A Retrospective Analysis of Dissemination Biases in the Brief Alcohol Intervention Literature

Emily E. Tanner-Smith and **Joshua R. Polanin**
Peabody Research Institute, Vanderbilt University

Abstract

This study examined dissemination and reporting biases in the brief alcohol intervention literature. We used retrospective data from 179 controlled trials included in a meta-analysis on brief alcohol interventions for adolescents and young adults. We examined whether the magnitude and direction of effect sizes were associated with publication type, identification source, language, funding, time lag between intervention and publication, number of reports, journal impact factor, and subsequent citations. Results indicated that effect sizes were larger for studies that had been funded ($b = 0.14$, 95% confidence interval [CI] [0.04, 0.23]), had a shorter time lag between intervention and publication ($b = -0.03$, 95% CI [-0.05, $-.001$]), and cited more frequently ($b = 0.01$, 95% CI [$+0.00$, 0.01]). Studies that were cited more frequently by other authors also had greater odds of reporting positive effects (odds ratio = 1.10, 95% CI [1.02, 1.18]). Results indicated that time lag bias has increased recently: larger and positive effect sizes were published more quickly in recent years. We found no evidence, however, that the magnitude or direction of effects was associated with location source, language, or journal impact factor. We conclude that dissemination biases may indeed occur in the social and behavioral science literature, as has been consistently documented in the medical literature. As such, primary researchers, journal reviewers, editors, systematic reviewers, and meta-analysts must be cognizant of the causes and consequences of these biases, and commit to engage in ethical research practices that attempt to minimize them.

Keywords

brief alcohol intervention; dissemination bias; meta-analysis; publication bias; reporting bias

---

The evidence-based practice movements in psychology and medicine have highlighted the importance of using current best evidence for making research, practice, and policy decisions (APA Presidential Task Force, 2006; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Well-implemented randomized controlled trials with low levels of attrition are generally considered the gold standard research design for evaluating the causal effects of psychological interventions. Findings from individual research studies will rarely provide sufficient evidence for practice or policy recommendations, however, because those findings are subject to sampling error and other study-specific biases (Ioannidis, 2006).

Correspondence concerning this article should be addressed to Emily E. Tanner-Smith, Vanderbilt University, Peabody Research Institute, Department of Human and Organizational Development, Box 0181 GPC, Vanderbilt University, Nashville, TN 37203-5721; Phone 615-322-6304; Fax 615-322-0293; e.tanner-smith@vanderbilt.edu..

Meta-analyses that synthesize findings across multiple primary studies can provide more credible information regarding current best evidence than findings from any single study alone. As such, findings from meta-analyses are increasingly used to guide patient decision making, future research priorities, and the allocation of resources.

As with any type of scientific research, meta-analyses are not immune to problems, however, and thus it is crucial to understand the potential biases that may threaten their internal validity. Including biased estimates in a meta-analysis can lead to biased conclusions. Perhaps the most widely recognized type of bias that meta-analyses may be subject to is publication bias, defined broadly here as the case in which the publication or non-publication of research findings depends on the nature and direction of those findings (Sterne, Egger, & Moher, 2008). Publication bias is sometimes referred to as the file drawer problem, in reference to Rosenthal's (1979) statement that "journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant (e.g., p > .05) results." (p. 638). Indeed, there is empirical evidence of publication bias, such that studies with statistically significant, "positive" (i.e., in the expected direction), or confirmatory results are often more likely to be published than those with non-significant or "negative" results (e.g., Decullier, Lhéritier, & Chapuis, 2005; Dickersin, Min, & Meinert, 1992; Dwan et al., 2008; Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009).

Publication bias can be problematic because summaries of the current best evidence that are often used to advance policy and practice (e.g., meta-analyses based on systematic literature reviews) are assumed to represent an unbiased synthesis of the entire population of studies on a given research topic. In the presence of publication bias, however, a review might only include a subset of studies from the larger population, which could be systematically different from the larger population to which one wishes to generalize. Summary statements, therefore, might be biased, potentially over- or under-estimating the magnitude of effects in a given research literature. As a result, publication bias has the potential to hamper the falsification of psychological theories, undermine scientific principles of the replicability of research, misguide patient and policy decision-making, and potentially divert resources from more appropriate interventions. It is little surprise, then, that publication bias has been call the "800 lb gorilla in psychology's living room" (Ferguson & Heene, 2012; p. 556).

Yet publication bias is only one part of a larger problem—dissemination bias—defined here as the phenomenon whereby the selective reporting, selective publication, and/or selective inclusion of scientific evidence in a systematic review yields the incorrect, or biased answer, on average. As outlined by Bax and Moons (2011), dissemination bias can result from the selective outcome reporting among primary study authors (outcome reporting bias), the selective publication of scientific outcome findings (publication bias), and the selective inclusion of research findings in a synthesis of the research literature (selection or inclusion bias). This framework is similar to Chalmers, Frank, and Reitman's (1990) early discussion of the three stages of publication bias (prepublication, publication, and post publication bias), but it clarifies that scientific evidence is disseminated in sequential steps, that selection processes can occur during each of these steps (reporting, publication, and inclusion), and that each of these selection processes may or may not ultimately lead to bias.

Prior Research

Selective outcome reporting

Historically, one of the most difficult dissemination biases to study has been selective outcome reporting bias, given that it is often impossible to detect in a retrospective study (e.g., in a traditional retrospective meta-analysis of existing research). This has changed with the advent of prospective trial registration and protocol publishing. In theory, prospective trial registration should minimize selective outcome reporting by forcing primary study authors to specify their analytic strategies prior to data collection and analysis. Trial registration and protocol publishing are more common in the health and medical sciences, but have yet to fully take hold in the psychological, social, and behavioral sciences (but see Cooper & VandenBos, 2013). But even in medicine, many trials are not prospectively registered, and those that do may still engage in selective outcome reporting (Rasmussen, Lee, & Bero, 2009; Redmond et al., 2013). For instance, Mathieu and colleagues (2009) recently examined 323 randomized controlled trials in cardiology, rheumatology, and gastroenterology. They found that only 46% of the trials were adequately registered; among those, 31% of trials reported primary outcomes different from those that were registered; and for the small group of studies for which it could be assessed, there was evidence that the outcome reporting discrepancies consistently favored statistically significant results. In a recent systematic review, Dwan and colleagues (2011) reviewed studies of cohorts of randomized controlled trials that compared the content of study protocols with final reports. They located 16 studies assessing the effects on 54 trials (most of which were in health or medicine), and concluded that discrepancies between protocols and final study reports were indeed common.

There is also evidence that outcome reporting bias may be a problem in the social and behavioral sciences. John, Lowenstein, and Prelec (2012) recently reported findings from an anonymous survey of over 2,000 academic psychologists in the United States, in which an alarming majority of respondents self-admitted failing to report all of a study's dependent measures in a paper. Another recent study (Pigott, Valentine, Polanin, Williams, & Canada, 2013), compared the reported effects of educational interventions from 79 dissertations versus the effects that were subsequently published in journal articles, and found that study authors subsequently published statistically significant outcomes more often than non-significant findings.

Selective publication

Extensive empirical literature suggests that studies are more likely to be published if they report statistically significant, positive, or confirmatory (in the expected direction) results. This bias most often results from the actions of primary study authors, rather than journal editors or peer reviewers (Olson et al., 2002). Examination of selective publication bias typically involve cohort studies of protocols approved by research ethics committees (e.g., Decullier, et al., 2005; Hall, de Antueno, & Webber, 2007); meta-analyses synthesizing results from these types of studies indicate that studies with significant results are more likely to have been published than those with non-significant results (Egger & Smith, 1998; Hopewell et al., 2009). McLeod and Weisz (2004), for instance, examined whether reported

effects for clinical trials of psychotherapy for children and adolescents varied for published studies versus dissertations. Based on results from 121 dissertations and 134 published studies, they found that dissertations reported effects that were almost one-half the size of those reported in the published literature, even after controlling for other study method characteristics. Polanin and Tanner-Smith (2014) found similar results in a review of 31 meta-analyses published in psychology and education journals, such that average effect sizes were 0.40 standard deviations larger in published studies than unpublished studies.

Selective inclusion

Empirical evidence varies in terms of the presence and magnitude of selective inclusion biases. For instance, there is relatively consistent evidence of time lag bias, such that studies with positive or statistically significant findings are more likely to be published quickly relative to those with null or negative findings (Hopewell, Clarke, Stewart, & Tierney, 2007; Decullier et al., 2005; Ioannidis, 1998; Liebeskind, Kidwell, Sayre, & Saver, 2006; Stern & Simes, 1997). Almost all of this research has been conducted in the medical sciences, however, thus to date there is limited evidence regarding time lag bias in the social and behavioral sciences.

Few studies have examined duplicate publication bias, or the bias resulting from the multiple publication of research findings resulting from the nature and/or direction of research results. Easterbrook, Goplan, Berlin, and Matthews (1991) conducted a retrospective survey of projects approved by a research ethics committee and found that studies with significant results were more likely to be published in multiple reports. In a retrospective cohort of protocols approved by a research ethics committee, Decullier et al. (2005) found no evidence that studies with confirmatory or validating results were more or less likely to be reported in multiple publications. Yet other studies have also documented that trials reporting larger treatment effects were more likely to be reported in multiple publications (Tramèr, Reynolds, Moore, & McQuay, 1997).

Evidence has also been mixed for location bias, or the bias whereby the nature and/or direction of research findings is associated with different ease of access to reports. The most commonly used measure to explore location bias has been journal impact factor, with the presumption that articles located in journals with higher impact factors will be more easily identified, located, and accessible. Whereas some empirical studies have reported that significant and/or positive results are more likely to be published in higher impact factor journals (De Oliveira, Chang, Kendall, Fitzgerald, & McCarthy, 2012; Etter & Stapleton, 2009; Jannot, Agoritsas, Gavet-Ageron, & Perneger, 2013; Timmer, Hillsden, Cole, Hailey, & Sutherland, 2002), others have found no such evidence (Callaham, Wears, & Weber, 2002; Fanelli, 2010; Hall et al., 2007).

There has also been inconsistent evidence of citation bias, or bias resulting from studies with large or significant studies being more likely to be cited by other scholars. Citation bias can be potentially problematic for meta-analyses if more frequently cited articles are more easily identifiable through non-traditional literature searching. Indeed, most meta-analyses supplement electronic bibliographic literature searches by scanning the reference lists of other studies and review articles. If grey literature or difficult-to-locate studies with large or

statistically significant results are more likely to be cited than other grey literature studies with small or non-significant effects, this could potentially bias the findings of the larger meta-analysis. Several studies have documented that studies with significant or positive results are more likely to be cited by other studies (Etter & Stapleton, 2009; Jannot et al., 2013; Nieminen, Rucker, Miettunen, Carpenter, & Schumacher, 2007). However, in an analysis of articles submitted to the Society for Academic Emergency Medicine meeting, Callaham, Wears, and Weber (2002) found no evidence that positive study outcomes were correlated with subsequent citations. Again, however, most of the research on citation bias has been conducted in the field of medicine, so less is known regarding its presence in the social and behavioral science literatures.

Similar to the evidence on citation bias, findings of language bias have varied (Egger et al., 1997; Grégoire, Derderian, & Le Lorier, 1995). For instance, Jüni et al. (2002) examined 309 Cochrane reviews to explore language bias and found that effects were on average 16% more beneficial in non-English studies than those in English languages. Trials published in non-English languages were generally smaller and more likely to report statistically significant results, but were also of lower quality.

Finally, there has been consistent evidence that studies receiving external funding are more likely to get published and/or publish significant or positive results compared to unfunded studies (e.g., Cunningham, Warme, Schaad, Wolf, & Leopold, 2006; Dickersin et al., 1992; Easterbrook et al., 1991), whereas studies receiving funding from pharmaceutical companies are less likely to be published compared to those not funded by pharmaceutical companies (Hall et al., 2007). Again, however, most of this research has been conducted in the health and medical sciences.

Changes in dissemination biases over time

Given Chalmers and colleagues' (1990) call to action for the reduction of publication bias in research literature, we might expect that dissemination biases have become less common over the past two decades as researchers, editors, and peer reviewers have become increasingly aware of this issue and therefore insisted on research practices that reduce these possible biases. There is evidence that systematic reviewers and meta-analysts have indeed increased their effort to search for unpublished studies, and to use methods to test for publication bias in their analyses (Parekh-Bhurke et al., 2011). Although these findings are encouraging, there is also recent evidence that the reporting of positive outcomes in research papers has increased over time, particularly in psychology and psychiatry (Fanelli, 2012).

Yet few empirical studies have examined whether specific types of dissemination biases have changed over time. One notable exception is the study conducted by Liebeskind and colleagues (2006), which examined the characteristics of 178 controlled clinical trials on acute ischemic stroke that were reported between 1955 and 1999. In their analysis, they found no evidence of time-lag bias among studies, nor any evidence that time-lag bias might have changed over time. Rather, they found that trials became more complex over time, and had longer enrollment phases; longer time lags simply reflected longer enrollment phases, not longer periods between study completion and study publication (with review and editorial phases in publication actually becoming shorter over time). We are otherwise

unaware of any other studies that have empirically examined whether different dissemination biases have changed over time.

In summary, understanding and preventing dissemination biases should be a priority in the psychological sciences. Dissemination biases are problematic in their own right, but also troublesome given their potential distorting effects on systematic reviews and meta-analyses used to guide patient, practitioner, and policy decision-making. With few exceptions (McLeod & Weisz, 2004; Niemeyer, Musch, & Pietrowsky, 2013; Pigott et al., 2013), most empirical research on dissemination biases has been conducted in the fields of health and medicine. Indeed, in the seminal edited book on publication bias in meta-analysis, Rothstein, Sutton, and Borenstein (2005) called for social scientists to attend to issues of publication bias, given the clear documentation of these potential biases in the medical literature (pg. 55). Furthermore, the extant empirical research on dissemination biases has often only focused on one specific bias (e.g., time lag bias) at a time, rather than examining several dissemination biases in tandem. Although this piecemeal approach has yielded valuable information regarding individual types of dissemination biases, it does little to advance a more comprehensive understanding of the various types of biases that may be at work. As such, there is a need for further empirical research examining the existence and possible effects of multiple forms of dissemination biases in the social and behavioral sciences, which has important implications for the advancement of psychological science as well as systematic reviewing and meta-analysis methods.

Study Objectives

This study involved a retrospective examination of dissemination biases in the literature on brief alcohol interventions for youth. We used data from a recently completed, large-scale systematic review and meta-analysis on the effectiveness of brief alcohol interventions for adolescents and young adults to address the following research questions: (a) Is the magnitude of effect sizes associated with the following reporting characteristics: publication type, identification source, language, funding, time lag between intervention and publication, number of reports, impact factor, and subsequent citations? (b) Is the direction (positive/negative) of effect sizes associated with the following reporting characteristics: publication type, identification source, language, funding, time lag between intervention and publication, number of reports, impact factor, and subsequent citations? (c) Have the associations between the magnitude/direction of effect sizes and reporting characteristics changed over time? The first two research questions assess the existence and magnitude of possible dissemination biases in this literature. The third research question assesses whether dissemination biases have changed over time.

Methods

Inclusion and Exclusion Criteria

This study analyzed data from a recent systematic review and meta-analysis on the effectiveness of brief alcohol interventions for adolescents and young adults (see Tanner-Smith & Lipsey, 2014 for more details). We included experimental or quasi-experimental research designs that examined the effects of brief alcohol interventions (i.e., no more than 5

hr of total contact time) relative to a comparison condition of no treatment, wait-list control, or treatment as usual. Eligible studies included adolescent (age 18 and under) and young adult (ages 19-25, or collegiate undergraduate students) participant samples, and eligible studies were required to report at least one post-intervention outcome related to alcohol consumption or an alcohol-consumption related problem (e.g., drunk driving). There were no geographic, language, or publication status restrictions on eligibility, but studies must have been conducted in 1980 or later.

Search Strategies

Primary studies were identified using an extensive literature search strategy that aimed to identify the entire population of published and unpublished studies that met the aforementioned inclusion criteria. The depth and breadth of the literature search were strengths of this meta-analysis, which make it particularly well-suited for use in a retrospective examination of possible dissemination biases. The following electronic bibliographic databases were searched through December 31, 2012: CINAHL, Clinical Trials Register, Dissertation Abstracts International, ERIC, International Bibliography of the Social Sciences, NIH RePORTER, PsycARTICLES, PsycINFO, PubMed, Social Services Abstracts, Sociological Abstracts, and WorldWideScience.org. Numerous additional sources were searched in an attempt to locate grey literature (e.g., Australasian Medical Index, Canadian Evaluation Society's Grey Literature Database, Index to Theses in Great Britain and Ireland, International Clinical Trials Registry, KoreaMed, Social Care Online, SveMed +; see Tanner-Smith & Lipsey, 2014 for full list). The search strategy also included harvesting of references from bibliographies of all screened and eligible studies, bibliographies in prior narrative reviews and meta-analyses, hand-searching of key journals, and online curriculum vitae (when available) of first authors of eligible studies.

Coding Procedures

Six researchers screened all abstracts and titles to eliminate any clearly irrelevant study reports. Staff first screened the abstracts/titles of 500 randomly selected reports; all disagreements were discussed until 100% consensus was reached. The remaining abstracts/titles were screened by the same staff, and the first author reviewed all screening decisions as a second coder. Disagreements between the research staff and the first author were discussed until consensus was reached. Ambiguity about the potential eligibility of a report based on the abstract (or title, when abstracts were not available) resulted in the retrieval of the full text report to make the eligibility decision. Full text versions of reports were then retrieved for all reports that were not explicitly ineligible at the abstract screening phase. The same six researchers then screened the full text reports to make final eligibility decisions using the same procedure. Again, the first author served as a second coder for all full text eligibility decisions.

Coding was conducted by six individuals who participated in several weeks of initial training led by the first author as well as weekly coding meetings. During initial training, five studies were coded by all coders, which convened as a group to resolve any coding discrepancies until 100% consensus was attained on all coded variables. After the training period, all coding questions were addressed in weekly meetings and decided via consensus

with the group. The first author double-checked all of the coded items, and resolved any coding discrepancies via consensus with the individual coder.

Statistical Methods

Effect size metric—We used the Hedges' *g* standardized mean difference effect size (Hedges, 1981) to index post-intervention differences in alcohol consumption for youth who received brief alcohol interventions relative to those in comparison conditions. All effect sizes were coded so that positive values represented better outcomes (e.g., lower alcohol consumption, higher abstinence). For binary outcomes (e.g., group differences in abstinence), the Cox transformation outlined by Sánchez-Meca and colleagues (2003) was used to convert log odds ratio effect sizes into standardized mean difference effect sizes. Effect size and sample size outliers were Winsorized to less extreme values to prevent distortion of the meta-analysis results (Lipsey & Wilson, 2001). We inflated the standard errors of effect size estimates that originated from cluster-randomized trials when the authors did not properly account for the cluster design in their own analyses.

Effect size moderators—Eight reporting characteristics were explored as potential effect size moderators in the analysis. Publication type was measured for each report within a study sample, so that the same study could be reported in multiple publication types; it was categorized as *journal article*, *dissertation/thesis*, or *other publication type* (e.g., technical report, book chapter).

Identification source indexed the database or location at which each report was located; it was categorized as a *main bibliographic database* (CINAHL, ERIC, IBSS, ProQuest, PsycINFO, PubMed, Social Services Abstracts, or Sociological Abstracts), *grey literature database* (Alcohol Concern, Google Scholar, Index to Theses), *other grey literature source* (author curriculum vitae search, forward citation search, INEBRIA conference proceeding, JMATE conference proceeding, journal hand-search), or *reference list* (harvested from a bibliography/reference list).[1]

Language was measured for each report (i.e., document) within a study sample, such that the same study sample could be described in multiple reports in multiple languages; it was categorized as *English* or *non-English* (e.g., Spanish, Portuguese, German).

Funding of study was measured for each study sample and categorized with two dummy variables indicating whether any report associated with the study acknowledged receipt of funding (*1 = yes*; *0 = no*); and whether any report associated with the study acknowledged receiving funding from the alcohol beverage industry (*1 = yes*; *0 = no*).

Time lag was measured as the time lag (in years) between the year the intervention was delivered and the year the study results first appeared in print. If interventions were

---

[1]The identification source variable indicated the way in which the report was located during the literature search for the parent meta-analysis, and was independently coded from the publication type variable noted above. For instance, if a published journal article was identified in a reference list but not identified in the electronic bibliographic database search (e.g., it was relevant but missed for some reason) it would have been coded as being identified in a *reference list,* and the publication type code would have been a *journal article*.

delivered over the span of multiple years, we were conservative in our coding and used the most recent year for the intervention year (e.g., intervention delivered between 1994-1996 would have been coded as having an intervention year of 1996).

Number of study reports was coded as the total number of reports (i.e., documents) located in the literature search and identified as being linked to the common study sample.

Impact factor was coded as the one-year Thomson Reuters impact factor for each journal publication, based on the JCR Sciences and Social Sciences Editions in Web of Knowledge. To account for variation in journal ratings over time, impact factors were coded separately for each publication year of a given journal. We were unable to locate impact factors for a few journals (see Appendix), which were treated as missing and deleted listwise from analyses. Sensitivity analyses (not shown, but available from authors upon request) using a time-invariant journal impact factor based on the 2012 ISI ratings yielded substantively similar results to those reported here.

Forward citations was measured as the number of subsequent citations each study report received in Google Scholar, per year since publication. Sensitivity analyses using forward citation data collected from Web of Science (not shown, but available from authors upon request) yielded substantively similar results to those reported here.

Although not explicitly framed as effect size moderators, the following study method characteristics were also used in analyses as covariates to control for any potential confounding between the aforementioned reporting characteristics and the quality of the primary studies: randomized study design (1 = yes; 0 = non-randomized quasi-experimental design); whether the authors explicitly reported conducting intent-to-treat analysis and there was no evidence to the contrary in the study report (*1 = yes; 0 = no*); overall attrition (% from pretest to first follow-up); publication year; and the total number of follow-up waves in each study.

Missing data—We contacted primary study authors when studies failed to include enough statistical information needed to estimate effect sizes. Overall, we had an excellent response rate from authors, most of whom provided the needed information. However, not all information could be collected for all variables: there were missing data for three of the reporting characteristics of interest (45% of studies were missing time lag information because the year in which the interventions were conducted was not reported; 13% of studies reported in journal articles were missing impact factor data; and 3% of studies were missing forward citation counts). Missing data were not imputed; all analyses using these variables were estimated using listwise deletion.

Analytic strategies—Most studies reported multiple measures of alcohol consumption (e.g., frequency of consumption, quantity consumed, blood alcohol concentration). Therefore, we used meta-regression models with robust variance estimates to handle the statistical dependencies in the data (Hedges, Tipton, & Johnson, 2010). Given presumed heterogeneity in interventions and participants in the primary studies, random effects statistical models were used for all analyses, implemented with weighted analyses using

inverse variance weights that included both within-study and between-study sampling variance components. The model was conceived as the following:

$$T_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \ldots + \beta_p X_{pij} + u_{ij} + \epsilon_{ij},$$
$$u_{ij} + e_{ij} \tilde{} N(0, v^*)$$

where T_{ij} represented the effect size i in study j, $\beta_p X_{1ij}$ represented the relationship between the predictor X_p and effect size i in study j, and $u_{ij} + \epsilon_{ij}$ were assumed normally distributed with a mean of zero and a variance that included a random effect for the between study variance component (Raudenbush, 2009). Weighted logistic regression models with clustered robust standard errors (Williams, 2000) were used to examine whether the reporting characteristics were associated with the odds of a reported effect size being positive (greater than zero). Finally, multiplicative interaction terms between each reporting characteristic and publication year (all grand mean centered) were used to explore whether the association between reporting characteristics and effect size magnitude and/or direction varied over time.

To address our research questions, we first estimated a series of regression models examining each reporting characteristic separately (Models I – VIII), as well as in a comprehensive model (Model IX) that included all reporting characteristics and covariate controls (note, however, that journal impact factor was omitted from the comprehensive model because it was validly missing for all effect sizes published in non-journal articles). All models controlled for the age of participant samples (adolescent vs. young adults), study design (randomized vs. quasi-experimental), intent-to-treat analysis, attrition, and the total number of study waves, to account for possible confounding between reporting characteristics and study quality measures (Lipsey, 2009). These control variables were included because they were either identified in the larger meta-analysis as correlated with effect size magnitude (Tanner-Smith & Lipsey, 2014), or because they were identified a priori as potentially confounded with any of the reporting characteristics of interest (e.g., time lag and total number of study follow-up waves).

Results

Study Reporting Characteristics

As reported in Tanner-Smith & Lipsey (2014), the literature search from the original meta-analysis identified 7,593 research reports. Of those, 2,467 were duplicates and dropped from consideration, 2,642 were screened out as ineligible at the abstract phase, and 2,171 were screened out as ineligible at the full-text screening phase. The original meta-analysis synthesized findings from 185 study samples that reported at least one post-intervention effect size indexing group differences in either alcohol consumption or an alcohol-related problem (e.g., driving under the influence). Results from the meta-analysis indicated that brief alcohol interventions resulted in significant reductions in alcohol consumption among adolescents ($\bar{g} = 0.27$, 95% CI [0.16, 0.38], $\tau^2 = 0.04$, $Q = 46.92$) and young adults ($\bar{g} = 0.17$, 95% CI [0.13, 0.20], $\tau^2 = 0.02$, $Q = 334.74$). The current study focuses only on the 179 studies that reported at least one alcohol consumption related outcome (i.e., 6 samples only

reported post-intervention effects on alcohol-related problem outcomes, which were not included in the current analysis). See the appendix for a list of references to the primary study reports providing effect size data included in the current analysis.

Most studies were published recently and used a randomized design (89%), few studies (39%) reported conducting intent-to-treat analyses, and the average sample attrition was .22 ($SD = 0.21$) across studies. The majority of reports (75%) were published in journal articles, approximately 76% were identified from a main electronic bibliographic database (i.e., CINAHL, ERIC, IBSS, ProQuest, PsycINFO, PubMed, Social Services Abstracts, Sociological Abstracts), and 11% were identified by harvesting the references in the bibliographies of previously identified studies. Almost all (97%) of the documents were reported in English, although a handful of study reports were available in another language (most commonly Spanish, Portuguese, or German). The majority of studies (69%) acknowledged funding for at least a portion of their research, the average time lag between intervention implementation and first report of study findings was 2.64 years ($SD = 1.46$), and the average study presented findings in 1.72 reports ($SD = 1.24$). Finally, most reports were subsequently cited by other study authors 5.62 times per year ($SD = 7.71$), and the average journal impact factor for reports published in journal articles was 2.99 ($SD = 2.92$).

As shown in Table 1, several of the study characteristics were intercorrelated at moderate levels. For instance, randomized trials were more likely to be published in English; journal articles had longer time lags between intervention and publication, a higher number of associated study reports, and were subsequently cited more often than non-journal articles (e.g., dissertations/theses, technical reports, book chapters). Funded studies were also more likely to have a higher number of associated study reports, more subsequent citations, and be published in journals with higher impact factors. Studies with more associated study reports were also cited more often, and published in journals with higher impact factors.

Relationship between Reporting Characteristics and the Magnitude of Effects

Our first objective was to examine the relationship between the magnitude of effect sizes and the eight measured reporting characteristics: publication type, identification source, language, funding, time lag between intervention and publication, number of reports, impact factor, and forward citations. Figures 1 and 2 present the distribution of reported effect sizes by reporting characteristic, for the categorically and continuously measured reporting characteristics, respectively. Table 2 presents unstandardized coefficients and 95% confidence intervals (using robust standard errors) from mixed-effects meta-regression models estimating the association between these reporting characteristics and effect size magnitude, after controlling for the five study characteristics.

Results from the models that examined each reporting characteristic separately (Models I – VIII, Table 2) indicated that journal articles reported larger effect sizes ($b = 0.10$, 95% CI [0.01, 0.19]) than those published in other formats (e.g., theses/dissertations, book chapters, technical reports). For instance, the average standardized mean difference effect size as reported in journal articles was 0.15 (95% CI [0.12, 0.20]), versus 0.07 for theses (95% CI [−0.05, 0.11]), and 0.03 for other publication types (95% CI [−0.05, 0.11]) (see Figure 1).

Studies that acknowledged receiving funding had larger effect sizes than those that did not receive funding ($b = 0.08$, 95% CI [0.01, 0.16]). The average effect size as reported in funded studies was 0.16 (95% CI [0.11, 0.20]), versus 0.08 for unfunded studies (95% CI [0.02, 0.14]) (see Figure 1). Because funded studies may have larger sample sizes (and therefore higher statistical power to detect effects) we conducted post-hoc analyses that additionally controlled for primary study sample size; results were substantively unchanged and indicated that funded studies had larger effect sizes than unfunded ones, even after controlling for sample size ($b = 0.09$, 95% CI [0.01, 0.17]). The majority of funded studies (73%) acknowledged funding from the National Institute on Alcohol Abuse and Alcoholism, with the remainder often citing other government funding agencies. Notably, three studies acknowledged receipt of funding from a private foundation supported by the malt beverage industries of the United States and Canada. We therefore conducted post-hoc analyses to examine whether this industry funding was correlated with effect size magnitude (results not shown in tables). Studies funded by the malt beverage industry reported significantly smaller effect sizes than those that did not; the average effect across the 17 effect sizes in those three studies was practically zero, at 0.01 (95% CI [−0.03, 0.04]).

Results provided no evidence that identification source, language, time lag, number of study reports, forward citations, and journal impact factor were associated with effect size magnitude (Table 2). However, the comprehensive model (IX) indicated that after controlling for the other reporting characteristics and control variables, funded studies continued to be correlated with larger effects ($b = 0.14$, 95% CI [0.04, 0.23]), longer time lags were associated with smaller effects ($b = -0.03$, 95% CI [−0.05, −0.01], and studies that were subsequently cited more frequently were those with larger effects (see also Figure 2; $b = 0.01$, 95% CI [+0.00, 0.01]). Both of these significant effects were relatively modest in practical terms, however, equivalent to a 0.03 standard deviation change with each additional year of time lag and only a 0.01 standard deviation change with each additional citation per year. Finally, the association between publication type and effect size magnitude that emerged in Model I was attenuated in the comprehensive models; this attenuation was likely caused by the intercorrelations between publication type, funding, time lag, and forward citations (see Table 1).

Relationship between Reporting Characteristics and the Direction of Effects

The second objective of the study was to examine whether study reporting characteristics were associated with the direction of intervention effects. Table 3 presents odds ratios and 95% confidence intervals (using robust standard errors) from logistic regression models predicting whether the effect size was positive (i.e., greater than zero).

We found no evidence that publication type, location source, language of report, funding source, time lag, or journal impact factor were associated with the odds of a positive effect size. The number of reports associated with a study ($OR = 1.22$, 95% CI [1.02, 1.47]) and the number of subsequent forward citations ($OR = 1.06$, 95% CI [1.02, 1.10]), however, were both positively associated with the odds of a positive effect size (Models VI and VII). The association between number of study reports and effect size direction was attenuated in the comprehensive model (IX), which was driven largely by the correlation between number

of study reports and subsequent citations. Indeed, each additional citation per year was associated with a 10% increase in the odds of a study reporting a positive effect size ($OR = 1.10$, 95% CI [1.02, 1.18]). Figure 3 depicts these results from model IX in Table 3, and shows the predicted probabilities (and corresponding 95% confidence intervals) of a study reporting a positive effect size, across varying levels of subsequent citations per year. Studies that were subsequently cited by other scholars were more likely to report a positive effect size relative to those that were cited less frequently.

Changes in Dissemination Biases over Time

The third objective of the study was to examine whether the association between reporting characteristics and effect size magnitude and/or direction has changed over time. To address this objective we re-estimated the models shown in Tables 2 and 3, but also included publication year and multiplicative interaction terms between publication year and each reporting characteristic. Overall, there was no consistent evidence of an interaction between publication year and each of the reporting characteristics (results not shown; available from authors upon request). The only exception to this was for time lag between intervention and publication, with results indicating a possible increase in time lag bias over time. This result was consistent in the model predicting effect size magnitude ($b_{Year} = 0.01$, $p_{Year} = .22$, 95% CI [−0.01, 0.02]; $b_{Lag} = -0.03$, $p_{Lag} = .02$, 95% CI [−0.05, −0.01]; $b_{YearXLag} = -0.01$, $p_{YearXLag} = .07$, 95% CI [−0.01, 0.00]) and the model predicting effect size direction ($OR_{Year} = 0.95$, $p_{Year} = .36$, 95% CI [0.86, 1.06]; $OR_{Lag} = 0.82$, $p_{Lag} = .08$, 95% CI [0.66, 1.03]; $OR_{YearXLag} = 0.94$, $p_{YearXLag} = .06$, 95% CI [0.87, 1.00]).

Figure 4 shows how time lag bias has changed over time, depicted as the association between time lag and the magnitude of effect sizes from 1995 to 2013. Results are shown as the relationship between predicted effect size and time lag between intervention and publication, at varying publication years (estimated in three-year increments since 1995). Recall that a negative slope reflects possible time lag bias, such that larger effect sizes are published more quickly than smaller effect sizes. As shown in Figure 4, the correlation between time lag and effect size magnitude has shifted in direction over time, with increased time lag bias in more recent years. The slopes are positive for studies published prior to 2004, indicating that studies with larger effect sizes took longer to get published. After 2004, however, this relationship shifted in direction (i.e., slopes of the lines became negative) and studies with larger effect sizes began to be published more quickly than those with smaller effect sizes.

Discussion and Conclusions

We conducted a retrospective analysis of data from a large systematic review and meta-analysis that examined the effectiveness of brief alcohol interventions for adolescents and young adults (Tanner-Smith & Lipsey, 2014). Using data from that meta-analysis, we examined whether the magnitude and/or direction of reported effects were associated with various reporting characteristics of the studies, and whether those associations changed over time. Overall, we found evidence of possible dissemination biases. Funded studies reported effects that were 0.14 standard deviations larger than those in unfunded studies. Results indicated that studies with larger effect sizes were published more rapidly, such that each

additional year a publication was delayed equated to a 0.03 standard deviation reduction in effect size. This time lag bias appears to have increased over the past decade. Finally, studies with larger and/or positive effects were also more likely to be cited by other scholars. We found no evidence, however, that the magnitude or direction of reported effect sizes was a function of other reporting characteristics such as location source, language of report, and journal impact factor.

As with any study, readers should interpret these results in light of the current study's strengths and limitations. The primary strengths of this study were its use of data from a large-scale meta-analysis based on a comprehensive literature search (including extensive grey literature searching) that synthesized findings from a large primary research literature in the social and behavioral sciences (i.e., brief alcohol interventions for adolescents and young adults), its simultaneous examination of several types of dissemination biases, and the exploration of potential changes in dissemination bias over time.

The main limitation of this study was its retrospective design. Ideally, studies of dissemination bias use a prospective research design whereby a sample of research studies are identified at study inception (e.g., during the stage of human subjects research approval) and followed prospectively over time, such that reported study findings could be compared to study protocols or proposals. Prospective designs, such as those used by Cooper, DeNeve, and Charlton (1997), are preferred because the theoretical population of potentially published studies is known at the outset. In contrast, retrospective designs (like the one used in the current study) make it impossible to know whether the final sample includes all eligible published and unpublished studies, no matter how comprehensive or extensive the literature search. It is impossible to know the findings of the truly "lost" studies. If we assume those lost studies are more likely to have null or negative findings, then findings from the current study might present a conservative picture of dissemination biases. Despite this limitation, our retrospective design did permit more in-depth analysis of several types of dissemination biases (e.g., time lag bias, multiple publication bias) that may not be possible to examine in prospective study designs due to shorter follow-up periods.

Another limitation of this study is the inability to determine whether the observed effects are truly indicative of dissemination bias and not some other underlying mechanisms. For instance, results indicated the possibility of time lag bias, such that studies with larger effect sizes were published more quickly. Although this could be caused by time lag bias, another plausible alternative could be that studies with larger effects are more likely to be published in journals with shorter peer review and editorial lags. For instance, journals in the health and medical sciences often have shorter peer review times than those in the social and behavioral sciences, so some of the observed time lag effects could be caused more by journal editorial practices than primary study authors' publication submission celerity. We were unable to empirically test this proposition in the current study, so this is an important direction for future research (particularly in interdisciplinary literatures such as addiction science).

Similarly, the finding that funded studies reported larger effect sizes might be caused by other mechanisms unrelated to dissemination bias. In the medical sciences, funding bias is

often tied to receipt of funding from pharmaceutical companies that have potential financial conflicts of interest. With alcohol intervention effectiveness research, receipt of funding from federal agencies (in this case, primarily the National Institutes of Health) is unlikely to yield financial conflicts of interest that would result in dissemination bias. Rather, a more plausible explanation is that funded studies may have larger sample sizes and be higher in overall quality (in terms of both conduct and reporting quality). We attempted to address this alternative explanation by statistically controlling for crude measures of study quality and sample size, but nonetheless still found evidence that funded studies yielded larger average effects. Notably, however, funded studies were no more likely to report positive effects, which suggest that larger sample sizes (and power to detect effects) are unlikely to explain the current findings. Other plausible alternatives may include (real or perceived) pressures for funded researchers to report larger effect sizes, in the hopes of obtaining future funding for follow-up studies. Our study did not have data to test these propositions, however, so clearly more research is needed to examine potential dissemination biases associated with external funding. This is particularly important for social and behavioral science research, where conflicts of interest (financial or otherwise) may not be as transparent as in the medical sciences.

This study did not examine selective outcome reporting bias, given our reliance on retrospective data. Selective outcome reporting is an important issue that deserves additional empirical examination, particularly in the social and behavioral sciences, but we were unable to examine it here. Another limitation of the current study is the potential range restriction for some of the reporting characteristics examined. As such, the true underlying relationships between some of the variables of interest (e.g., effect size magnitude and impact factor) may have been attenuated due to range restriction and thus the reported relationships may be conservative in nature.

Despite these limitations, evidence from this study supports prior research documenting that positive intervention or treatment effects appear more frequently in the literature than null or negative effects (e.g., Decullier et al., 2005; Dwan et al., 2008; Hopewell et al., 2009; Pigott et al., 2013). Such inflated estimates of treatment effects can be detrimental to the promotion of evidence-based practice – potentially leading to the overestimation of beneficial effects (or in the worst-case scenario, incorrectly attributing beneficial effects to harmful interventions). Moreover, this could lead to depriving individuals in need of interventions from other types of services that may be more effective, or at least more cost-effective.

What then, can be done to minimize dissemination biases in the psychological sciences? We believe that researchers and primary study authors, editors, peer reviewers, systematic reviewers, and meta-analysts alike have professional and ethical obligations to engage in research practices that seek to minimize biases. In terms of dissemination biases related to selective outcome reporting and selective publication, primary study authors should understand their ethical obligation to report fully all research findings, regardless of the direction or magnitude of those results. Prospective registration of research studies may be one way to promote complete outcome reporting, although recent research suggests this may not be enough (Dwan et al., 2011; Mathieu et al., 2009; Rasmussen et al., 2009; Redmond et al., 2013). Indeed, a larger cultural shift may be necessary, such that doctoral research

training programs sufficiently emphasize the importance of complete outcome reporting rather than statistical significance chasing (Schmidt, 1996). Such a shift will undoubtedly be difficult to implement given the publish or perish culture that implicitly encourages junior researchers to focus on the publication of new and innovative research findings, rather than the publication of null results or replication studies.

Systematic reviewers and meta-analysts, in turn, can attempt to minimize dissemination biases related to selective publication and selective inclusion by actively identifying and including grey or unpublished literature in reviews. An important question, of course, is whether including grey or unpublished literature in a meta-analysis might threaten the validity of the review findings, if those unpublished studies are of lower quality or higher risk of bias. Although some researchers have recently advocated for the systematic exclusion of dissertations or other unpublished grey literature from meta-analyses (Ferguson & Brannick, 2012), we agree with other scholars (Rothstein & Bushman, 2012) who contend that systematically excluding grey literature from meta-analyses will likely lead to exaggerated estimates of intervention effects (McAuley, Pham, Tugwell, & Moher, 2000; McLeod & Weisz, 2004). Indeed, if an unpublished or grey literature source meets all pre-specified eligibility criteria for a review or meta-analysis (including minimum quality or design thresholds), the onus is on the reviewer to include that study in the review, to code each study's method quality or risk of bias, and to assess whether such characteristics may influence or bias the results of the review. The issue of whether grey or unpublished literature is of lower quality than commercially published literature is an empirical question, and the answer is likely to vary across different disciplines and research topics. Fortunately, this question is one that systematic reviewers and meta-analysts are well poised to answer if they routinely collect and analyze publication type and method quality data. Of course, researchers should strive to conduct high quality studies, and we are not advocating that researchers, editors, or peer reviewers permit the literature to be flooded with low quality studies.

As recently noted by Francis (2012), another possible solution to minimize dissemination biases is for primary study authors to shift from the frequentist hypothesis testing framework to a Bayesian statistical framework. Bayesian hypothesis testing avoids the null hypothesis paradigm altogether, and can quantify the amount of evidence in support of both null and alternative hypotheses. Bayesian approaches can also avoid problems of data mining or fishing, and problems with multiple significance tests. However, even within the frequentist statistical approach, researchers can embrace the "new statistics" of psychology by avoiding null hypothesis significance testing in favor of effect size estimates and their confidence intervals (Cumming, 2014).

As systematic reviews and meta-analyses play an increasing role in summarizing the current best research evidence that informs evidence-based practice decisions, it is crucial for researchers and readers to be attentive to the issues of publication and dissemination bias. This is especially true in the social/behavioral sciences where publication bias has been researched less often relative to medicine (Rothstein et al., 2005). Readers of social and behavioral science meta-analyses may assume that a large number of included studies results in unbiased review conclusions. Yet bias in a meta-analysis may *increase* with the number

of studies included, which might occur if studies are selected from the published literature only. A comprehensive search, screen, and coding process are therefore paramount for minimizing potential biases.

We emphasize the comprehensiveness of the searching, coding, and analysis of meta-analyses for grey literature because we believe it is the only way to ensure valid reviews. As other social scientists have suggested, preventing publication bias in the first place may be the best approach to deal with it (Rothstein, 2008). At a minimum, systematic reviewers should be cognizant of such biases and attempt to code and test for the influence of variables potentially related to publication bias. Systematic reviewers and meta-analysts, however, must also be willing to conduct comprehensive literature searches to identify the grey literature and primary researchers should either prospectively register their studies or at least make a serious effort to report all study outcomes regardless of study findings.

Such efforts can only improve the validity of reviews that are used to inform evidence-based practice. As social scientists, we must continue to improve our awareness of these issues in a similar fashion to the medical research community. Efforts to decrease dissemination bias will suffer if they are not recognized by the community as important and apparent. The resulting efforts will deliver greater validity to reviews and will provide favorable future outcomes for policymakers, practitioners, and more importantly, the individuals we ultimately seek to help.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

APA Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. American Psychologist. 2006; 61:271–285. doi:10.1037/0003-066X.61.4.271. [PubMed: 16719673]

Bax L, Moons KG. Beyond publication bias. Journal of Clinical Epidemiology. 2011; 64:459–462. doi:10.1016/j.jclinepi.2010.09.003. [PubMed: 21194886]

Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA. 2002; 287:2847–2850. doi:10.1001/jama.287.21.2847. [PubMed: 12038930]

Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. JAMA. 1990; 263:1392–1395. doi:10.1001/jama.1990.03440100104016. [PubMed: 2406473]

Cooper H, DeNeve K, Charlton K. Finding the missing science: The fate of studies submitted for review by a human subjects committee. Psychological Methods. 1997; 2:447–452. doi: 10.1037//1082-989X.2.4.447.

Cooper H, VandenBos GR. Archives of Scientific Psychology: A new journal for a new era. Archives of Scientific Psychology. 2013; 1:1–6. doi:10.1037/arc0000001.

Cumming G. The new statistics: Why and how. Psychological Science. 2014; 25:7–29. doi: 10.1177/0956797613504966. [PubMed: 24220629]

Cunningham MR, Warme WJ, Schaad DC, Wolf FM, Leopold SS. Industry-funded positive studies not associated with better design or larger size. Clinical Orthopaedics and Related Research. 2006; 457:235–241. doi:10.1097/BLO.0b013e3180312057. [PubMed: 17195818]

De Oliveira GS, Chang R, Kendall MC, Fitzgerald PC, McCarthy RJ. Publication bias in the anesthesiology literature. Anesthesia & Analgesia. 2012; 114:1042–1048. doi:10.1213/ANE. 0b013e3182468fc6. [PubMed: 22344237]

Decullier E, Lhéritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: Retrospective cohort study. BMJ. 2005; 331:19. doi:10.1136/bmj.38488.385995.8F. [PubMed: 15967761]

Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. JAMA. 1992; 267:374–378. doi:10.1001/jama.1992.03480030052036. [PubMed: 1727960]

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLOS One. 2008; 3:e3081. doi:10.1371/journal.pone.0003081. [PubMed: 18769481]

Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR. Comparison of protocols and registry entries to published reports for randomised controlled trials. Cochrane Database of Systematic Reviews. 2011:1. doi:10.1002/14651858.MR000031.pub2.

Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR. Publication bias in clinical research. The Lancet. 1991; 337:867–872. doi:10.1016/0140-6736(91)90201-Y.

Egger M, Smith GD. Bias in location and selection of studies. BMJ. 1998; 316:61–66. doi:10.1136/bmj.316.7124.61. [PubMed: 9451274]

Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. The Lancet. 1997; 350:326–329. doi:10.1016/S0140-6736(97)02419-7.

Etter JF, Stapleton J. Citations to trials of nicotine replacement therapy were biased toward positive results and high-impact-factor journals. Journal of Clinical Epidemiology. 2009; 62:831–837. doi:10.1016/j.jclinepi.2008.09.015. [PubMed: 19128941]

Fanelli D. "Positive" results increase down the hierarchy of the sciences. PLOS One. 2010; 5:e10068. doi:10.1371/journal.pone.0010068. [PubMed: 20383332]

Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics. 2012; 90:891–904. doi:10.1007/s11192-011-0494-7.

Ferguson CJ, Brannick MT. Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. Psychological Methods. 2012; 17:120–128. doi:10.1037/a0024445. [PubMed: 21787082]

Ferguson CJ, Heene M. A vast graveyard of undead theories publication bias and psychological science's aversion to the null. Perspectives on Psychological Science. 2012; 7:555–561. doi:10.1177/1745691612459059. [PubMed: 26168112]

Francis G. Too good to be true: Publication bias in two prominent studies from experimental psychology. Psychonomic Bulletin & Review. 2012; 19:151–156. doi:10.3758/s13423-012-0227-9. [PubMed: 22351589]

Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: Is there a Tower of Babel bias? Journal of Clinical Epidemiology. 1995; 48:159–163. doi:10.1016/0895-4356(94)00098-B. [PubMed: 7853041]

Hall R, de Antueno C, Webber A. Publication bias in the medical literature: A review by a Canadian Research Ethics Board. Canadian Journal of Anesthesia. 2007; 54:380–388. doi:10.1007/BF03022661. [PubMed: 17470890]

Hedges LV. Estimation of effect sizes from a series of independent experiments. Psychological Bulletin. 1981; 92:490–499. doi:10.1037/0033-2909.92.2.490.

Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods. 2010; 1:39–65. doi:10.1002/jrsm.5. [PubMed: 26056092]

Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. Cochrane Database of Systematic Reviews. 2007; 2 Art. No.: MR000011. doi:10.1002/14651858.MR000011.pub2.

Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database of Systematic Reviews. 2009; 1 Art. No.: MR000006. doi:10.1002/14651858.MR000006.pub3.

Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA. 1998; 279:281–286. doi:10.1001/jama.279.4.281. [PubMed: 9450711]

Ioannidis JPA. Evolution and translation of research findings: From bench to where? PLOS Clinical Trials. 2006; 1:e36. doi:10.1371/journal.pctr.0010036. [PubMed: 17111044]

Jannot AS, Agoritsas T, Gayet-Ageron A, Perneger TV. Citation bias favoring statistically significant studies was present in medical research. Journal of Clinical Epidemiology. 2013; 66:296–301. doi: 10.1016/j.jclinepi.2012.09.015. [PubMed: 23347853]

John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychological Science. 2012; 23:524–532. doi: 10.1177/0956797611430953. [PubMed: 22508865]

Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. International Journal of Epidemiology. 2002; 31:115–123. doi:10.1093/ije/31.1.115. [PubMed: 11914306]

Liebeskind DS, Kidwell CS, Sayre JW, Saver JL. Evidence of publication bias in reporting acute stroke clinical trials. Neurology. 2006; 67:973–979. doi:10.1212/01.wnl.0000237331.16541.ac. [PubMed: 17000963]

Lipsey, MW. Identifying interesting variables and analysis opportunities. In: Cooper, H.; Hedges, LV.; Valentine, JC., editors. The handbook of research synthesis and meta-analysis. Russell Sage Foundation; New York, NY: 2009. p. 147-158.

Lipsey, MW.; Wilson, DB. Practical meta-analysis. Sage; Thousand Oaks, CA: 2001.

Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA. 2009; 302:977–984. doi:10.1001/jama. 2009.1242. [PubMed: 19724045]

McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? The Lancet. 2000; 356:1228–1231. doi: 10.1016/S0140-6736(00)02786-0.

McLeod BD, Weisz JR. Using dissertations to examine potential bias in child and adolescent clinical trials. Journal of Consulting and Clinical Psychology. 2004; 72:235–251. doi:10.1037/0022-006X. 72.2.235. [PubMed: 15065958]

Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. Journal of Clinical Epidemiology. 2007; 60:939–946. doi:10.1016/j.jclinepi.2006.11.014. [PubMed: 17689810]

Niemeyer H, Musch J, Pietrowsky R. Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for depression. Journal of Consulting and Clinical Psychology. 2013; 81:58–74. doi:10.1037/a0031152. [PubMed: 23244368]

Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, Pace B. Publication bias in editorial decision making. JAMA. 2002; 287:2825–2828. doi:10.1001/jama.287.21.2825. [PubMed: 12038924]

Parekh-Bhurke S, Kwok CS, Pang C, Hooper L, Loke YK, Ryder JJ, Song F. Uptake of methods to deal with publication bias in systematic reviews has increased over time, but there is still much scope for improvement. Journal of Clinical Epidemiology. 2011; 64:349–357. doi:10.1016/ j.jclinepi.2010.04.022. [PubMed: 20800992]

Pigott TD, Valentine JC, Polanin JR, Williams RT, Canada DD. Outcome-reporting bias in education research. Educational Researcher. 2013; 42:424–432. doi:10.3102/0013189X13507104.

Polanin JR, Tanner-Smith EE. Estimating the difference between published and unpublished effect sizes: A meta-review. Manuscript under review. 2014

Rasmussen N, Lee K, Bero L. Association of trial registration with the results and conclusions of published trials of new oncology drugs. Trials. 2009; 10:116. doi:10.1186/1745-6215-10-116. [PubMed: 20015404]

Raudenbush, SW. Analyzing effect sizes: Random-effects models. In: Cooper, H.; Hedges, LV.; Valentine, JC., editors. The handbook of research synthesis and meta-analysis. Russell Sage Foundation; New York, NY: 2009. p. 295-315.

Redmond S, von Elm E, Blümle A, Gengler M, Gsponer T, Egger M. Cohort study of trials submitted to ethics committee identified discrepant reporting of outcomes in publications. Journal of Clinical Epidemiology. 2013; 66:1367–1375. doi:10.1016/j.jclinepi.2013.06.020. [PubMed: 24075598]

Rosenthal R. The file drawer problem and tolerance for null results. Psychological Bulletin. 1979; 86:638–641. doi:10.1037/0033-2909.86.3.638.

Rothstein HR. Publication bias as a threat to the validity of meta-analytic results. Journal of Experimental Criminology. 2008; 4:61–81. doi:10.1007/s11292-007-9046-9.

Rothstein HR, Bushman BJ. Publication bias in psychological science: Comment on Ferguson and Brannick (2012). Psychological Methods. 2012; 17:129–136. doi:1010.1037/a0027128. [PubMed: 22369520]

Rothstein, HR.; Sutton, AJ.; Borenstein, M., editors. Publication bias in meta-analysis: Prevention, assessment and adjustments. John Wiley & Sons; Chichester, UK: 2005.

Sackett DL, Rosenberg WM, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: What it is and what it isn't. BMJ. 1996; 312:71–72. doi:10.1136/bmj.312.7023.71. [PubMed: 8555924]

Sánchez-Meca J, Marín-Martínez F, Chacón-Moscoso S. Effect-size indices for dichotomized outcomes in meta-analysis. Psychological Methods. 2003; 8:448–467. doi:10.1037/1082-989X. 8.4.448. [PubMed: 14664682]

Schmidt FL. Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. Psychological Methods. 1996; 1:115–129. doi:10.1037//1082-989X. 1.2.115.

Stern JM, Simes RJ. Publication bias: Evidence of delayed publication in a cohort study of clinical research projects. BMJ. 1997; 315:640–645. doi:10.1136/bmj.315.7109.640. [PubMed: 9310565]

Sterne, JA.; Egger, M.; Moher, D. Addressing reporting biases. In: Higgins, JPT.; Green, S., editors. Cochrane handbook for systematic reviews of interventions: Cochrane book series. John Wiley & Sons; Chichester, UK: 2008. p. 297-333.doi:10.1002/9780470712184.ch10

Tanner-Smith EE, Lipsey MW. Brief intervention effects on adolescent and young adults' alcohol consumption: A systematic review and meta-analysis. Manuscript under review. 2014

Timmer A, Hilsden RJ, Cole J, Hailey D, Sutherland LR. Publication bias in gastroenterological research–a retrospective cohort study based on abstracts submitted to a scientific meeting. BMC Medical Research Methodology. 2002; 2:7. doi:10.1186/1471-2288-2-7. [PubMed: 11978183]

Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: A case study. BMJ. 1997; 315:635–640. doi:10.1136/bmj.315.7109.635. [PubMed: 9310564]

Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000; 56:645–646. doi:10.1111/j.0006-341X.2000.00645.x. [PubMed: 10877330]
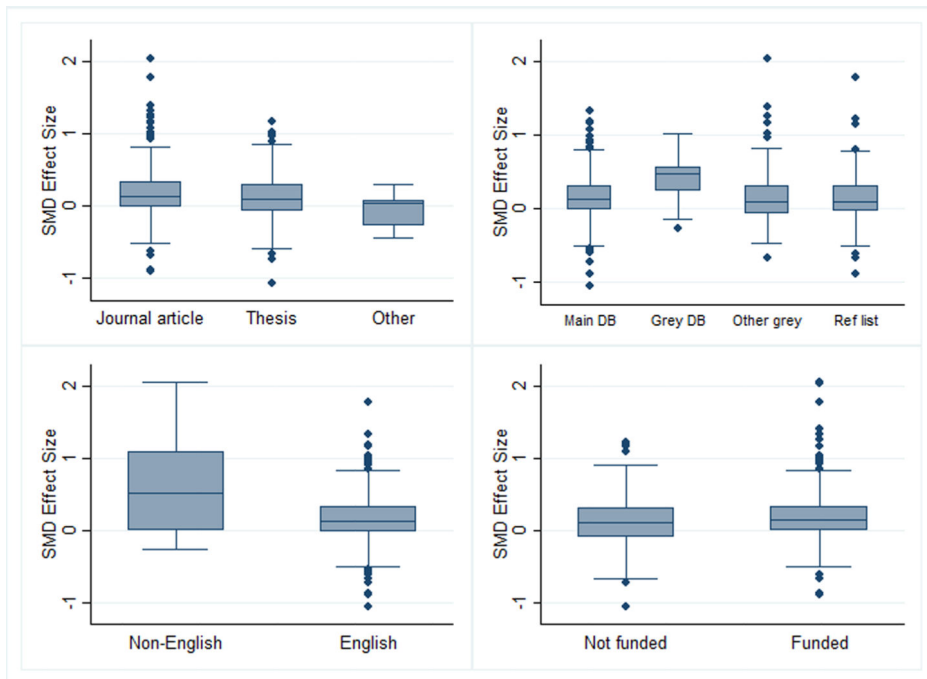
Figure 1.
Distribution of effect sizes by publication type, identification source, language of report, and funding status of study.

SMD = Standardized mean difference. *DB* = database. *Ref list* = reference list.
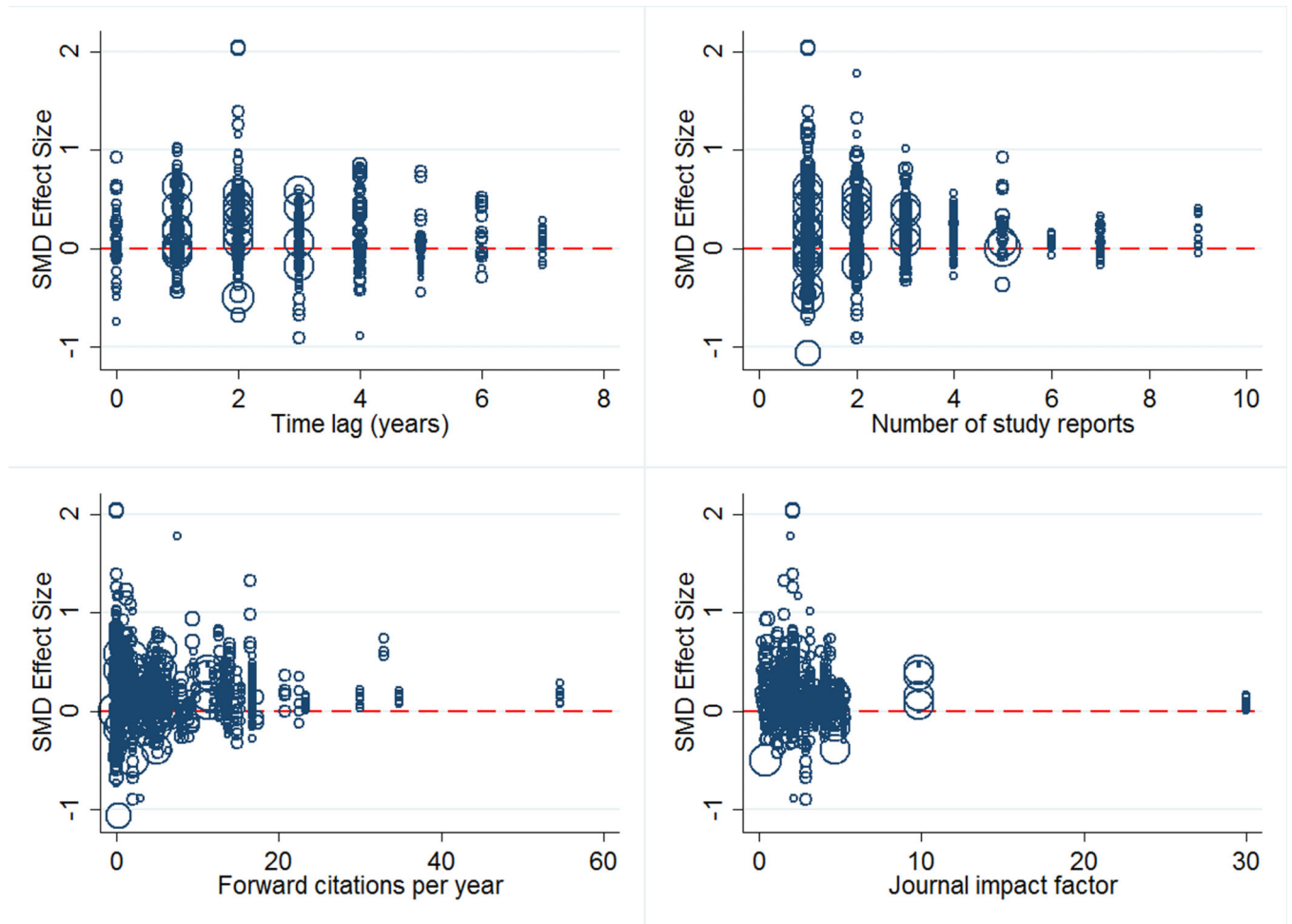
Figure 2.
Associations between effect sizes, time lag, number of study reports, forward citations, and journal impact factor.

SMD = Standardized mean difference.
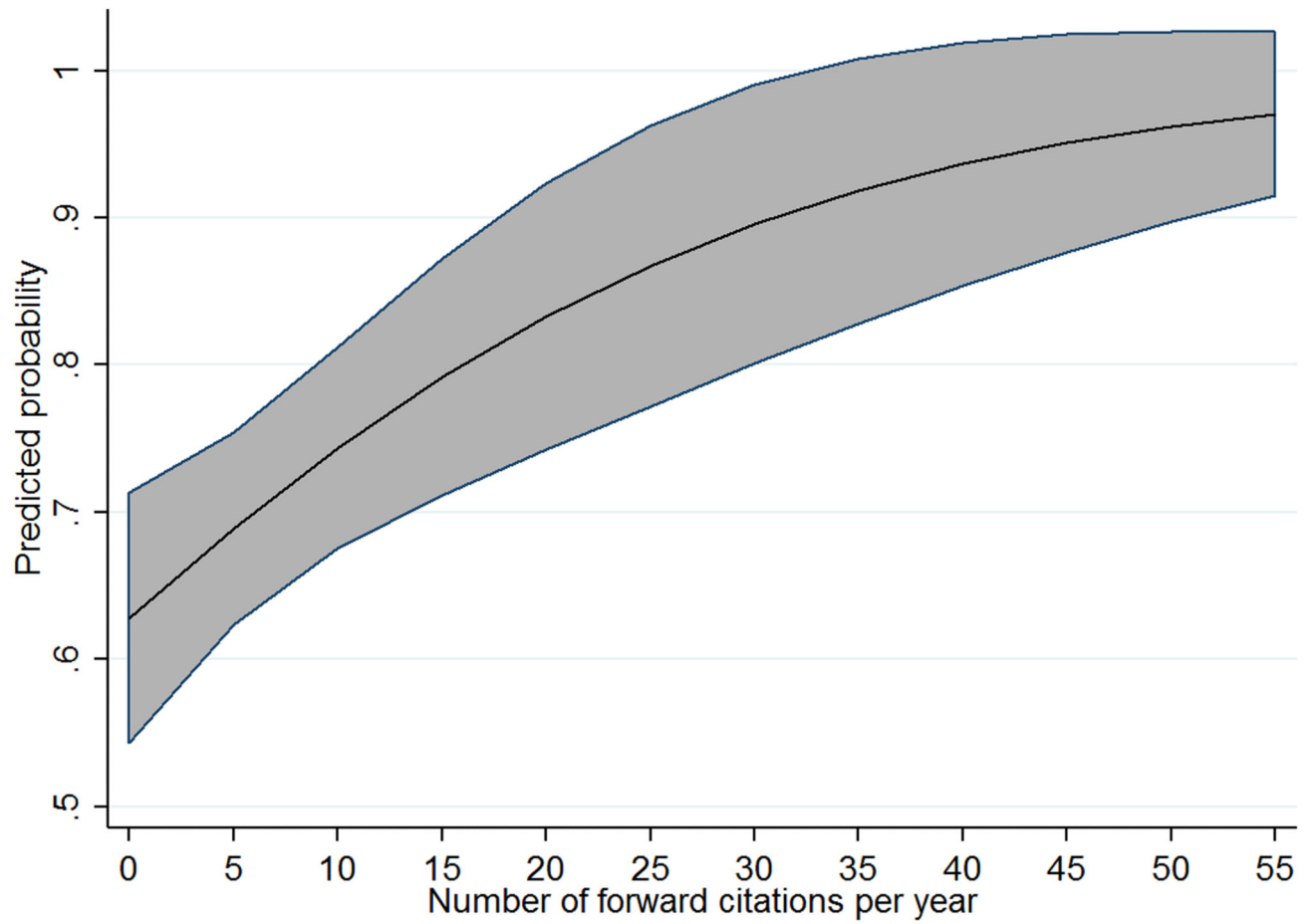
Figure 3.
Predicted marginal probabilities and corresponding 95% confidence intervals for the
probability of reporting a positive effect size, by the number of subsequent forward citations
per year.

Figure 4.
Predicted marginal effect size and time lag between intervention and publication, by publication year.

Table 1

Bivariate Correlations and Descriptive Statistics of Study and Reporting Characteristics

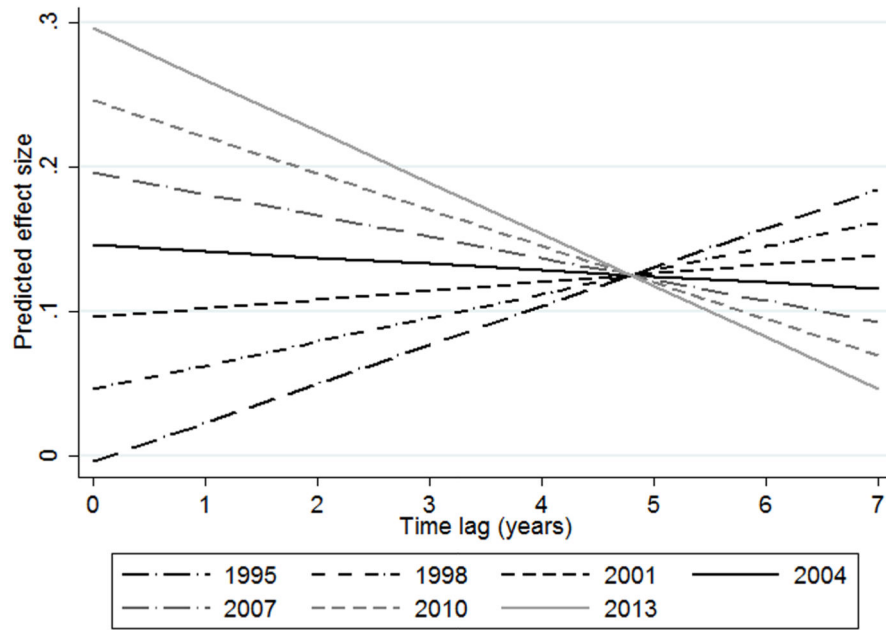|	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Publication year [a]	1.0														
2. Randomized trial (1 = yes) [b]	-.02	1.0													
3. Intent-to-treat analysis (1 = yes) [b]	.00	.13	1.0												
4. Attrition rate [b]	.08	.07	-.23	1.0											
5. Number of study waves [b]	.11	.11	.06	-.15	1.0										
6. Adolescent sample (1 = yes) [b]	-.17	-.09	.04	-.16	-.08	1.0									
7. Journal article (1 = yes) [a]	.13	.07	.08	-.23	.14	.07	1.0								
8. Main electronic database (1 = yes) [a]	-.22	.18	-.10	.10	.03	-.11	.14	1.0							
9. Reference list (1 = yes) [a]	-.07	-.13	.00	-.08	-.09	.17	.01	-.62	1.0						
10. English report (1 = yes) [a]	-.04	.44	-.07	.12	.06	-.20	.03	.19	-.05	1.0					
11. Funded study (1 = yes) [b]	.12	.13	.26	-.21	.22	.09	.42	.05	-.12	.18	1.0				
12. Time lag (years) [b]	.01	.05	.22	-.05	.14	-.11	.33	.11	.02	-.01	.24	1.0			
13. Number of study reports [b]	-.01	.09	.20	-.20	.14	.16	.36	.19	-.04	.05	.26	-.02	1.0		
14. Number of citations per year [a]	-.22	.16	.27	-.16	.22	-.06	.38	.29	-.18	.11	.34	.32	.41	1.0	
15. Journal impact factor	.23	.12	.34	-.16	.02	.21	-.08	.03	-.08	.10	.20	-.09	.35	.25	1.0
Mean	2007	.89	.39	.22	1.84	.13	.75	.76	.11	.97	.69	2.64	1.72	5.62	2.99
SD	4	0.32	0.49	0.21	2.47	0.34	0.43	0.43	0.31	0.17	0.46	1.46	1.24	7.71	2.92

Notes.

[a] Variable measured at the effect size level (k = 1,439).

[b] Variable measured at the study level (n = 179). SD = standard deviation; k = number of effect sizes; n = number of studies. All correlations |0.06| are significantly different from zero at $p < .05$.

Table 2

Unstandardized Coefficients and 95% Confidence Intervals from Meta-Regression Models Predicting Effect Size Magnitude: Main Effects of Reporting Characteristics

Model	I	II	III	IV	V	VI	VII	VIII	IX
Reporting characteristics									
Journal article	0.10* [0.01, 0.19]								0.05 [−0.06, 0.17]
Main database		−0.01 [−0.13, 0.12]							−0.11 [−0.26, 0.04]
Reference list		.00 [−0.14, 0.15]							−0.05 [−0.23, 0.12]
English report			−0.17 [−0.48, 0.13]						−0.38 [−1.02, 0.26]
Funded study				0.08* [0.01, 0.16]					0.14* [0.04, 0.23]
Time lag (years)					−0.02 [−0.05, 0.00]				−0.03* [−0.05, −0.01]
Number of study reports						0.01 [−0.01, 0.02]			0.01 [−0.02, 0.03]
Forward citations							0.00 [−0.00, 0.01]		0.01* [+0.00, 0.01]
Impact factor								−0.01 [−0.02, 0.00]	–
Control variables									
Adolescent sample	0.00 [−0.13, 0.13]	0.00 [−0.14, 0.13]	−0.01 [−0.14, 0.11]	−0.00 [−0.14, 0.13]	−0.00 [−0.14, 0.13]	−0.01 [−0.14, 0.13]	0.01 [−0.13, 0.15]	−0.01 [−0.17, 0.15]	−0.12* [−0.24, −0.01]
Randomized design	−0.02 [−0.20, 0.16]	0.01 [−0.19, 0.22]	0.04 [−0.17, 0.24]	−0.01 [−0.20, 0.17]	−0.01 [−0.20, 0.17]	0.01 [−0.18, 0.19]	0.01 [−0.18, 0.20]	−0.19 [−0.50, 0.13]	0.14 [−0.03, 0.31]
Intent-to-treat design	0.00 [−0.08, 0.08]	0.00 [−0.09, 0.09]	0.01 [−0.07, 0.08]	−0.01 [−0.08, 0.07]	−0.01 [−0.08, 0.07]	0.00 [−0.08, 0.08]	−0.00 [−0.08, 0.08]	−0.01 [−0.08, 0.07]	−0.09 [−0.19, 0.01]
Attrition rate	−0.09 [−0.24, 0.07]	−0.11 [−0.27, 0.04]	−0.10 [−0.25, 0.06]	−0.09 [−0.24, 0.07]	−0.09 [−0.24, 0.07]	−0.11 [−0.26, 0.04]	−0.12 [−0.28, 0.03]	−0.13 [−0.31, 0.05]	−0.04 [−0.29, 0.20]
Number of study waves	−0.01* [−0.01, −0.00]	−0.01* [−0.01, −0.00]	−0.01* [−0.01, −0.00]	−0.01* [−0.01, −0.00]	−0.01* [−0.01, −0.00]	−0.01* [−0.01, −0.00]	−0.01 [−0.04, 0.01]	−0.00* [−0.01, −0.00]	−0.03 [−0.07, 0.01]

Model	I	II	III	IV	V	VI	VII	VIII	IX
k	1,439	1,439	1,439	1,439	983	1,439	1,414	973	961
n	179	179	179	179	98	179	173	119	93

Note. Results estimated from mixed-effect meta-regression models with robust standard errors. k = number of effect sizes; n = number of studies.

*
$p < .05$.

Table 3

Odds Ratios and 95% Confidence Intervals from Logistic Regression Models Predicting Whether Effect Size was Positive: Main Effects of Reporting Characteristics

Model	I	II	III	IV	V	VI	VII	VIII	IX
Reporting characteristics									
Journal article	1.49 [0.78, 2.84]								0.65 [0.26, 1.61]
Main database		0.96 [0.42, 2.16]							0.64 [0.22, 1.91]
Reference list		0.84 [0.35, 2.01]							0.57 [0.19, 1.71]
English report			0.70 [0.20, 2.46]						0.90 [0.19, 4.19]
Funded study				1.42 [0.81, 2.49]					1.20 [0.55, 2.66]
Time lag (years)					0.85 [0.64, 1.12]				0.90 [0.72, 1.12]
Number of study reports						1.22[*] [1.02, 1.47]			1.09 [0.86, 1.37]
Forward citations							1.06[*] [1.02, 1.10]		1.10[*] [1.02, 1.18]
Impact factor								1.04 [0.94, 1.15]	--
Control variables									
Adolescent sample	0.80 [0.41, 1.59]	0.84 [0.45, 1.58]	0.80 [0.40, 1.60]	0.79 [0.39, 1.58]	0.64 [0.30, 1.35]	0.73 [0.37, 1.45]	0.94 [0.48, 1.81]	0.79 [0.32, 1.96]	0.69 [0.34, 1.38]
Randomized design	0.98 [0.44, 2.16]	1.02 [0.45, 2.33]	1.16 [0.51, 2.60]	0.96 [0.42, 2.23]	1.81 [0.78, 4.23]	0.94 [0.42, 2.12]	0.85 [0.37, 1.94]	0.70 [0.19, 2.62]	1.80 [0.57, 5.72]
Intent-to-treat design	1.14 [0.65, 2.00]	1.13 [0.64, 2.01]	1.15 [0.65, 2.01]	1.06 [0.60, 1.88]	1.17 [0.64, 2.15]	1.06 [0.62, 1.84]	0.97 [0.55, 1.72]	.64 [0.34, 1.21]	0.93 [0.48, 1.81]
Attrition rate	1.87 [0.39, 8.92]	1.50 [0.30, 7.53]	1.56 [0.32, 7.68]	1.69 [0.36, 7.82]	2.99 [0.34, 26.58]	1.85 [0.38, 8.87]	1.66 [0.38, 7.21]	2.01 [0.20, 19.95]	3.24 [0.42, 24.98]
Number of study waves	1.01 [0.94, 1.08]	1.01 [0.94, 1.09]	1.02 [0.94, 1.09]	1.00 [0.94, 1.07]	1.00 [0.94, 1.05]	1.00 [0.94, 1.07]	1.00 [0.86, 1.15]	1.03 [0.97, 1.10]	0.70 [0.44, 1.11]
k	1,439	1,439	1,439	1,439	983	1,439	1,414	973	961
n	179	179	179	179	98	179	173	119	93

Note. Results estimated from mixed-effect logistic regression models with clustered robust standard errors. *k* = number of effect sizes; *n* = number of studies.

[*] *p* < .05.