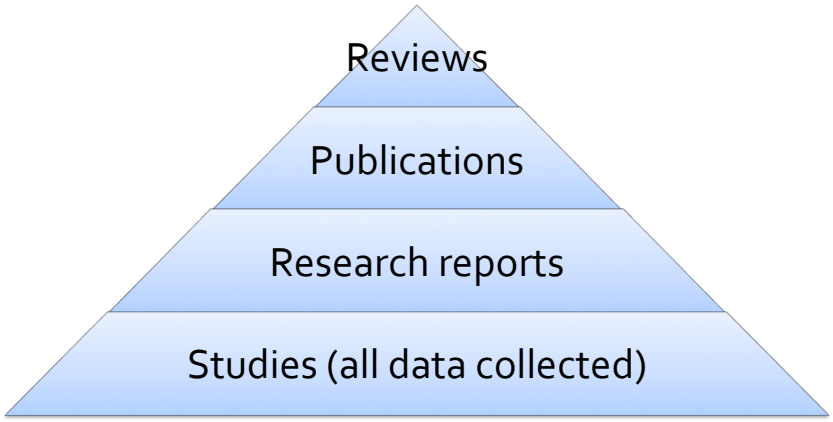
Overview

1. Empirical evidence of reporting, publication, and dissemination biases in the scholarly literature
2. Strategies for limiting these biases in the literature
   - Small group discussion
3. Methods for limiting these biases in reviews
4. Assessing and adjusting for biases in reviews
   - Group discussion

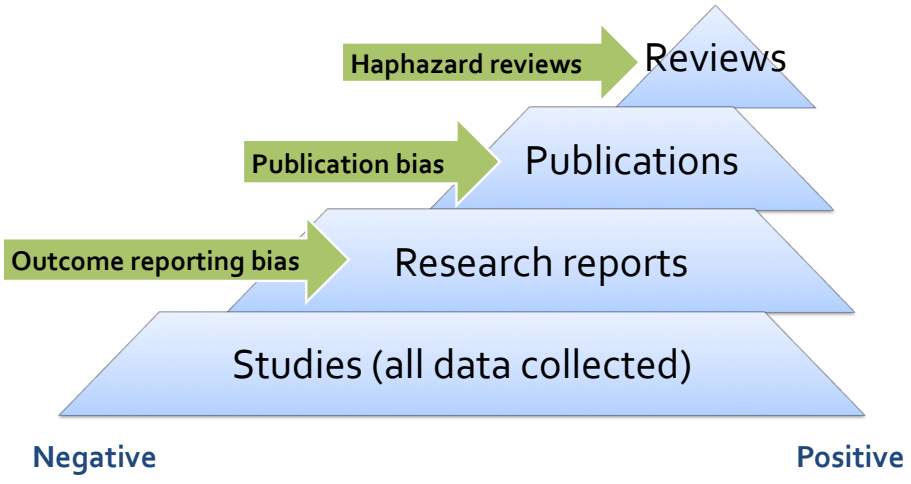
1. Empirical evidence of bias

- *Bias* is a systematic error that distorts results from the truth.
- The reporting, publication, and dissemination of research results is a biased process (Song et al., 2009, 2010).
- This presentation focuses on:
  - Outcome reporting bias
  - Publication bias
  - Dissemination biases
  - Biases that arise in research reviews (selection, inclusion, confirmation)

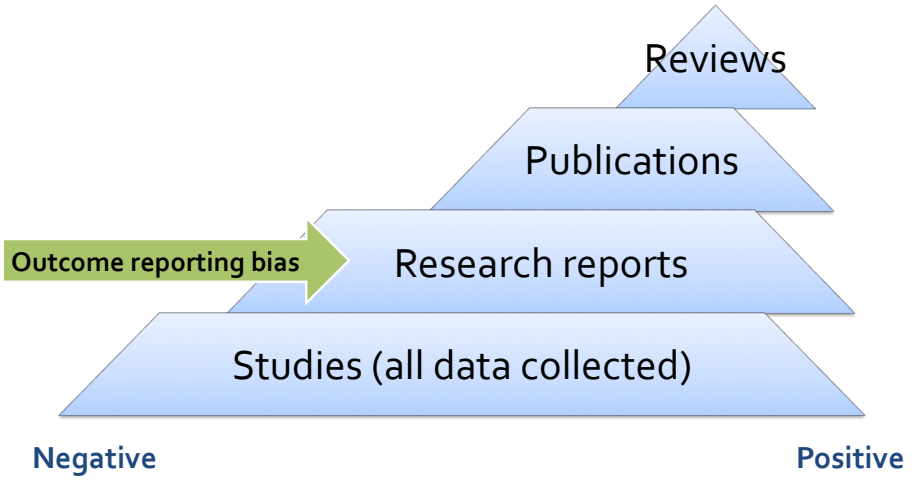
Research, reports, and reviews: Ideal



Research, reports, and reviews: Reality



Outcome reporting bias



Outcome reporting bias (ORB)

- Reporting of results is influenced by their direction and/or statistical significance
- "Cherry picking"

Evidence of ORB - 1

- Statistically significant and positive results are more likely to be
  - reported (mentioned at all)
  - fully reported (data provided)
- These reporting biases occur within studies (Chan et al., 2004a, 2004b; Chan & Altman, 2005; Dwan et al., 2008; Hahn et al., 2002; Pigott et al., 2011; Williamson et al., 2006)
- Unrelated to study or outcome "quality" (Chan et al., 2004, 2005; Pigott et al., 2011; Williamson et al., 2006)

Evidence of ORB - 2

Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias

Kerry Dwan[1]*, Douglas G. Altman[2], Juan A. Arnaiz[3], Jill Bloom[4], An-Wen Chan[5], Eugenia Cronin[6], Evelyne Decullier[7], Philippa J. Easterbrook[8], Erik Von Elm[9,10], Carrol Gamble[1], Davina Ghersi[11], John P. A. Ioannidis[12,13], John Simes[14], Paula R. Williamson[1]

- Statistically significant outcomes are more likely to be reported than nonsignificant outcomes
- Odds ratios 2.2 to 4.7 (Dwan et al., 2008)

Evidence of ORB - 3

Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists

R M D Smyth, research associate,[1,2] J J Kirkham, research associate,[1] A Jacoby, professor of medical sociology,[2] D G Altman, professor of statistics in medicine,[3] C Gamble, senior lecturer,[1] P R Williamson, professor of medical statistics[1]

- BMJ (2010)
- "The prevalence of incomplete reporting is high. Trialists seem generally unaware of the implications for the evidence base of not reporting all outcomes…"

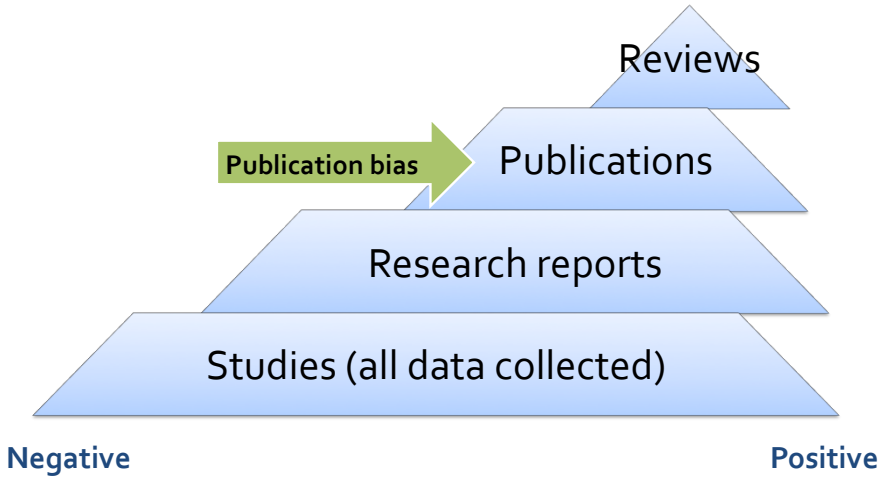
Evidence of ORB - 4

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]

- BMJ (2010)
- 19/42 (45%) of meta-analyses had substantial errors due to ORB
  - 8 (19%) became non-significant after adjusting for ORB
  - 11 (26% overestimated treatment effect by 20% or more

Publication bias



Publication rates

- 50% of completed studies are published (Dwan et al., 2008; Jones et al., 2013)
- Publication rates may be lower in social sciences, observational studies, and low/middle income countries
- 31% publication rate in psychology

Study ID	Total published (percentage)
Easterbrook, 1991 [26]	138/285 (48%)
Dickersin, 1992 [27]	390/514 (76%)
Dickersin, 1993 [3]	184/198 (93%)
Stern, 1997 [4]	[illegible]
Cooper, 1997 [32]	38/121 (status known for 117/121) (31%)
Wormald, [21]	[illegible] completed trials) (49%)
Ioannidis, 1998 [5]	36/66 (55%)
Pich, 2003 [28]	26/123 (21%)
Cronin, 2004 [31]	28/70 (40%)
Decullier, 2005 [29]	205/649 (32%) (status known for 248[1])
Decullier, 2006 [30]	48/93 (status known for 47/51 completed trials) (52%)
Hahn, 2002 [13]	18/27 (67%)
Chan, 2004a [14]	48/105 (46%)
Chan, 2004b [15]	102/274 (37%)
Ghersi, 2006 [17]	103/226 (46%)
Von Elm, 2008 [18]	233/451 (52%)

Publication status

- Publication status is not a proxy for methodological quality (McLeon & Weitz, 2004; Moyer et al., 2010)
- Should never be used as an inclusion criteria in reviews (Chandler et al., 2013; Higgins & Green, 2011; Institute of Medicine, 2011)

Evidence of publication bias

- Studies with statistically significant, positive results are 2-3 times more likely to be published than similar studies with null or negative results (Song et al., 2009, 2010)
  - likelihood of publication is related to direction and significance of results--net of influence of other variables
  - (Begg, 1994; Cooper et al., 1997; Coursol & Wagner, 1986; Dickersin, 1987, 2005; Dwan et al., 2008; Easterbrook et al., 1991; Hopewell et al., 2007, 2009; Scherer et al., 2007; Song et al., 2000, 2009, 2010; Torgerson, 2006; Vecchi et al., 2009)

Sources of publication bias

- Sources of publication bias are complex
  - Investigators
    - don't think null/negative results are worthwhile and/or don't expect these results to be accepted/published
    - are less likely to submit null results for conference presentations (Song et al., 2009) and publication (Dickersin, 2005; Song et al., 2009)
  - Peer reviewers & editors may be less likely to accept/ publish null results? (Mahoney, 1977 vs. Song et al., 2009)
- "Publication bias appears to occur early, mainly before the presentation of findings at conferences or submission of manuscripts to journals" (Song et al., 2009).

Evidence of effects of publication bias

- Publication bias appears to inflate overall effect size estimates in some meta-analyses (Lipsey & Wilson, 1993; Sutton et al., 2000)
- A recent example…

BJPsych The British Journal of Psychiatry (2010) 196, 173–178. doi: 10.1192/bjp.bp.109.066001

Review article

Efficacy of cognitive–behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias

Pim Cuijpers, Filip Smit, Ernst Bohlmeijer, Steven D. Hollon and Gerhard Andersson

Background
It is not clear whether the effects of cognitive–behavioural therapy and other psychotherapies have been overestimated because of publication bias.

Aims
To examine indicators of publication bias in randomised controlled trials of psychotherapy for adult depression.

Method
We examined effect sizes of 117 trials with 175 comparisons between psychotherapy and control conditions. As indicators of publication bias we examined funnel plots, calculated adjusted effect sizes after publication had been taken into account using Duval & Tweedie's procedure, and tested the symmetry of the funnel plots using the Begg & Mazumdar rank correlation test and Egger's test.

Results
The mean effect size was 0.67, which was reduced after adjustment for publication bias to 0.42 (51 imputed studies). Both Begg & Mazumbar's test and Egger's test were highly significant ($P<0.001$).

Conclusions
The effects of psychotherapy for adult depression seem to be overestimated considerably because of publication bias.

Declaration of interest
None.

Dissemination bias

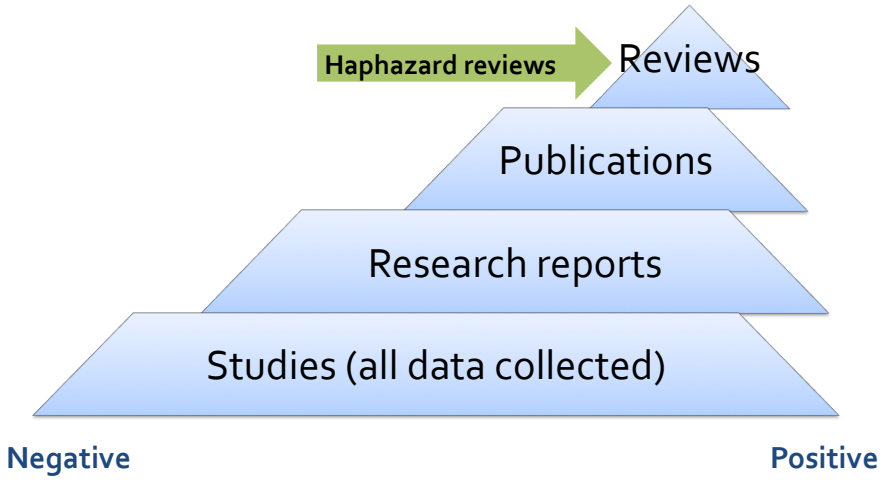
- Studies with significant results are
  - Published faster (Hopewell et al., 2001)
  - Cited and reprinted more often (Egger & Smith)
- Easier to locate (esp. in English)

Reporting, publication, dissemination

Reporting, publication, dissemination biases

- Are ubiquitous
- Are cumulative
- Inflate effect size estimates
- (Altman, 2006; Hopewell et al., 2005, 2007, 2009; Song et al., 2009)

Biases in haphazard reviews



Bias and error in the review process

- Can occur at several stages, including:
  - Searching for studies
  - Selection of studies
  - Data extraction
  - Data analysis
  - Synthesis of results across studies
- Some examples…

Searching

- Bibliographic databases
  - Largely limited to published studies
  - Search results are likely to be affected by reporting, publication, and citation biases

Selection/inclusion bias

- Trivial properties of studies or reports affect recall and evaluation of information
- Memorable titles (Bushman & Wells, 2001)

Data extraction

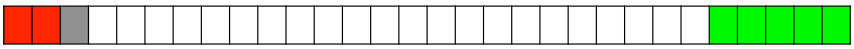
- Extracting data from studies is difficult
- Errors are common (Gøtzsche et al., 2007)
- Initial agreement is low (Tendal et al., 2009)
- Experimental evidence shows that duplicate extraction reduces errors (Buscemi et al., 2006)

Synthesis

- Narrative synthesis is
  - Unduly influenced by trivial properties of studies (Bushman & Wells, 2001)
  - Less accurate than meta-analysis (Bushman & Wells, 2001; Cooper & Rosenthal, 1980; Mann, 1994)
- Vote counting is not a good alternative
  - Does not consider sample size or heterogeneity
  - E.g., 10 studies: 6 positive, 2 null, 2 negative
    - Overall results depend on N and SE
    - Overall effect could be positive, null, or negative

Evidence of bias in narrative reviews

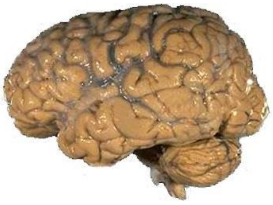
- Analysis of 14 published reviews of results of one RCT (Littell, 2008)
- Results of the RCT were mixed.
  - 30 outcomes: 2 negative, 1 missing, 22 null, 5 positive



- Most (12/14) reviewers used a single phrase to characterize results of this study

  - Highlighting advantages of one approach
  - Ignoring valuable information on relative advantages, disadvantages, and equivalent results of different approaches.

Traditional reviews and well-meaning experts can be misleading



- Scholars are human
- Rely on "natural" methods to filter and synthesize data
- The human brain is
  - Good at detecting patterns, maintaining homeostasis, defending territory
  - Bad at complex math, revising beliefs (Runciman, 2007)
- Research synthesis is too complex for informal methods, "cognitive algebra"
- Vulnerable to many sources of bias.

Bias in social work literature

- Under-investigated.
- Opportunities for bias may be greater because our research tends to use:
  - Observational designs: case reports and series, cross sectional, case-control, and cohort studies;
  - Smaller sample sizes; and
  - Larger number of tested relationships.

Summary

- Bias and error are common at every stage
  - Reporting
  - Publication
  - Dissemination
  - Reviews

2. Limiting biases in the literature

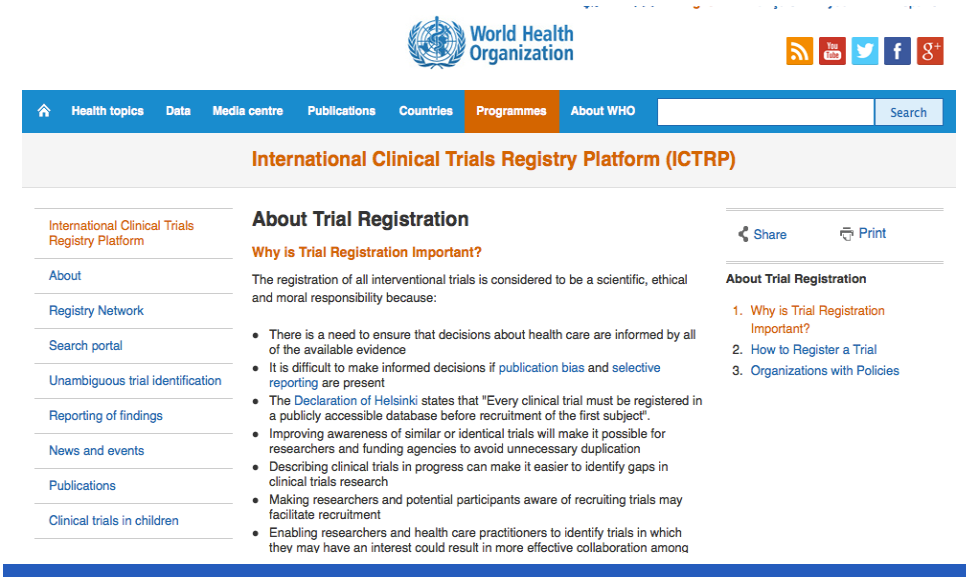
Strategies include:

1. Prospective registration of clinical intervention studies;
2. Submit null and negative results for publication;
3. Cite relevant unpublished reports; and
4. Cite null and negative results.

Prospective registration

- Prospective registration of all clinical trials required by:
  - International Committee of Medical Journal Editors; and
  - NIH: Clinicaltrials.gov
    - Remains a challenge: only 22% of trials mandated by the FDA reported results
- WHO global platform links prospective registries
  - http://www.who.int/ictrp/trial_reg/en/

WHO ICTRP



Make all results public

- Alltrials.net
  - Movement (largely in UK and EU) to require public access to all results for all trials involving humans
  - Prospective and retrospective

What can investigators do?

- Submit null and negative results
  - What makes it difficult for investigators to submit null or negative results?
    - For conference presentations?
    - For publication?
- Cite relevant unpublished reports
  - How do we find these?
- Cite relevant null and negative results
  - How can we counteract biases toward positive, significant results?

Small group discussion

- What role do you play in creating and perpetuating publication and reporting biases?
- Feasibility of strategies for limiting biases in literature?
- Other ideas?

3. Methods for limiting biases in SRs

Strategies include:

1. Comprehensive search strategies;
2. Risk of bias (ROB) assessment; and
3. Outcome reporting bias in trials (ORBIT) rubric.

Comprehensive search strategies

- Why use them?
  - Because they can reduce the likelihood of publication bias in reviews.
- Search multiple sources for individual studies including:
  - Electronic databases; and
  - Grey literature. Types include:
    - Abstracts;
    - Unpublished data;
    - Book chapters; and
    - Other.

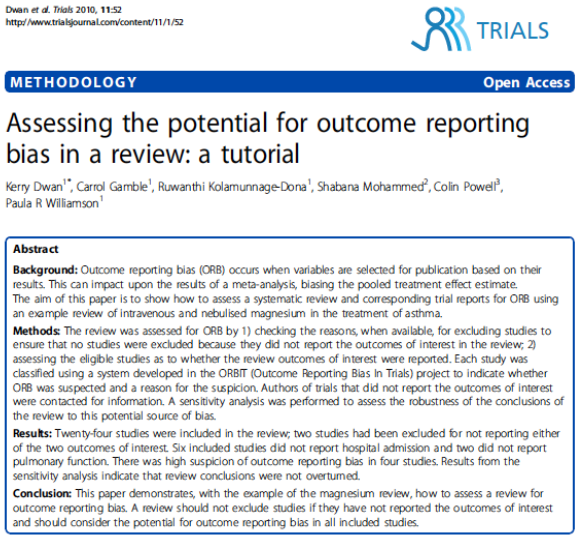
Risk of bias assessment

- Strategies include:

1. Rate risk of several types of bias for each study:
   - Selection bias;
   - Performance bias;
   - Detection bias;
   - Attrition bias; and
   - Reporting bias. (Here we focus only on reporting bias.)

2. Use moderator analysis to assess potential effects of specific biases on results

Outcome reporting bias assessment



Dwan et al. Trials 2010, 11:52
http://www.trialsjournal.com/content/11/1/52

TRIALS

METHODOLOGY Open Access

Assessing the potential for outcome reporting bias in a review: a tutorial

Kerry Dwan[1*], Carrol Gamble[1], Ruwanthi Kolamunnage-Dona[1], Shabana Mohammed[2], Colin Powell[3], Paula R Williamson[1]

Abstract

Background: Outcome reporting bias (ORB) occurs when variables are selected for publication based on their results. This can impact upon the results of a meta-analysis, biasing the pooled treatment effect estimate. The aim of this paper is to show how to assess a systematic review and corresponding trial reports for ORB using an example review of intravenous and nebulised magnesium in the treatment of asthma.

Methods: The review was assessed for ORB by 1) checking the reasons, when available, for excluding studies to ensure that no studies were excluded because they did not report the outcomes of interest in the review; 2) assessing the eligible studies as to whether the review outcomes of interest were reported. Each study was classified using a system developed in the ORBIT (Outcome Reporting Bias In Trials) project to indicate whether ORB was suspected and a reason for the suspicion. Authors of trials that did not report the outcomes of interest were contacted for information. A sensitivity analysis was performed to assess the robustness of the conclusions of the review to this potential source of bias.

Results: Twenty-four studies were included in the review; two studies had been excluded for not reporting either of the two outcomes of interest. Six included studies did not report hospital admission and two did not report pulmonary function. There was high suspicion of outcome reporting bias in four studies. Results from the sensitivity analysis indicate that review conclusions were not overturned.

Conclusion: This paper demonstrates, with the example of the magnesium review, how to assess a review for outcome reporting bias. A review should not exclude studies if they have not reported the outcomes of interest and should consider the potential for outcome reporting bias in all included studies.

ORBIT rubric

- Matrix of studies and outcomes.
- Code for reporting (for each cell):
  - Full reporting for comparisons of interest;
  - Partial reporting (e.g., *p*-value only); and
  - No reporting.
- Code suspicion of ORB:
  - High, low, or no risk.

Outcome reporting bias assessment

Table 2 The ORBIT classification system for missing or incomplete outcome reporting [10]

Classification	Description	Level of reporting	Level of suspicion of ORB
Clear that the outcome was measured and analysed			
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	Low 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
Clear that the outcome was measured			
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
Unclear that the outcome was measured			
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
Clear that the outcome was NOT measured			
I	Clear that outcome was not measured.	N/A	No risk

Considerations for ORBIT

- Need to consider multiple publications per study to understand whether outcome was measured, reported;
- Separate ORB ratings for each outcome
- ORB ratings may seem subjective.
  - Provide documentation for ratings.

4. Assessing and adjusting for bias in SRs

a. Failsafe N (or file drawer analysis)
b. Funnel plots
c. Trim and fill analysis
d. Simple statistical tests
e. Cumulative meta-analysis
f. Copas selection model
g. Contour-enhanced funnel plots

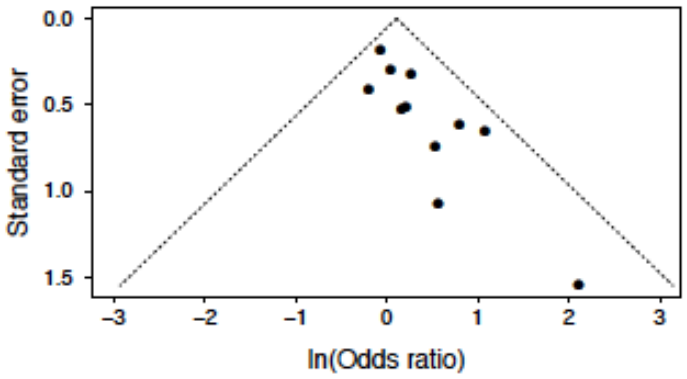
Failsafe N

- Failsafe N (Rosenthal, 1979) AKA file drawer analysis computes
  - Number of null/negative studies (of similar size) needed to overturn a significant result
- Several ways of calculating Failsafe N
- Focus on statistical not clinical significance
- All Failsafe N methods lead to widely varying estimates.
- Failsafe N should be abandoned in favor of better (more robust, reliable) methods (Becker, 2005)

Funnel plots

- *Funnel plots* are scatter plots of the treatment effects estimated from individual studies against a measure of precision (usually the SE of the ES).
- Light & Pillemer (1984)
- Plot of ES (x axis, low to high) by SE of ES (y axis, high to low)
- In absence of bias, we expect symmetry in the plot
  - Asymmetry results from a variety of sources, including non-publication of small studies with null or negative effects

Funnel plot with pseudo 95% CIs

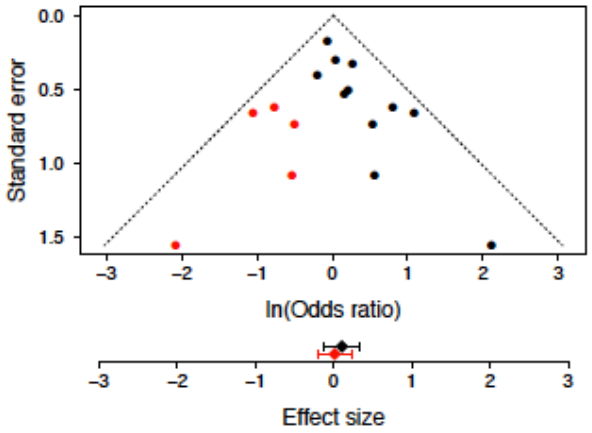


Inter-ocular analysis

- Visual assessment ("eyeballing") of funnel plots alone is unreliable
  - Is the plot symmetrical or asymmetrical? Inter-rater reliability is low.
- Shape of the plot depends on metric used in y axis
  - Use SE (Sterne & Egger, 2001)

Trim and fill analysis

Trim and fill analysis estimates missing studies and recalculates pooled ES (a form of sensitivity analysis)



(Riebler, 2008)

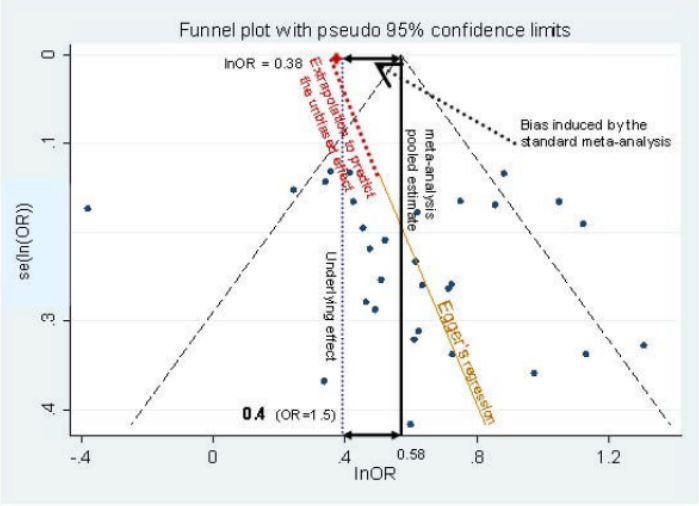
Trim & fill procedure

- Builds on the idea behind the funnel plot – that is, in the absence of bias the plot would be symmetric around the summary effect.
- The procedure imputes missing studies, adds them the analysis, and then re-computes the summary effect (Duvall & Tweedie, 2000).
- Performs poorly with substantial between-study heterogeneity and in meta-analyses with few (<10) studies
- Limitations:
  - We assume that the missing studies are the most negative.
  - Robustness of estimators with very negative effects.

Simple statistical tests

- Begg's rank correlation test, Egger's linear regression test, other regression tests
  - Quantify the bias captured by the funnel plot using the actual values of the effect sizes and their precision;
  - Have low statistical power
  - Regression methods tend to outperform trim-and-fill, but all methods deteriorate with smaller n of studies and unexplained heterogeneity (Moreno, Sutton, Ades, et al., 2009)

Funnel plot with Egger's regression test



(Moreno, Sutton, Ades, et al. 2009)

Figure 1
Regression line and standard meta-analysis on a funnel plot of simulated asymmetrical data.

Cumulative meta-analysis

- Studies sorted in forest plot in sequence by
  - Sample size (largest n to smallest n) or
  - Precision (smallest SE to largest SE)
- Cumulative meta-analysis conducted
- If ES estimate is stable after inclusion of large studies and does not change with addition of small studies, there is no evidence of publication bias
- If ES estimate changes with addition of small studies, there is evidence that bias might be present; need to investigate reasons for this (Bornstein, 2005)

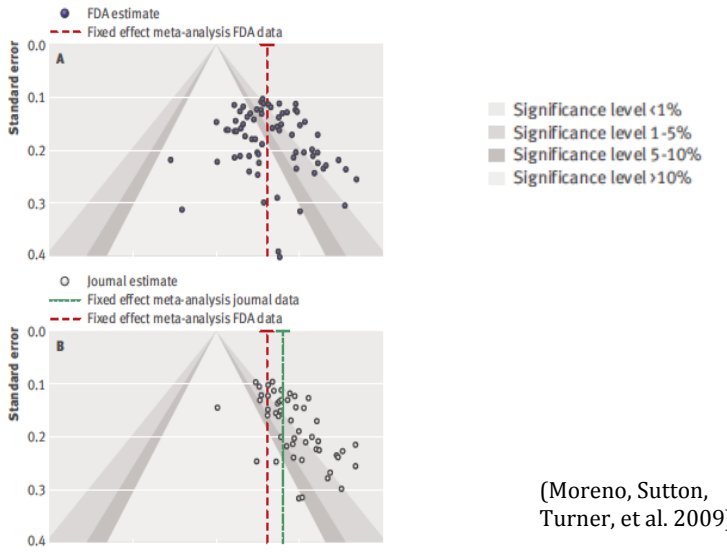
Copas selection model

- Two components, based on Heckman selection (two-stage regression) model (Copas, 1999; Copas & Shi, 2000, 2001)
  1. Random effects model for the outcome
  2. Selection model of the probability that study is observed or published
  - Correlation between these two components models the extent of selection/publication bias
- Performs better than trim & fill analysis (Schwarzer et al., 2010)
- Bayesian application and extension to network meta-analysis available (Mavridis et al., 2013)

Contour-enhanced funnel plots

- Aims to disentangle publication bias from other sources of asymmetry.
- Contours partition funnel into areas of statistical significance and non-significance
- Moreno, Sutton, Turner, et al. (2009)

Contour-enhanced funnel plots - 2



(Moreno, Sutton, Turner, et al. 2009)

Contour-enhanced funnel plots - 3

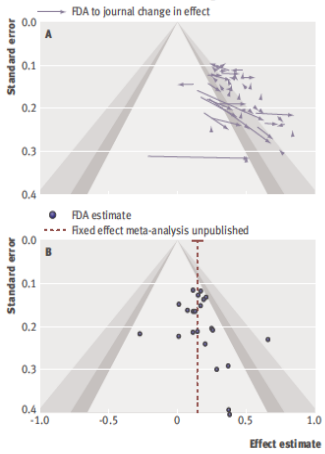


Fig 2| Contour enhanced funnel plots displaying discrepancy between Food and Drug Administration (FDA) data and journal data. (A) Arrows joining effect results from same studies when both were available from FDA and journals. (B) Estimates of effect only available from FDA (not journal published studies)

(Moreno, Sutton, Turner, et al. 2009)

Summary

1. Extensive evidence of outcome reporting, publication, and dissemination biases in the professional literature.
2. Efforts underway to limit these biases in literature – with mixed results to date
3. Methods to limit bias in reviews
   a. Comprehensive search strategies can be effective; time consuming
   b. ROB and ORBIT rubrics require judgment; understudied
4. Methods to assess and adjust for bias in reviews are under development (no consensus on best methods)

Recommended reading

- Rothstein, Sutton, & Bornstein (2005)

Evidence-based standards for reviews

- Cochrane MECIR standards (Chandler et al., 2013)
  - http://www.editorial-unit.cochrane.org/mecir
- Cochrane Handbook (Higgins & Green, 2011)
  - http://handbook.cochrane.org/
- Institute of Medicine (IOM, 2011)
  - http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-systematic-Reviews.aspx
- PRISMA (Moher et al., 2009)
  - http://www.prisma-statement.org/

Discussion

Thank you!

Contact information:

Julia H. Littell: jlittell@brynmawr.edu

David L. Albright: albrightd@missouri.edu