

2014

The Science of Research Synthesis: Limiting Bias and Error in Reviews

Julia H. Littell

Bryn Mawr College, jlittell@brynmawr.edu

Brandy R. Maynard



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).

[Let us know how access to this document benefits you.](#)

Follow this and additional works at: http://repository.brynmawr.edu/gsswsr_pubs

 Part of the [Social Work Commons](#)

Citation

Littell, Julia H. and Maynard, Brandy R., "The Science of Research Synthesis: Limiting Bias and Error in Reviews" (2014). *Graduate School of Social Work and Social Research Faculty Research and Scholarship*. Paper 59.
http://repository.brynmawr.edu/gsswsr_pubs/59

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/gsswsr_pubs/59

For more information, please contact repository@brynmawr.edu.

The Science of Research Synthesis: Limiting Bias and Error in Reviews



Julia H. Littell

Bryn Mawr College

Brandy R. Maynard

Saint Louis University

Society for Social Work and Research

January 19, 2014



Agenda

- Purposes and types of research reviews
- Empirical evidence for review methods
- Evidence- and consensus-based standards for systematic reviews and meta-analysis
- Current sources of information and global partnerships
- Potential uses of research synthesis to inform policy, practice, and research



Purposes of Reviews

- Summarize existing empirical research to:
 - Take stock of the body of research
 - Identify and address knowledge gaps
 - Organize knowledge (master the information tsunami)
 - Provide directions for further research
 - Inform policy and practice



Questions

- What do we know and how do we know it?
- Possible topics include
 - **Rates and trends** (e.g., incidence/prevalence, differences over time/place/subgroups)
 - **Correlates and causes** (e.g., risk and protective factors)
 - **Prevention and treatment** (e.g., outcomes, impacts, cost effectiveness, comparative effectiveness)
 - **Diagnosis** (e.g., accuracy of various dx categories/tests)
 - **Prognosis** (e.g., predictively validity of categories/tests)
 - **Methods and measures** (e.g., reliability, validity)



Questions and Methods

- Different review questions call for
 - Different types of evidence
 - Different synthesis methods
- Evidence hierarchies do not work across questions



Research Synthesis

Combining results of multiple studies

1. Provides more compelling evidence than results of any single study

- Single studies can have undue influence on practice & policy
- We don't use single subject (N=1) designs to assess public opinion, shouldn't rely on single studies to answer important questions

2. Provides new opportunities to investigate

- What works for whom under what conditions
- Why results may vary across studies
- Using analyses that capitalize on natural variations across studies



Building Evidence

“Science is suppose to be cumulative, but scientists only rarely accumulate evidence scientifically” (Chalmers, Hedges, & Cooper, 2002, p. 12)

Scientific methods of research synthesis are

- Available
- Rapidly advancing



The Problem: Studies Pile Up



- **“What can you build with thousands of bricks?” (Lipsey, 1997)**
- Many studies are conducted on the same topic
- Which one(s) do we use? How do we use them?



Types of Research Synthesis

- Reviews vary in amount of planning, transparency, and rigor
- Different approaches to research synthesis:
 - Traditional, narrative reviews (may include “vote counting”)
 - Systematic reviews (aim to minimize bias)
 - Some “systematic reviews” aren’t
 - Meta-analysis (quantitative synthesis)
 - Rapid evidence assessment (AKA rapid reviews) - hybrids

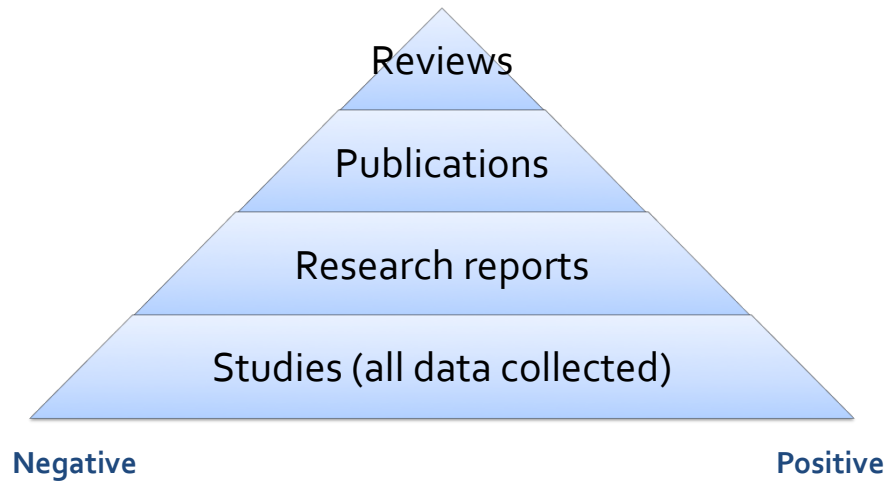


Traditional Narrative Reviews

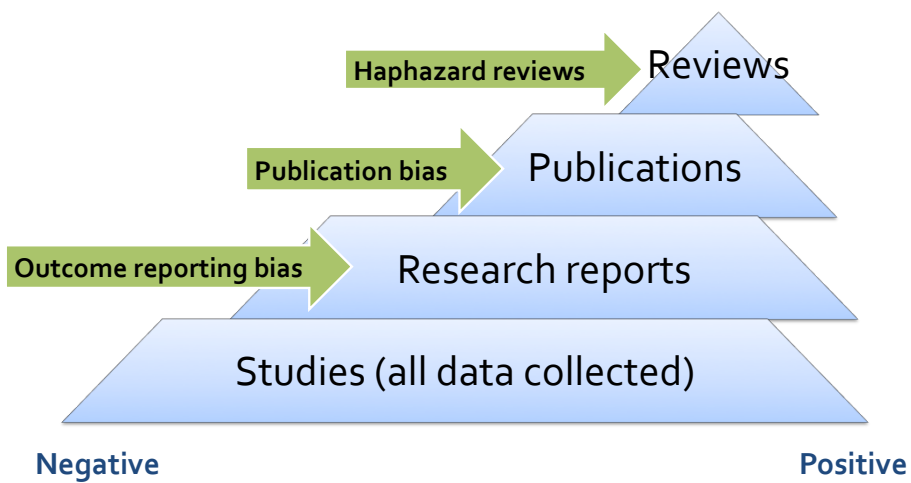
- Convenience samples of published studies
- Narrative description of studies
- Cognitive algebra or “vote counting” to synthesize results
 - How many studies had positive, null, negative, or mixed results
 - Relies on statistical significance in primary studies
 - Counting the wrong thing. Significance depends, in part, on sample size (studies may be too small to detect meaningful effects, large studies can detect differences that are meaningless)
- Decision rules are not transparent
- Vulnerable to many sources of bias and error



Research, Reports, and Reviews: Ideal



Research, Reports, and Reviews: Reality



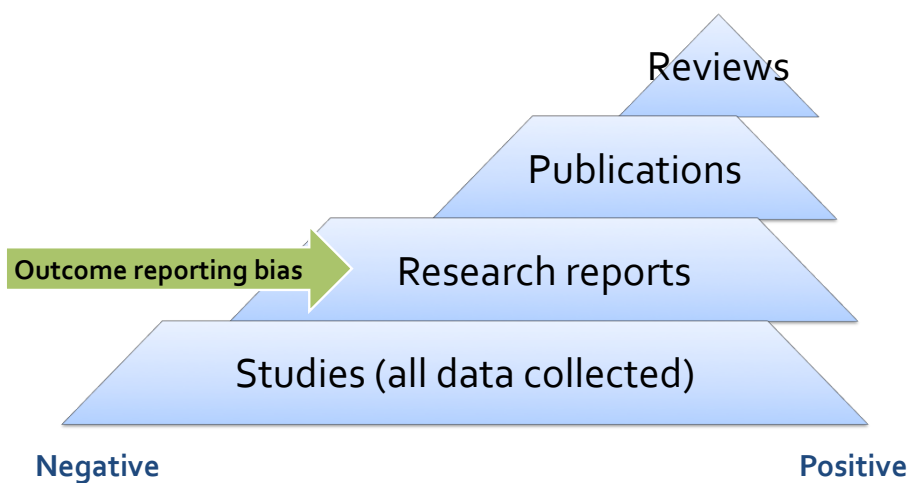


Empirical Evidence of Bias

- Dissemination of research results is a biased process (Song et al., 2009, 2010)
 - Selective reporting, publication, citation, selection of evidence
 - Confirms favored theories
 - Overestimates benefits and underestimates harms of favored treatments



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. Altman², Juan A. Arnaiz³, Jill Bloom⁴, An-Wen Chan⁵, Eugenia Cronin⁶, Evelyne Decullier⁷, Philippa J. Easterbrook⁸, Erik Von Elm^{9,10}, Carrol Gamble¹, Davina Gherzi¹¹, John P. A. Ioannidis^{12,13}, John Simes¹⁴, Paula R. Williamson¹

- 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 statistics¹

- 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

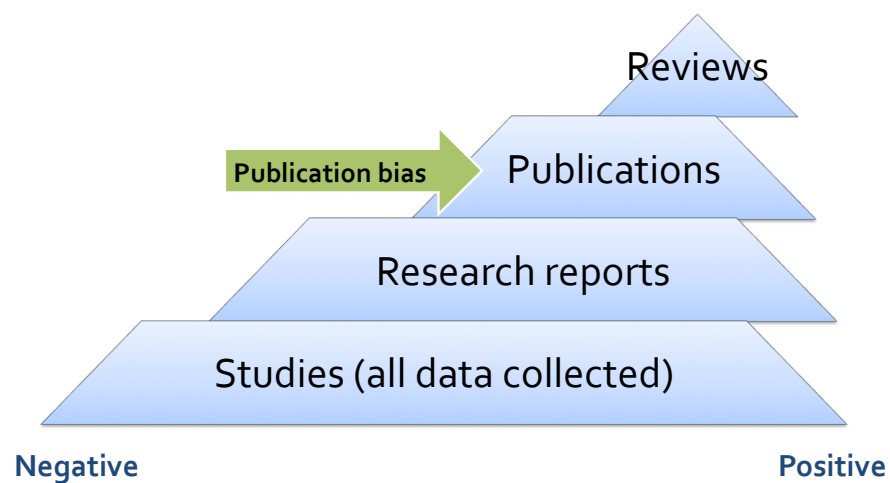
The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews

Jamie J Kirkham,¹ Kerry M Dwan,¹ Douglas G Altman,² Carrol Gamble,¹ Susanna Dodd,¹ Rebecca Smyth,³ Paula R Williamson¹

- 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 [14]	103/327 (31%)
Cooper, 1997 [32]	38/121 (status known for 117/121) (31%)
Wormaid, 2001 [25]	30/61 (status known for 55 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...



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 & Mazumdar'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)

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and importantly, the ratio of true to no relationships among the relationships tested in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller, when effect sizes are smaller, when there is a greater number and lesser prevalence of tested relationships where there is greater flexibility in analysis definitions, and when there is greater financial and other interest and publicity, and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

It can be most difficult to find a research claim to be false than true.

Published research findings are sometimes refuted by subsequent evidence, with resulting confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6,7]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here, I will examine the key factors that influence the probability of a research claim to be true or false. I will also discuss the implications of these problems for the conduct and interpretation of research.

Modeling the Positive Bias

Several methodological factors that influence the probability of a research claim to be true or false are: study power and bias, the number of other studies on the same question, and importantly, the ratio of true to no relationships among the relationships tested in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller, when effect sizes are smaller, when there is a greater number and lesser prevalence of tested relationships where there is greater flexibility in analysis definitions, and when there is greater financial and other interest and publicity, and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

The Winner's Curse

In auction theory, under certain conditions, the bidder who wins is to have overpaid. Consider oil firm bidding for drilling rights; compare estimates of the size of the reserves, as estimates differ across firms. The average of all the firms' estimates usually approximates the true reserve size. Since the firm with the highest estimate bids the most, the auction winner systematically overestimates reserves, sometimes so substantially as to be money in net terms [1]. When bid are equivalent of the statistical procedure of estimators and bids, they correct the winner's curse by shading them down. This is why experienced bid sometimes avoid the curse, as opposed to inexperienced ones [1–4]. Yet, in numerous studies, bidder behavior appears consistent with the winner's curse.

Conclusion

Citation is both an impartial scholarly method and a powerful form of social communication. Through distortions in its social use that include bias, amplification, and invention, citation can be used to generate information cascades resulting in unfounded authority of claims. Construction and analysis of a claim-specific citation network may clarify the nature of a published belief system and expose distorted methods of social citation.

ABSTRACT

Objective To understand belief in a specific scientific claim by studying the pattern of citations among papers stating it.

Design A complete citation network was constructed from all PubMed indexed English literature papers addressing the belief that β amyloid, a protein accumulated in the brain in Alzheimer's disease, is produced by and injures skeletal muscle of patients with inclusion body myositis. Social network theory and graph theory were used to analyse this network.

Main outcome measures Citation bias, amplification, and invention, and their effects on determining authority.

Results The network contained 242 papers and 675 citations addressing the belief, with 220/533 citation paths supporting it. Unfounded authority was established by citation bias against papers that inflated or weakened the belief; amplification, the marked expansion of the belief system by papers presenting no data addressing it; and forms of invention such as the conversion of hypothesis into fact through citation alone. Extension of this network into text within grants funded by the National Institutes of Health and obtained through the Freedom of Information Act showed the same phenomena present and sometimes used to justify requests for funding.

Conclusion Citation is both an impartial scholarly method and a powerful form of social communication. Through distortions in its social use that include bias, amplification, and invention, citation can be used to generate information cascades resulting in unfounded authority of claims. Construction and analysis of a claim-specific citation network may clarify the nature of a published belief system and expose distorted methods of social citation.

KEYWORDS: citation bias, amplification, invention, authority, β amyloid, inclusion body myositis, social network theory, graph theory

INTRODUCTION

Published research findings are sometimes refuted by subsequent evidence, with resulting confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6,7]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here, I will examine the key factors that influence the probability of a research claim to be true or false. I will also discuss the implications of these problems for the conduct and interpretation of research.

Modeling the Positive Bias

Several methodological factors that influence the probability of a research claim to be true or false are: study power and bias, the number of other studies on the same question, and importantly, the ratio of true to no relationships among the relationships tested in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller, when effect sizes are smaller, when there is a greater number and lesser prevalence of tested relationships where there is greater flexibility in analysis definitions, and when there is greater financial and other interest and publicity, and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

The Winner's Curse

In auction theory, under certain conditions, the bidder who wins is to have overpaid. Consider oil firm bidding for drilling rights; compare estimates of the size of the reserves, as estimates differ across firms. The average of all the firms' estimates usually approximates the true reserve size. Since the firm with the highest estimate bids the most, the auction winner systematically overestimates reserves, sometimes so substantially as to be money in net terms [1]. When bid are equivalent of the statistical procedure of estimators and bids, they correct the winner's curse by shading them down. This is why experienced bid sometimes avoid the curse, as opposed to inexperienced ones [1–4]. Yet, in numerous studies, bidder behavior appears consistent with the winner's curse.

Conclusion

Citation is both an impartial scholarly method and a powerful form of social communication. Through distortions in its social use that include bias, amplification, and invention, citation can be used to generate information cascades resulting in unfounded authority of claims. Construction and analysis of a claim-specific citation network may clarify the nature of a published belief system and expose distorted methods of social citation.

KEYWORDS: citation bias, amplification, invention, authority, β amyloid, inclusion body myositis, social network theory, graph theory

Open access, freely available online

RESEARCH

Why Current Publication Practices May Distort Science

Neal S. Young*, John P. A. Ioannidis, Omar Al-Ubaydli

Summary

The current system of publication in biomedical research provides a distorted view of the reality of scientific data that are generated in the laboratory and clinic. The system can...

How citation distortions create unfounded authority: analysis of a citation network

Steven A Greenberg, associate professor of neurology

ABSTRACT

Objective To understand belief in a specific scientific claim by studying the pattern of citations among papers stating it.

Design A complete citation network was constructed from all PubMed indexed English literature papers addressing the belief that β amyloid, a protein accumulated in the brain in Alzheimer's disease, is produced by and injures skeletal muscle of patients with inclusion body myositis. Social network theory and graph theory were used to analyse this network.

Main outcome measures Citation bias, amplification, and invention, and their effects on determining authority.

Results The network contained 242 papers and 675 citations addressing the belief, with 220/533 citation paths supporting it. Unfounded authority was established by citation bias against papers that inflated or weakened the belief; amplification, the marked expansion of the belief system by papers presenting no data addressing it; and forms of invention such as the conversion of hypothesis into fact through citation alone. Extension of this network into text within grants funded by the National Institutes of Health and obtained through the Freedom of Information Act showed the same phenomena present and sometimes used to justify requests for funding.

Conclusion Citation is both an impartial scholarly method and a powerful form of social communication. Through distortions in its social use that include bias, amplification, and invention, citation can be used to generate information cascades resulting in unfounded authority of claims. Construction and analysis of a claim-specific citation network may clarify the nature of a published belief system and expose distorted methods of social citation.

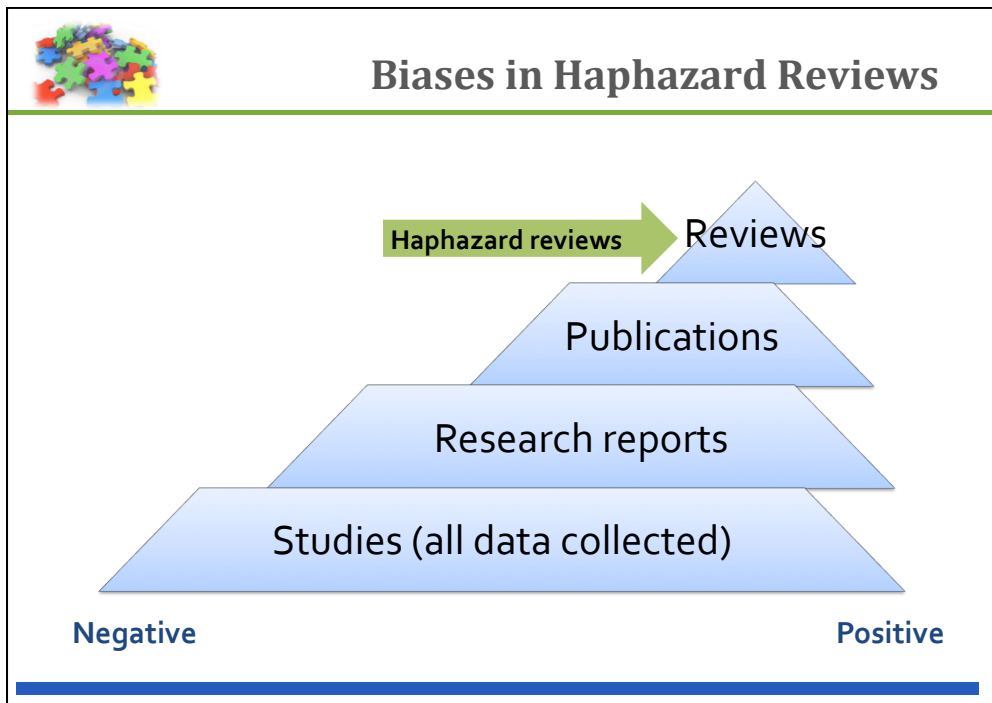
KEYWORDS: citation bias, amplification, invention, authority, β amyloid, inclusion body myositis, social network theory, graph theory

BMJ

Children's Hospital Informatics Program and Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

Correspondence to: S A Greenberg, sgreenberg@rics.bwh.harvard.edu

DOI: 10.1136/bmj.e1090





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 publication bias



Selection 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)



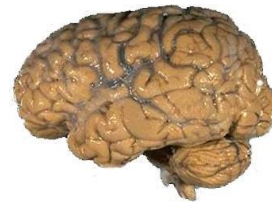
Synthesis

- 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



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.



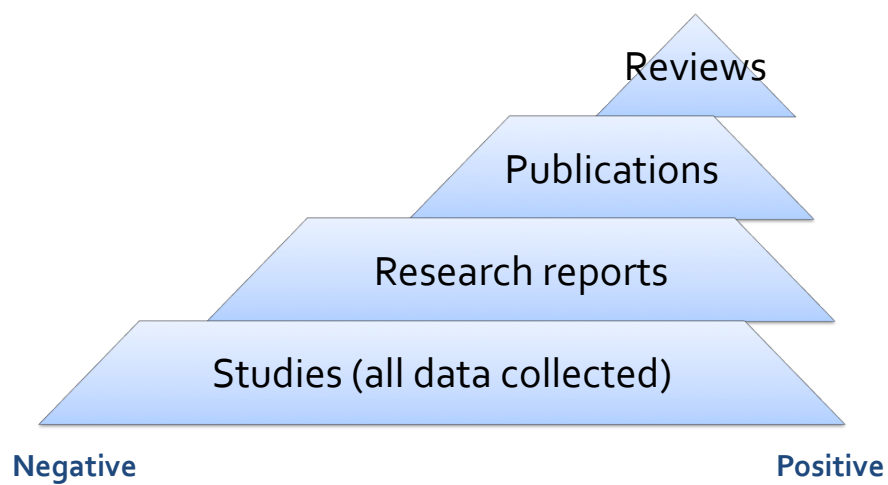


Summary

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



Research, Reports, and Reviews: Reality





Really?

- A concrete example
- One study published in 1987
 - How did investigators make use of data?
 - How did reviewers make use of data?
- Littell, 2008



An Example

Parent Training vs Multisystemic Therapy (Brunk et al., 1987)

Comparison of Multisystemic Therapy and Parent Training in the Brief Treatment of Child Abuse and Neglect

Molly Brunk
College of William and Mary

Scott W. Henggeler and James P. Whitton
Marshall University

This study evaluated the relative efficacy of two promising treatments of child abuse and child neglect: parent training and multisystemic therapy. Subjects included 43 abused/neglected children and 13 caregiver families who were randomly assigned to the treatment conditions. Self-report and observational measures were used to evaluate the effects of treatment on deviant behaviors that have been associated with child maltreatment: individual functioning, family relations, and stress/social support. Statistical analyses revealed that families who received either treatment showed decreased overall problem behavior, reduced overall stress, and a reduction in the severity of identified problems. Analyses of ecological observational measures revealed that multisystemic therapy was more effective than parent training at increasing positive child behaviors. These findings have clear implications for multisystemic therapy as reducing identified social problems. The differential influence of the two treatments were probably associated with differences in their respective treatment contexts and implementation.

In spite of the serious consequences of child maltreatment (Gelles, 1973; Marmor & Ruchelstein, 1985), relatively few investigators have developed effective treatment strategies for this problem. In recent years, however, two conceptual models of maltreatment have emerged that may provide a basis for the development of more effective interventions. These approaches include the social-situational model (Parker & Collier, 1975) and the ecological model (Bokley, 1985). Within both models, abused and neglected children are assumed to have a constellation of predisposing variables. For example, in Wade (1985) model, factors such as the child's developmental characteristics such as depression and social perceptions of child behavior; child characteristics such as difficult temperament; and social system variables such as environmental stress, social isolation, and dysfunctional family interactions.

Although the social-situational and ecological models both acknowledge the multidimensional nature of child maltreatment, the models have several key differences. The social-situational model was developed from the perspective of learning theory. Patterson and his colleagues (Patterson, 1977; Patterson, Reid, Jones, & Coatsworth, 1977) have suggested that the use of control punishment exacerbates the child's aversive behavior,

which subsequently triggers further abuse and traps the parent and child in a "coercive cycle" that maintains the abuse. Factors that contribute to this process include limited and ineffective parental control repertoire and high rates of aversive-child behavior (Gelles, Marmor, & Williams, 1978). Hence, treatment attempts to develop parents' capacity to increase positive parent-child interactions and to reduce aversive child behavior. Outcome results have indicated that parent training is effective in reducing negative parent and child behavior (Gelles, Jones, & Reid, 1984; Wade, Sullivan, & Kuhlman, 1983). It is not clear, however, that parent training increases positive behavior (Egert, 1977) that is an effective intervention for child neglect.

Within the ecological model (Bokley, 1985; Bronfenbrenner, 1979; Garbarino, 1977), child maltreatment is viewed from a systems perspective. Maltreatment results from the interaction of multiple factors that are nested within four ecological levels: the background of the parent, family relations, family interactions with extrafamilial systems, and cultural variables that support maltreatment. Although the ecological model provides an understanding of child maltreatment in the larger context of the family and culture, there have been no empirical evaluations of treatment based on this model. The multisystemic model of therapy developed by Henggeler and his colleagues (Henggeler, 1982; Henggeler & Brunk, in press; Henggeler et al., 1984), however, is consistent with the ecological model of maltreatment. Individuals are viewed as embedded in multiple systems that are ecologically nested within one another and that have both direct and indirect influences on behavior. The multisystemic model is similar to family therapy approaches (Egert, 1978; Minuchin, 1974) in its recognition that child behavior problems must be understood within their systemic context. In contrast with most family therapy models, however, the multisystemic model also treats the role of cognitive variables and extrafamilial variables in maintaining behavior problems.

This research was supported by funding to the Department of Psychology at Marshall State University granted through the Centers of Excellence Program of the State of West Virginia. The authors wish to thank the parents and to James T. Johnson, Director of Children and Youth Services, Northern Marshall Health Center, for his cooperation in this research. In addition, we greatly appreciate the editorial suggestions of Jeffrey Brunk and anonymous reviewers.

Correspondence concerning this article should be addressed to Molly Brunk, Center for Psychological Services, College of William and Mary, Williamsburg, Virginia 23185.

171

- 43 families of abused/neglected children randomly assigned to
- Parent training (PT) groups or Multisystemic Therapy (MST)
- 33 /43 families completed treatment and provided data on outcomes immediately after treatment
- 30 outcomes (scales and subscales)



Investigators' Summary

Journal of Consulting and Clinical Psychology, 2008, Vol. 76, No. 3, 111-119 Copyright 1987 by the American Psychological Association, Inc. 0893-3200/08/\$12.00 DOI: 10.1037/0893-3200.76.3.111

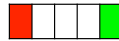
Comparison of Multisystemic Therapy and Parent Training in the Brief Treatment of Child Abuse and Neglect

Molly Brunk
College of William and Mary

Scott W. Henggeler and James P. Whelan
Memphis State University

This study evaluated the relative efficacy of two promising treatments of child abuse and child neglect: parent training and multisystemic therapy. Subjects included 13 abusive families and 13 neglectful families who were randomly assigned to the treatment conditions. Self-report and observational measures were used to evaluate the effects of treatment at time points that have been associated with child maltreatment: individual functioning, family relations, and stress/social support. Statistical analyses revealed that families who received either treatment showed decreased parental psychiatric symptoms, reduced overall stress, and a reduction in the severity of identified problems. Analyses of sequential observational measures revealed that multisystemic therapy was more effective than parent training at restructuring parent-child relations. Parent training was more effective than multisystemic therapy at reducing identified social problems. The differential influences of the two treatments were probably associated with differences in their respective treatment contexts and opportunities.

- Both groups showed decreased psychiatric symptoms, reduced stress, and reduced severity of identified problems.
- MST was more effective than PT at restructuring parent-child relations.
- PT was more effective than MST at reducing identified social problems.



Reviewers' Summaries

- Analysis of 14 published reviews of Brunk et al. (Littell, 2008)
- Most (12/14) reviewers used a single phrase to characterize results of this study, highlighting advantages of one approach (MST)
- Ignoring valuable information on relative advantages, disadvantages, and equivalent results of different approaches





Methods Matter

- Haphazard reviews may be hazardous to public health and wellbeing
 - Over-estimate positive effects of interventions
 - Under-estimate or ignore potential harmful effects
 - Minimize or ignore viable alternatives
 - Promote ineffective or potentially harmful interventions



Systematic Reviews

- Treat review process as a form of survey research and follows basic steps in research process
 - Research reports, rather than people, are surveyed
 - Each research report is “interviewed” by a coder who codes information and quantitative findings
- Aim to sum up the best available evidence in a way that minimizes errors and biases
- Use explicit, replicable research methods to identify relevant studies and objective techniques to analyze those studies
 - Develop and follow pre-determined plan (protocol)
 - Secondary analysis of existing data



Systematic Reviews (SRs)

Steps to reduce bias and error:

1. Set explicit inclusion/exclusion criteria
2. Develop and document strategies for locating all relevant studies (regardless of publication status)
3. Inter-rater agreement (reliability) on key decisions, data extraction, coding
4. Formal study quality assessment (risk of bias)
5. Meta-analysis (when possible) to synthesize results across studies



1. Explicit Criteria

- Protocol (detailed plans) for reviews made public
 - Check on review bias
- Study inclusion/exclusion criteria
 - PICOS: Populations, Interventions, Comparisons, Outcomes, Study designs
- Search strategy plans are explicit and replicable
- Data extraction forms are public
- Plans for analysis and synthesis are explicit



2. Comprehensive Search

- To reduce publication bias in reviews
- Aim to get all relevant studies, not just published studies
- Broad range of sources in addition to electronic databases
 - Personal contacts, listserves
 - Websites, research registries, research centers, government and professional organizations
 - Conference abstracts
 - Reference lists



3. Reliable Data Extraction and Coding

- Reduce bias and error in data collection and analysis
- Double screening, selection, extraction, and coding
- Inter-rater agreement (disagreements resolved with 3rd person)
- Cohen's Kappa
- May not be possible to double code everything, identify key item and decisions -- a priori



4. Formal Risk of Bias Assessment

Identify potential sources of bias in studies:

Selection bias - Systematic differences between groups at baseline

Performance bias - Something other than the intervention affects groups differently

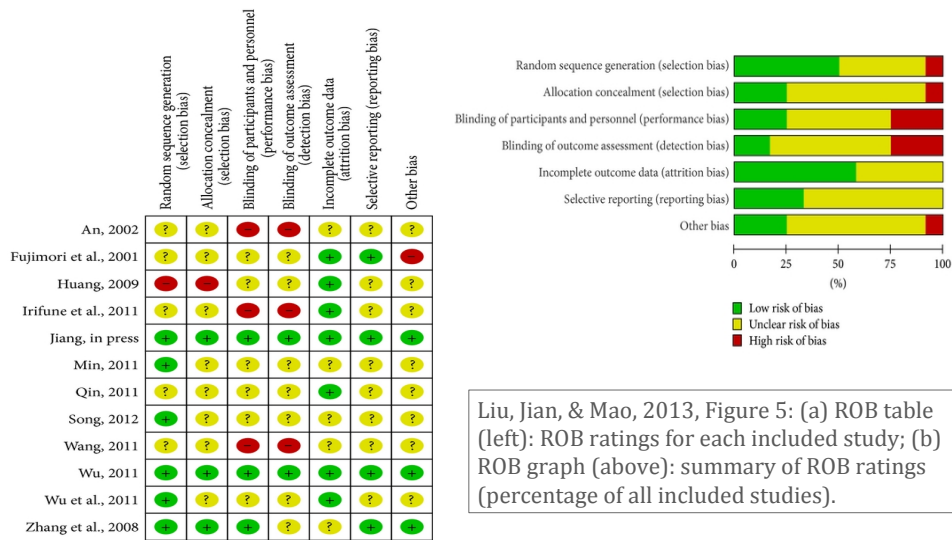
Attrition bias - Participant loss affects initial group comparability

Detection bias - Method of outcome assessment affects group comparisons

Reporting bias - Selective reporting of outcomes



Risk of Bias (ROB) Tables and Graphs





5. Reliable Synthesis

Using appropriate techniques, meta-analysis if possible

- Meta-analysis is possible with 2 or more studies that report quantitative results
- Sensitivity analysis to check on effects of decisions made during the review
 - Especially departures from the protocol
 - Run synthesis under different assumptions



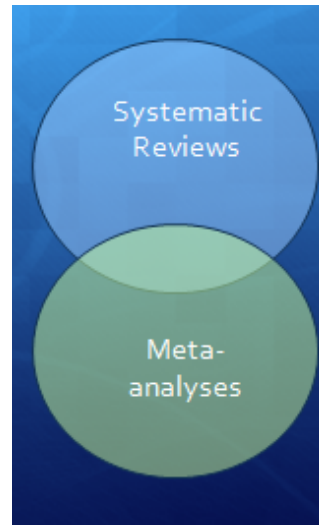
Meta-Analysis (MA)

Set of statistical procedures used to assess

- Averages across studies
- Variations across studies
- Potential sources of variation (moderators)
- Risk of bias (e.g., tests for publication & small sample bias)



- Systematic reviews don't always include meta-analysis
 - Might include narrative synthesis (or no synthesis)
 - Can include multiple meta-analyses
- Meta-analyses are not always based on systematic reviews
 - Many use convenience sample of published studies
 - Vulnerable to publication and dissemination biases



Some “Systematic Reviews” Aren’t

- Evidence-based standards for SRs & MA
 - Based on methodological research (Cochrane Library)
- Standards for conduct of SRs
 - Developed by Cochrane and Campbell Collaborations (Higgins & Green, 2011; IOM, 2011; MECIR, 2013)
- Standards for reporting SRs & MA
 - PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses; Moher et al., 2009)
- Standards not followed by US Evidence-based Practice Centers, most peer-reviewed journals, etc.



Minimum Standards Important

- Reviews should take reporting and publication biases into account
 - Include extensive search for grey literature
 - Assess risk of bias (in original studies and in review)
 - Many methods available for assessing potential effects of reporting and publication biases in reviews (Dwan et al., 2010; Rothstein, Sutton, Bornstein, 2005; Moreno et al., 2009)
- Reviews should use adequate synthesis methods
 - Narrative reviews are unreliable
 - Vote counting is inadequate
 - Meta-analysis is the best available method for quantitative data
 - And it's not hard to do



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/>



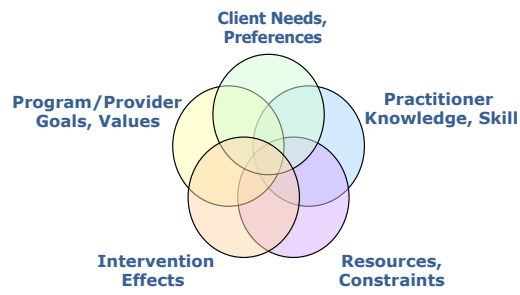
Using Evidence in Context

Policy and practice are informed by

- Many types of evidence (qualitative, quantitative, anecdotal) on
- Many topics

Evidence isn't enough

- Need to consider values, preferences, resources, ethics, legal constraints, etc.



Adapted from: Gibbs (2003), Davies (2004)



Thank You

Questions/Comments?

Contact Information

Julia Littell: jlittell@brynmawr.edu
Brandy Maynard: bmaynar1@slu.edu



Websites

- Cochrane Collaboration:
www.cochranecollaboration.org
- Campbell Collaboration:
www.campbellcollaboration.org
- David B. Wilson's effect size calculator:
http://www.campbellcollaboration.org/resources/effect_size_input.php



Books

- Borenstein et al. (2009). Introduction to meta-analysis.
- Card (2011). Applied meta-analysis for social science research.
- Cooper (2009). Research synthesis and meta-analysis: A step-by-step approach.
- Cooper, Hedges, & Valentine (2009). Handbook of Research Synthesis and Meta-Analysis.
- Higgins & Green (2011). Cochrane Handbook.
- Lipsey & Wilson (2001). Practical Meta-analysis.
- Littell, Corcoran, & Pillai (2008). Systematic Reviews and Meta-Analysis.
- Petticrew & Roberts (2005). Systematic Reviews in the Social Sciences: A Practical Guide.
- Rothstein, Sutton, & Borenstein (2005). Publication bias in meta-analysis: Prevention, Assessment, and Adjustment.