Human Communication Research ISSN 0360–3989

SPECIAL ISSUE ARTICLE

Unresolved Heterogeneity in Meta-Analysis: Combined Construct Invalidity, Confounding, and Other Challenges to Understanding Mean Effect Sizes

Timothy R. Levine[1] & René Weber[2]

1 Department of Communication Studies, The University of Alabama at Birmingham, Birmingham, AL 35294, USA

2 Department of Communication, Media Neuroscience Lab, University of California Santa Barbara, Santa Barbara, CA 93106, USA

We examined the interplay between how communication researchers use meta-analyses to make claims and the prevalence, causes, and implications of unresolved heterogeneous findings. Heterogeneous findings can result from substantive moderators, methodological artifacts, and combined construct invalidity. An informal content analysis of meta-analyses published in four elite communication journals revealed that unresolved between-study effect heterogeneity was ubiquitous. Communication researchers mainly focus on computing mean effect sizes, to the exclusion of how effect sizes in primary studies are distributed and of what might be driving effect size distributions. We offer four recommendations for future meta-analyses. Researchers are advised to be more diligent and sophisticated in testing for heterogeneity. We encourage greater description of how effects are distributed, coupled with greater reliance on graphical displays. We council greater recognition of combined construct invalidity and advocate for content expertise. Finally, we endorse greater awareness and improved tests for publication bias and questionable research practices.

Keywords: Meta-Analysis, Effect Heterogeneity, Effect Sizes, Construct Validity

doi:10.1093/hcr/hqz019

The meta-analysis is a valuable tool for cumulating quantitative results—most often effect sizes—across studies. Heterogeneity of effects exists when effects vary from primary study to primary study more than would be expected by sampling error alone. When unresolved, heterogeneity reduces confidence in the mean effect size as an estimate of the population effect. The greater the amount of unresolved heterogeneity, the greater the cause for concern.

Corresponding author: René Weber; e-mail: renew@ucsb.edu

Van Erp, Verhagen, Grasman, and Wagenmakers (2017) created a database containing 705 between-study heterogeneity estimates in meta-analyses published in *Psychological Bulletin*. Our analyses of the Van Erp et al. database shows that substantial heterogeneity is evident in approximately four out of five meta-analyses. Consequently, Van Erp et al.'s data document that between-study effect heterogeneity is the norm rather than the exception in psychology (for similar studies, see Rhodes, Turner, & Higgins, 2015; Turner, Davey, Clarke, Thompson, & Higgins, 2012). The same conclusion may hold for most meta-analyses in communication. Our concern is not the mere ubiquity of heterogeneity, but that between-study effect heterogeneity is seldom fully resolved or explained.

Heterogeneity of effects can be produced by at least three broad classes of mechanisms. First and perhaps most widely recognized, effects are often contingent on other constructs. We will call this category substantive moderators. Constructs that are substantively different are ones that have non-isomorphic conceptual definitions. Conceptually or theoretically, they are qualitatively different types of things. Substantive moderators are when the effects of one construct (x1) on another construct (y) varies as a function of one or more additional constructs (x2, x3, and so forth).

Substantive moderators are usually tested by coding primary studies for certain theoretically relevant features and cumulating results separately within levels of those features. For example, Feeley, Anker, and Aloe (2012) examined 10 substantive moderators of the effectiveness of the door-in-the-face compliance strategy. Feeley et al. report (as have others) that the door-in-the-face strategy requires that the same person make both the large and subsequent smaller requests.

As a second example, Reimer, Reimer, and Czienskowski (2010) examined the frequency in which groups discussed shared and unshared information in hidden profile experiments. Substantive moderators resolve heterogeneity when effect sizes vary systematically by level of the moderator and when within-level effect sizes are no longer heterogeneous. Reimer et al. found that two substantive moderators—discussion time and decision alternatives—accounted for much of the heterogeneity. Effects were considerably weaker with two alternatives and less than 30 minutes of discussion time than other combinations of these variables. While findings remained heterogeneous when there were three alternatives coupled with less than 30 minutes of discussion time, the findings were homogenous in the other conditions.

A second category of moderators that can lead to heterogeneity of effects are various methodological artifacts. Poor measurement reliability, weak experimental manipulations, and restrictions in range can all lower effect sizes (Baugh, 2002; Hunter & Schmidt, 1990, 1994). If, for example, primary studies use measures that vary in reliability, effects will vary as a result. Other examples of method-induced heterogeneity can be differences in research design (e.g., between- or within-groups designs), questionable research practices (Vermeulen & Hartman, 2015), publication bias (e.g., Thornton & Lee, 2000), or when meta-analysts combined different types of effect sizes (e.g., the synthesized effect includes Pearson's zero-order correlations aside correlations transformed from frequency tables; see Hunter & Schmidt, 1990, 1994). Methodological artifact moderators can resolve heterogeneity when effects, once corrected for artifacts, are no longer heterogeneous.

Combined construct invalidity (Hunter & Schmidt, 1994) is a third potential cause of heterogeneous effects, which is under-recognized and under-studied in communication meta-analyses but is especially pernicious. Similar to Carpenter (2020), we believe the aggregation of apples and oranges is problematic. Are all primary findings being averaged in meta-analyses indeed of the same kind? We argue here that this issue can be best understood as an issue of confounding. When substantively different effects are averaged to form one mean, that mean can be said to be confounded and, as a consequence, the mean effect can be said to exhibit combined construct invalidity. If confounded constructs exhibit different effects, then confounding produces heterogeneity that adds to existing heterogeneity beyond that due to substantive moderators or methodological artifacts.

A confound occurs when there is more than one type of thing that are mashed together. It is common to think of confounds as a concern in experimental inductions (researcher-manipulated independent variables), where a nuisance variable varies concomitantly with the treatment and controls. An experiment obsessively looking at the effects of x on y compares a treatment involving both x and z with a control involving neither x nor z, thus confounding x and z. But importantly, and frequently ignored, measured variables and outcome variables can be confounded as well. For instance, scale items that tap more than one latent variable can be said to be confounded. Similarly, average effect sizes can be confounded if the effects are derived from different variables. What all these types of confounds have in common is that they make the interpretation of findings ambiguous. When recognized, they create rival explanations that cannot be empirically parsed. When they go undetected, conclusions about findings can be misleading.

For instance, when we meta-analyze the relationship between some message feature (e.g., message sensation value) and some message outcome (e.g., behavioral intentions), the effect sizes being averaged all need to be based on assessments of the same message feature and the same outcome. It would be invalid, for example, to combine effect sizes from assessments of behavioral intentions and actual behavior, as we know that intentions are not the same thing as behaviors. Similarly, combining effect sizes that are based on alternative assessments of message sensation value (e.g., information introduced; I^2) would be problematic. Further, the empirical finding that a message feature has similar effects on both intentions and behaviors does not mean that intentions and behaviors are the same construct (cf., O'Keefe, 2013). While combined construct invalidity can create heterogeneous effects, homogeneity of effects does not preclude conceptual confounding.

Combined construct invalidity is an especially slippery issue in meta-analyses, because it is a matter of critical thinking and topic expertise rather than something subject to mere statistical testing. Carpenter (2020), for instance, argues that in order to detect combined construct invalidity it is necessary to evaluate whether a construct possesses the necessary level of concreteness. If a construct lacks concreteness—that is, is abstract and thereby covers an entire class of constructs—then it is potentially invalid and a poor candidate for meta-analytical consideration. Obviously, this judgment is based on theoretical and, to some extent, measurement considerations. There cannot be a statistical test of construct concreteness. Thus, absent a generalizable strategy to determine the concreteness of a construct, it is possible that decisions regarding when or when not to summarize effect sizes in meta-analyses are subject to substantial researcher degrees of freedom, with effect heterogeneity as a likely result.

An example of combined construct invalidity and confounded effect sizes can be seen in the Vrij, Fisher, and Blank (2017) meta-analysis examining the effectiveness of the cognitive approach to lie detection (for a detailed critique, see Levine, Blair, & Carpenter, 2018). The cognitive approach to lie detection includes at least two conceptually different strategies to improve lie detection: imposing a cognitive load on senders and encouraging senders to say more. The meta-analysis also included two conceptually different forms of outcome assessments: using statistical algorithms to classify truths and lies based on coded sender behaviors, and human receivers making truth-lie assessments scored for accuracy. Meta-analyzing the aggregate effect of imposing a cognitive load on senders plus encouraging senders to say more on statistical modeling plus human judgments leads to a confounded mean effect size. It is important to note here that statistical tests indicating effect homogeneity do not inform the issue at hand. The issues are whether adding a cognitive load is conceptually the same as encouraging an interviewee to say more, and whether human lie detection accuracy is the same as a statistical analysis of interviewee behavioral cues.

A content analysis of current practice

To provide evidence of our assertions and to get a feel for current practices in communication research, we examined a sample of meta-analyses published in four leading communication journals (*Communication Monographs*, *Communication Research*, *Human Communication Research*, and *Journal of Communication*) dating back 10 years to 2008 and extending to 2015. The idea was that if heterogeneity is prevalent and frequently unresolved, then examples should be evident even in a small sample. Further, if unresolved heterogeneity was commonplace in our best journals, then readers might benefit from our four suggestions provided at the end of our essay. That is, we seek a small demonstration that the concerns we articulate have a basis in recent published work and that practice can be improved.

Selection criteria and coding

The rationale of our selection criteria was straightforward: the time frame was selected to capture the current state of the art while still having a large enough sample to separate common practice from an idiosyncratic instance. We limited the sample to A-level journals so that the works we examined passed not only peer review but a relatively high bar for importance and quality. Meta-analyses were selected using a database of 149 meta-analyses of communication phenomena compiled by Rains, Levine, and Weber (2018).

There were 14 meta-analyses that met the criteria and were included (for a list, see the Supporting Information). The selected meta-analyses were read, and the following coding system was created based on the goals of this article. One of the authors did the coding. Three research purposes (population estimates, significance testing, and moderator searches) were coded as present or absent. The reporting of tests of between-study effect heterogeneity, the results of the heterogeneity tests, and whether the heterogeneity was resolved with moderator analyses were noted. Whether fixed-effects or random-effect analyses, correction for measurement artifacts, and tests of publication bias were conducted were also coded as present or absent. Finally, the meta-analyses were examined for how the distribution of effects were described or graphed, if at all.

Results

The meta-analyses we examined focused primarily on three goals: (1) estimating population effect sizes; (2) testing mean effects against a nil/null hypothesis for statistical significance; and (3) searching for moderators (i.e., conditions under which effect sizes vary). All the meta-analyses in our sample calculated (and typically weighted) mean effect sizes. Of the 14, 10 tested the mean effect sizes against a nil/null hypotheses of no effect, while 12 looked for moderators.

Most (64%) of the meta-analyses we examined tested for heterogeneity. Of those meta-analyses testing for heterogeneity, all found it. Importantly, most (78%) meta-analyses were unable to fully resolve the heterogeneity with moderator searches or artifact corrections. Although searching for moderators was ubiquitous in the meta-analyses we examined, descriptions of the distributions of effect sizes were typically limited to central tendency, dispersion, and alphabetical listings of effects. Only two meta-analyses provided forest plots. No funnel plots, histograms, or stem-and-leaf plots were included. In short, there was very little in the way of describing or graphing the shapes of distributions. Graphing is critical because different shapes of distributions may reveal the potential causes of heterogeneity (e.g., a bimodal distribution might suggest a single, critical moderator). We also observed that testing for publication bias has not yet become normative.

We graphed five sets of effect sizes from the meta-analyses. We initially selected distributions from meta-analyses that provided tables of effects sizes from primary studies and that involved relatively more effects. The five initially selected sets of effect sizes were sufficient to exemplify the importance of how effects are distributed (see Figure 1).

In one case (O'Keefe & Jensen, 2009), the distribution of effects was symmetrical and approximately normal (Shapiro-Wilk = .98; $p < .575$). With most primary studies tightly clustered around the mean effect, in this case the interpretation of

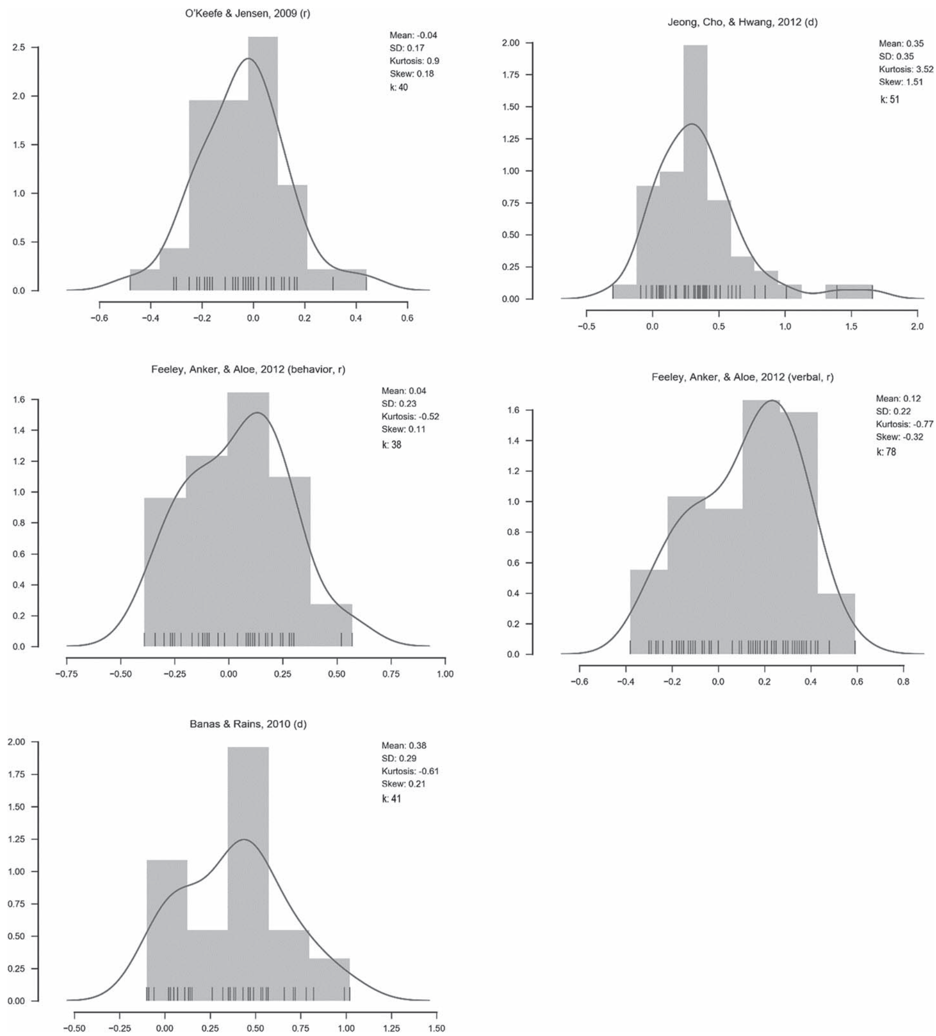

Figure 1 Histograms of primary effects from five recent communication meta-analyses.

the mean effect is straightforward because, in a distribution that is approximately normal (symmetrical, unimodal, bell-shaped), the mean is a valid representation of the distribution's central tendency. A second distribution (Jeong, Cho, & Hwang, 2012) had a substantial, positive skew, which likely suggests publication bias and an inflated mean effect size. A third distribution (Banas & Rains, 2010) was clearly bimodal, suggesting an unknown moderator. The remaining two distributions were somewhere in between these three extremes.

Our reviewers correctly noted that we have omitted four additional meta-analyses published in our time frame: Rains, Peterson, and Wright (2015); Sun, Pan, and Shen (2008); Wright, Tokunaga, and Kraus (2015); and Yang, Aloe, and Feeley (2014).

Rains et al. (2015) aggregates counts in content analyses rather than more typical effect sizes: the arguments we make in this manuscript do not apply. Wright et al. (2015) does not report heterogeneity tests, and Yang et al. (2014) mention the shape of effect distributions and the heterogeneity of effects, but the heterogeneous findings remained unresolved. Sun et al. (2008) focused on mean effect sizes and found heterogeneous effects. Moderators resolved some, but not all heterogeneity. Taken together, these four additional meta-analyses reflect exactly the issues we report from the included studies.

Suggestions for future practice

Based on our arguments and results, we encourage four improvements for future meta-analytical practice in communication research. First, it should become standard practice that all communication meta-analyses consider between-study heterogeneity of effects. Heterogeneity should not only be reported as merely present or absent, but instead as a matter of degree.

Researchers should use widely accepted measures of effect variation, combined with the appropriate statistical tests, such as Higgin's I^2 (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003), which, in contrast to Cochran's Q, allows researchers both to quantify and test the degree of heterogeneity (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). If researchers provide statistical evidence for the absence of effect heterogeneity, then simply referring to a non-significant result at an alpha level of 5% is insufficient because of the risk of a Type II error. This is especially true when the number of primary studies in a meta-analysis is small. More appropriate alpha levels (maximum tolerable Type I errors in statistical decisions against a nil/null hypothesis) are 10% or even 20% (Higgins & Thompson, 2002). Better still is the application of an equivalence test approach (e.g., Weber & Popova, 2012), in which previously reported estimates of effect size variation across studies (e.g., Van Erp et al., 2017) are used as parameters of the H1 distribution for between-study heterogeneity (for a Bayesian approach, see Gronau, Van Erp, Heck, Cesario, Jonas, & Wagenmakers, 2017; Kpekpena & Muthukumarana, 2018).

Second, we suggest that meta-analyses report and preferably graph the distributions of primary effects. Van Erp et al. (2017, p. 2) report that most (194 out of 255) published meta-analyses did not provide an estimate of between-study effect size variation and many failed to provide effect size distributions. This is an unfortunate practice. Creating a stem-and-leaf plot of primary effects is simple and highly informative. Additionally, we strongly encourage adding forest plots (e.g., Sedgwick, 2015; for available software, see Gordon, 2019), which are especially informative, to all future communication meta-analyses. In addition to reporting mean effect sizes and their variations (first and second moment), we endorse including information on further distribution moments, such as skewness and kurtosis. If the distribution of primary effects in plots is highly skewed, unusually platykurtic, or if there is more than one mode, this should be noted and discussed in the context of between-study effect size heterogeneity. As distributions depart from symmetrical, bell-shaped distributions, the utility of the mean effect as a representation of the literature may be diminished. If primary studies' effect sizes are normally distributed around the mean effect size, then this should also be noted and offered as evidence that the mean effect size is likely (but not certainly, see below) an informative indication of central tendency and a valid estimate of a true population parameter. If research does provide specific information on between-study effect size heterogeneity, such as τ (tau), then it is important for the researchers to explain which estimator for τ or τ^2 they have used, as different estimators are available in the literature and in meta-analysis software (e.g., DerSimonian-Laird; maximum likelihood; Hunter-Schmidt), and different estimators can lead to different and even conflicting test results for between-study heterogeneity (Viechtbauer, 2005).

Third, along with Carpenter (2019), we encourage greater awareness of the apples-to-oranges, confounded effects, and combined construct invalidity concerns in meta-analyses. We emphasize that while combined construct invalidity can create heterogeneous effects, homogeneity of effects does not preclude combined construct invalidity. For a purposely extreme example, Wilson, Norris, Shi, and Rack (2010) report that non-maltreated children engage in more positive behaviors with parents, relative to maltreated children (Cohen's d = .42–.45). These mean effect sizes are quite similar to the mean effect size (d = .42) of various lie detection strategies in Vrij et al. (2017). We hope it is sufficiently obvious that the similarity in effect sizes across these vastly different literatures would in no way justify pooling the Wilson et al. children's data with the Vrij et al. deception research to create an average effect size. Such an aggregate would make no substantive sense, and no finding of homogeneity would change that.

As a solution to combined construct invalidity, we advocate that meta-analyses with the intention to specify a population parameter in a particular research field must always include a content expert. A content expert is someone well versed in the topic of the meta-analysis, which includes the range of methodological practices and the relevant theory. Content experts in lie detection would know, for example, that judgments of human message receivers are conceptually distinct from statistical models of sender behavior in relevant theory (because senders and receivers play different roles in communication; see Levine et al., 2018). At the very least, if meta-analysts are unable to document topic expertise in the form of previously peer-reviewed research on the topic, then their meta-analysis article needs to include a comprehensive discussion of all relevant methodological practices and theoretical premises (perhaps even theoretical controversies), so that these discussions may allow for the necessary scrutiny from reviewers who may be expert in the topic. We do not mean to suggest that only the most published scholars in a research area are from now on "allowed" to produce meta-analyses. However, based on our claim that combined construct invalidity goes mostly undetected and requires content expertise, editors and reviewers should be wary of meta-analyses lacking a content expert.

Finally, we emphasize that so long as heterogeneity is unresolved, its causes are unknown. Unresolved heterogeneity might indicate an unrecognized substantive moderator, it might be the result of methodological artifacts, or it might stem from combined construct invalidity. And, of course, these three classes of causes are not mutually exclusive. All three can be present at the same time. Thus, we call for including complementary heterogeneity analyses in all future communication meta-analyses. At a minimum, such analyses should include investigations of publication bias beyond a fail-safe *N* (Orwin, 1983; Rosenthal, 1979; placeholder citation for publication bias essay), providing either standard funnel plots (Egger, Smith, Schneider, & Minder, 1997; Light & Pillemer, 1984), contour-enhanced funnel plots (Johnson & Hennessy, 2019), tests for skews in effect size distributions (e.g. studies with smaller sample size show larger effects; Stanley & Doucouliagos, 2014), or direct tests of publication bias by comparing effects from published journal articles with effects from the grey literature (e.g., unpublished dissertations). In addition, we suggest that researchers demonstrate their best efforts by including a discussion of how indications of questionable research practices—such as (a) optional removal of outliers; (b) optional selection between multiple dependent variables; (c) optional use of additional covariates; and (d) arbitrary decisions in stopping data collection in primary studies—may have affected between-study effect-size variations. As Carter, Schönbrodt, Gervais, and Hilgard (2019) have convincingly demonstrated in a large-scale simulation study of biased (primary) study results, such a discussion is essential not only for the identification of effect heterogeneity's potential causes, but also for determining the performance of different estimation models in meta-analyses (such as fixed- vs random-effect models). Carter et al. found that no particular estimation model demonstrated clear advantages over others; hence, they recommend adding sensitivity analyses of different estimation models in cases of between-study effect heterogeneity. Concretely, this means that researchers are advised to (a) conduct meta-analyses using more than one estimation method; (b) consider how the applied estimation methods may be superior or inferior in light of different types of questionable research practices; and then (c) report how their interpretations of results may change, dependent on which questionable research practices are most likely. While researchers can currently choose among many convenient software packages for meta-analyses that all provide different estimation methods, we realize that our last recommendation adds a substantial burden on the side of meta-analysts. However, given the severity of the problem, and in the spirit of best-practice recommendations, we believe these extra efforts are warranted in assuring the relevance and quality of future communication meta-analyses.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Please note: Oxford University Press is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

(References marked with an asterisk are included in the analysis.)

*Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281–311 doi: 10.1080/03637751003758193.

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and Psychological Measurement*, *62*(2), 254–263 doi:10.1177/0013164402062002004 .

Carpenter, C. J. (2020). Meta-analyzing apples and oranges: How to make applesauce instead of fruit salad. *Human Communication Research* (this issue). doi: 10.1093/hcr/hqz018.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. doi: 10.1177/2515245919847196.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. doi: 10.1136/bmj.315.7109.629.

*Feeley, T. H., Anker, A. E., & Aloe, A. M. (2012). The door-in-the-face persuasive message strategy: A meta-analysis of the first 35 years. *Communication Monographs*, *79*(3), 316–343. doi: 10.1080/03637751.2012.697631.

Gordon, M. (2019, June 25). *Introduction to forest plots*. Retrieved from https://cran.r-project.org/web/packages/forestplot/vignettes/forestplot.html.

Gronau, Q. F., Van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., & Wagenmakers, E. J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power. *Comprehensive Results in Social Psychology*, *2*(1), 123–138. doi: 10.1080/23743603.2017.1326760.

Higgins, J. P. T., & Thompson, S. G. (2002). Statistics in Medicine. In *Quantifying heterogeneity in a meta-analysis.* (Vol. 21, pp. 1539–1558). doi: 10.1002/sim.1186.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560. doi: 10.1136/bmj.327.7414.557.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. doi: 10.1037/1082-989X.11.2.193.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (1994). Correcting for sources of artificial variation across studies. In H. Cooper & L. V. Hedge (Eds.), *The handbook of research synthesis* (pp. 323–336). New York, NY: Russell Sage Foundation.

*Jeong, S. H., Cho, H., & Hwang, Y. (2012). Media literacy interventions: A meta-analytic review. *Journal of Communication*, *62*(3), 454–472. doi: 10.1111/j.14602466.2012.01643.x.

Johnson, B. T., & Hennessy, E. A. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine*, *233*, 237–251. doi: 10.1016/j.socscimed.2019.05.035.

Kpekpena, C., & Muthukumarana, S. (2018). Bayesian equivalence testing and meta-analysis in two-arm trials with binary data. *Computational and Mathematical Methods in Medicine*, *2018*, 8134132. doi: 10.1155/2018/8134132.

Levine, T. R., Blair, J. P., & Carpenter, C. J. (2018). A critical look at meta-analytic evidence for the cognitive approach to lie detection: A re-examination of Vrij, Fisher, and Blank (2017). *Legal and Criminological Psychology*, *23*(1), 7–19 doi: 10.1111/lcrp.12115.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

O'Keefe, D. J. (2013). The relative persuasiveness of different message types does not vary as a function of the persuasive outcome assessed: Evidence from 29 meta-analyses of 2,062 effect sizes for 13 message variations. *Annals of the International Communication Association*, *37*(1), 221–249. doi: 10.1080/23808985.2013.11679151.

*O'Keefe, D. J., & Jensen, J. D. (2009). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease detection behaviors: A meta-analytic review. *Journal of Communication*, *59*(2), 296–316. doi: 10.1111/j.1460-2466.2009.01417.x.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. doi: 10.2307/1164923.

Rains, S. A., Levine, T. R., & Weber, R. (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association*, *42*(2), 105–124. doi: 10.1080/23808985.2018.1446350.

Rains, S. A., Peterson, E. B., & Wright, K. B. (2015). Communicating social support in computer-mediated contexts: A meta-analytic review of content analyses examining support messages shared online among individuals coping with illness. *Communication Monographs*, *82*(4), 403–430. doi: 10.1080/03637751.2015.1019530.

*Reimer, T., Reimer, A., & Czienskowski, U. (2010). Decision-making groups attenuate the discussion bias in favor of shared information: A meta-analysis. *Communication Monographs*, *77*(1), 121–142. doi: 10.1080/03637750903514318.

Rhodes, K. M., Turner, R. M., & Higgins, J. P. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology*, *68*(1), 52–60. doi: 10.1016/j.jclinepi. 2014.08.012.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi: 10.1037/0033-2909.86.3.638.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Sedgwick, P. (2015). How to read a forest plot in a meta-analysis. *British Medical Journal*, *351*, h4028. doi: 10.1136/bmj.h4028.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. doi: 10.1002/jrsm.1095.

Sun, Y., Pan, Z., & Shen, L. (2008). Understanding the third-person perception: Evidence from a meta-analysis. *Journal of Communication*, *58*(2), 280–300. doi: 10.1111/j.1460-2466.2008.00385.x.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207–216. doi: 10.1016/S0895-4356(99)00161-4.

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., & Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane database of systematic reviews. *International Journal of Epidemiology*, *41*(3), 818–827. doi: 10.1093/ije/dys041.

Van Erp, S., Verhagen, J., Grasman, R. P. P. P., & Wagenmakers, E. J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in *Psychological Bulletin* from 1990–2013. *Journal of Open Psychology Data*, *5*, 1–5. doi: 10.5334/jopd.33.

Vermeulen, I., & Hartmann, T. (2015). Questionable research and publication practices in communication science. *Communication Methods and Measures*, *9*(4)), 189–192. doi: 10.1080/19312458.2015.1096331.

Viechtbauer, W. (2005). Bias and efficiency of metaanalytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. doi: 10.3102/10769986030003261.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21. doi: 10.1111/lcrp.12088.

Wright, P. J., Tokunaga, R. S., & Kraus, A. (2015). A meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies. *Journal of Communication*, *66*(1), 183–205. doi: 10.1111/jcom.12201.

Weber, R., & Popova, L. (2012). Testing equivalence in communication research: Theory and applications. *Communication Methods and Measures*, *6*(3), 190–213. doi: 10.1080/19312458.2012.703834.

*Wilson, S. R., Norris, A. M., Shi, X., & Rack, J. J. (2010). Comparing physically abused, neglected, and nonmaltreated children during interactions with their parents: A meta-analysis of observational studies. *Communication Monographs*, *77*(4), 540–575. doi: 10.1080/03637751.2010.502535.

Yang, Z. J., Aloe, A. M., & Feeley, T. H. (2014). Risk information seeking and processing model: A meta-analysis. *Journal of Communication*, *64*(1), 20–41. doi: 10.1111/jcom.12071.