USING META-SEN TO ASSESS THE TRUSTWORTHINESS OF META-ANALYTIC FINDINGS

James G. Field, Ph. D.
John Chambers College of Business and Economics
West Virginia University
Morgantown, WV
Email: j.g.field@hotmail.com
Phone: (304) 677-0791

Frank A. Bosco, Ph. D.
School of Business
Virginia Commonwealth University
Richmond, VA 23284-4000
Email: fabosco@vcu.edu
Phone: (804) 828-1602

ABSTRACT

To assess the robustness of meta-analytic inferences, one should compare results using different statistical techniques and assumptions. Two types of sensitivity analyses concerns examining the effect of outliers and publication bias on the obtained meta-analytic results. However, analyses to examine their independent and combined effects are rarely conducted, calling into question the trustworthiness of meta-analytic results. In this professional development institute, we introduce Meta-Sen (see https://metasen.shinyapps.io/gen1/), an open-source software that can be used to conduct a meta-analysis that adheres to recommended standards and best practices. We will demonstrate the tool's functionality and introduce new approaches to reporting sensitivity analysis results.

OVERVIEW OF THE PROFESSIONAL DEVELOPMENT INSTITUTE

Meta-analytic reviews are considered the primary means for generating cumulative scientific knowledge and their results are often used by practitioners to inform evidence-based practice. However, the results of published meta-analyses may be misestimated and, thus, untrustworthy because their robustness to the effects of outliers and publication bias (PB) is rarely examined. Evidence suggests that both phenomena can independently distort meta-analytic results. However, given that outliers can inflate the amount of residual heterogeneity in meta-analytic datasets, which can lead to biased meta-analytic and PB analysis results, a combined outlier and PB effect may also threaten meta-analytic results and conclusions. We introduce a tool that will facilitate the conduct of meta-analyses that adhere to recommended reporting standards and best practices. Specifically, we describe and demonstrate a cloud-based software (see https://metasen.shinyapps.io/gen1/) that allows users to upload a meta-analytic dataset and provides as output all essential meta-analytic and sensitivity analysis results before and after outlier removal. Together these results can be used to estimate the independent and combined effects of these phenomena.

An *outlier* is an observation that appears "to deviate markedly from other members of the sample in which it occurs" (Grubbs, 1969, p. 1). Outliers have long been acknowledged to have a potentially distorting influence on statistical analyses and their results, including meta-analytic ones (Viechtbauer & Cheung, 2010). For example, Rubenstein, Eberly, Lee, and Mitchell (2018) reported that the meta-analytic mean observed correlation for the "employee performance-voluntary turnover" relation changed from -.07 to -.17 (Δ = .10, or 143%) after removing a potential outlier. *Publication bias* (PB) occurs when there is a systematic suppression of research findings, which causes the available literature to be unrepresentative of all completed research on a relation of interest (Begg & Mazumdar, 1994) and, like outliers, has been shown to distort meta-analytic results. For example, a review of the strategic management literature found evidence for considerable levels of PB and, as a result, Harrison, Banks, Pollack, O'Boyle, and Short (2017, p. 400) suggested that "caution should be exercised when interpreting scientific conclusions regarding certain determinants of firm performance." Taken together, when not properly addressed, outliers and PB can lead to meta-analytic mean effect size estimates that are misestimated (Kepes, Banks, McDaniel, & Whetzel, 2012; Kepes & McDaniel, 2013). Moreover, both phenomena are often addressed as important ethical issues (Aguinis, Gottfredson & Joo, 2013) and can distort utility analyses (e.g., Hancock, Allen, Bosco, McDaniel & Pierce, 2013), which may impair evidence-based practice efforts (Kepes & McDaniel, 2015).

In this professional development institute (PDI), we begin by providing a brief introduction to the fundamentals of meta-analysis. Following this, we introduce a taxonomy of causes of outliers (see Table 1) and PB (see Table 2). Specifically, we focus on outcome-level and sample-level causes of outliers and PB. With regard to *outcome-level causes of outliers*, we describe the role played by a sample's effect size magnitude and *p*-value in determining whether or not it is labelled as an outlier. For instance, samples that have an effect size that diverges from all other samples in the dataset may need to be removed before performing a meta-analysis as they could introduce residual heterogeneity that may threaten its results (Kepes & McDaniel, 2015). In regard to *sample-level causes of outliers*, a study's sample size may play an important role in determining whether or not it is an outlier. Given that both the Hedges and Olkin (1985; see also Hedges & Olkin, 2014) and Schmidt and Hunter (2015) approaches to meta-analysis estimate the mean by giving more precise studies more weight, large samples can have an undue influence on the meta-analytic mean. As such, meta-analytic results with and without relatively large samples should be compared to determine the influence of large-sample studies. The PDI's discussion of outcome-level and sample-level causes of PB will be informed by Kepes et al.'s (2012) taxonomy of causes of PB.

Although evidence suggests that outliers and PB can have independent adverse downstream effects for research and practice (Kepes, Bennett, & McDaniel, 2014), there appears to be some degree of interdependence between the causes of outliers and the causes of PB. For instance, an effect size may be removed from a primary study manuscript before being submitted to a journal (i.e., author decision, outcome-level cause of PB; Table 2) because its corresponding *p*-value (i.e., outcome-level cause of outliers; see Table 1) was greater than the conventional statistical significance threshold ($p < .05$). In this case, an outlier-related phenomenon caused PB. Yet, to date, and to the best of our knowledge, sensitivity analyses of published meta-analytic results have largely failed to examine the combined effect of these phenomena (see Field, Bosco, & Kepes, 2021 for an exception that we are aware of). As such, this PDI will describe why it is likely important to account for both outliers *and* PB when assessing the trustworthiness of meta-analytic results.

Next, the PDI will review the strengths and weaknesses of methods used to detect and possibly adjust for outlier and PB effects. We will review two *outlier assessment methods* (one-sample removed analysis [Bornstein, Hedges, Higgins, & Rothstein, 2009] and Viechtbauer and Cheung's [2010, see also Viechtbauer, 2017] multivariate, multidimensional influence diagnostics) and five *PB assessment methods* (contour-enhanced funnel plots [Peters, Sutton, Jones, Abrams, & Ruston, 2008], Duval and Tweedie's [2000; see also Duval, 2005] trim and fill, cumulative meta-analysis by precision [Kepes et al., 2012], a priori selection models [Vevea

& Woods, 2005], and precision-effect test-precision effect estimate with standard error analysis [Stanley & Doucouliagos, 2014]).

Finally, we will demonstrate Meta-Sen (see https://metasen.shinyapps.io/gen1/; landing page and example output are displayed in Figures 1-6). Example data files will be provided so that attendees can interact with Meta-Sen during and after the PDI. In general, the purpose of the PDI is to illustrate that (1) outliers may distort meta-analytic results, (2) PB may distort meta-analytic results, and (3) PB results may change after removing outliers, indicating that outliers and PB can have a joint effect on the trustworthiness of meta-analytic results. The PDI will also introduce a new quantitative and visual way to summarize meta-analytic and sensitivity analysis results. With regard to the quantitative method, Meta-Sen illustrates how the degree of observed bias can be measured using a standardized mean difference and, thus, quantified using accepted benchmarks (i.e., $d = \sim.2$, .5, and .8 represent "small," "medium", and "large" degree of bias, respectively; Cohen, 1988). With regard to the visual method, Meta-Sen introduces a new way to display the range of meta-analytic and sensitivity analysis results before and after outlier removal. This new visualization shows if outliers and/or PB contributed to the range of results and, thus, the potential misestimation of the originally reported meta-analytic mean estimate.

To improve the transparency of meta-analytic findings, Meta-Sen allows the user to download all of the obtained results and plots. In addition, to better aid the user report the obtained quantitative results, table templates that adhere to American Psychological Association (APA) formatting requirements can be downloaded from their respective tabs. The PDI will conclude with recommendations for minimizing the impact of outliers and/or PB. These recommendations include changing author norms and the journal review processes. We will also encourage research registries and the ability to submit supplemental information to journals.

INTEREST TO SMA MEMBERSHIP

We assert that our PDI will be of interest to research methodologists as it introduces a comprehensive sensitivity analysis that is aligned with the APA's Meta-Analysis Reporting Standards (Appelbaum et al., 2018). Furthermore, the PDI discusses the importance of accounting for a statistical artifact that has been overlooked by nearly all PB assessments. Specifically, because the performance of PB methods (Peters, Sutton, Jones, Abrams & Rushton, 2007; Terrin, Schmid, Lau & Olkin, 2003) can be affected by outlier-driven heterogeneity (Kepes & McDaniel, 2015), the PDI demonstrates a software that illustrates how outlier presence may influence inferences regarding the presence of PB.

In addition, we contend that our PDI will be of interest to theorists, particularly those who use meta-analytic results as "building blocks of theory" (Schmidt, 1992, p. 1177). Finally, we propose that this PDI will be of interest to practitioners who use meta-analytic results to inform utility analyses (e.g., Hancock et al., 2013). Moreover, the PDI will introduce two new ways of communicating sensitivity analysis results. We suggest that these new methods are aligned with a consumer-centric (Aguinis et al., 2010) approach to reporting research results as it may help researchers and practitioners make better sense of sensitivity analysis results. Taken together, we assert that our PDI will be of interest to all research-active scholars and practitioners across all areas of the Southern Management Association.

DESCRIPTION OF PROFESSIONAL DEVELOPMENT INSTITUTE FORMAT

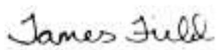
The format of the PDI will be as follows:

- Brief introduction to the presenters
- Discussion of fundamentals of meta-analytic procedures
- Discussion of causes of outliers and publication bias
- Discussion of independent and combined effects of outliers and PB on meta-analytic results.
- Review of two outlier assessment methods
- Review of five publication bias assessment methods
- Demonstration of Meta-Sen
    - Sample data files will be provided so that attendees can interact with the software
- Recommendations for minimizing the impact of outliers and/or PB
- Discussion/questions/comments from the audience

STATEMENT FROM ORGANIZER

I have received signed statements from all intended participants agreeing to participate in the entire symposium

Name: James Field

Signature: 

Date: 04/26/2021

REFERENCES

Aguinis, H., Gottfredson, R. K., & Joo, H. 2013. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16: 270-301.

Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. 2010. Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, 13: 515-539.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. 2018. Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73: 3-25.

Begg, C. B. & Mazumdar, M. 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50: 1088-1101.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. 2009. *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Cohen, J. 1988, *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition.* Hillsdale, NJ: Lawrence Erlbaum.

Duval, S., & Tweedie, R. 2000. A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95: 89-98.

Duval, S. J. 2005. The “trim and fill” method. In H. R. Rothstein, A. J. Sutton and M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments*: 127-144. West Sussex, UK: Wiley.

Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics,* 11: 1-21.

Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. 2013. Meta-analytic review of employee turnover as a predictor of firm performance. *Journal of Management,* 39: 573-603.

Harrison, J., S., Banks, G. C., Pollack, J., M. , O’Boyle, E., H. , & Short, J. 2017. Publication bias in strategic management research. *Journal of Management*, 43: 400-425.

Hedges, L. V., & Olkin, I. 1985. *Statistical method for meta-analysis*. New York, NY: Academic Press.

Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. 2012. Publication bias in the organizational sciences. *Organizational Research Methods,* 15: 624-662.

Kepes, S., Bennett, A. A., & McDaniel, M. A. 2014. Evidence-based management and the trustworthiness of our cumulative scientific knowledge: Implications for teaching, research, and practice. *Academy of Management Learning & Education,* 13: 446-466.

Kepes, S., & McDaniel, M. A. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology: Perspectives on Science and Practice,* 6: 252-268.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. 2007. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine,* 26: 4544-4562.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61: 991-996.

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. 2018. Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology,* 71: 23-65.

Schmidt, F. L. 1992. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47: 1173-1181.

Schmidt, F. L., & Hunter, J. E. 2015. *Methods of meta-analysis: Correcting error and bias in research findings.* (3rd ed.). Newbury Park, CA: Sage.

Stanley, T. D., & Doucouliagos, H. 2014. Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods,* 5: 60-78.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. 2003. Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, 22: 2113-2126.

Vevea, J. L., & Woods, C. M. 2005. Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods,* 10: 428-443.

Viechtbauer, W. 2017. Meta-analysis package for R: Package ‘metafor.’ R package version 2.0-0.

Viechtbauer, W., & Cheung, M. W. L. 2010. Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1: 112-125.

Table 1

Taxonomy of causes of outlier

Cause of outliers	Explanation
<u>Outcome-level causes</u>	
Effect size magnitude	Samples that have an effect size that diverges from the effect sizes of all other samples in the dataset may need to be removed before performing a meta-analysis as they could introduce residual heterogeneity that may threaten its results and conclusions.
P-value	An effect size may be labelled as an outlier if its corresponding *p*-value deviates noticeably from the other *p*-values in the dataset. Failing to remove such effect sizes may increase the degree of heterogeneity observed in a dataset and thus threaten its meta-analytic results.
<u>Sample-level causes</u>	
Sample size	Sample size is a characteristic that may determine whether or not an effect size is labelled as an outlier because both the Hedges and Olkin (1985; see also Hedges & Olkin, 2014) and Schmidt and Hunter (2015) approaches to meta-analysis estimate the meta-analytic mean by giving more precise studies more weight. Thus, relatively large samples can have an undue influence on the meta-analytic mean.
Sample type	In the context of a meta-analysis, an effect size that differs from all other effect sizes in regard to some sample type characteristic (e.g., incumbents vs. applicants, employees vs. students) may need to be removed before performing a meta-analysis as it could introduce residual heterogeneity that may threaten its results and conclusions. This may be especially true if theoretical evidence suggests the sample characteristic is a boundary condition.

Table 2

Taxonomy of causes of publications bias

Cause of publication bias	Explanation
Outcome-level causes	
Author decisions	Authors may decide to exclude some effect sizes prior to submitting a paper because the effects are not statistically significant, contrary to their expectations or theoretical position, contrary to past research, contrary to the position of the journal editor, etc.
Editorial review process	An editor may request that the author change the focus of the paper by making some results less relevant or request that the author drop the analyses yielding statistically non-significant effect sizes to "streamline" or "shorten" the paper.
Organizational constraints	Organizations who provide researchers with data cause outcome-level publication bias when they refuse to let authors published some results (e.g., demographic differences in pay or level of job performance)
Sample-level causes	
Author decisions	An author may contribute to publication bias if he/she works only on papers that have the highest chance of getting into the best journal; other papers may be abandoned and thus suppressed from the available literature.
Editorial review process	The editorial review process will reject papers that are poorly framed, papers without statistically significant findings, with results contrary to existing literature and current theory, and well done research that "didn't work." These editorial decisions result in suppression of effect sizes at the sample level.
Organizational constraints	An organization (e.g., employment test vendors) may force the suppression of entire studies if such studies damage the marketability of the organization's products.

Note. Adapted from Kepes et al. (2012)

Table 3

Analyses performed by Meta-Sen

Analysis/parameter
Meta-analysis
k (number of independent samples)[a]
N (sum of independent sample sizes)[a]
$\bar{r}_{oRE}$ (random effects meta-analytic mean effect size estimate)[a]
95% confidence interval[a]
90% prediction interval[a]
Q (weighted sum of squared deviations from the mean)[a]
I^2 (ratio of true heterogeneity to total variation)[a]
Tau (between-sample standard deviation)[a]
Outlier detection
One-sample removed[a]
Minimum, maximum, and median weighted mean observed correlation
Influence diagnostics[b]
Publication bias detection
Fixed-effects trim and fill model[a]
Side imputed
Number of imputed samples
Adjusted meta-analytic mean effect size estimate
Adjusted lower bound of 95% confidence interval
Random effects trim and fill model[a]
Side imputed
Number of imputed samples
Adjusted meta-analytic mean effect size estimate
Adjusted lower bound of 95% confidence interval
A priori selection model[a]
Moderate publication bias assumption
Back transformed adjusted meta-analytic mean effect size estimate
Severe publication bias assumption[a]
Back transformed adjusted meta-analytic mean effect size estimate
Precision-effect test-precision effect estimate with standard error (PET-PEESE)[a]
Weighted least squares approach
Final adjusted meta-analytic mean effect size estimate (two-tailed test)
Random effects meta-analysis (metafor; Viechtbauer [2017]) approach
Final adjusted meta-analytic mean effect size estimate (two-tailed test)
Cumulative meta-analysis by precision[a]

Note: [a] = estimated before and outlier removal; [b] = performed iteratively until all identified outliers are removed

Figure 1

Full view of the Meta-Sen graphical user interface

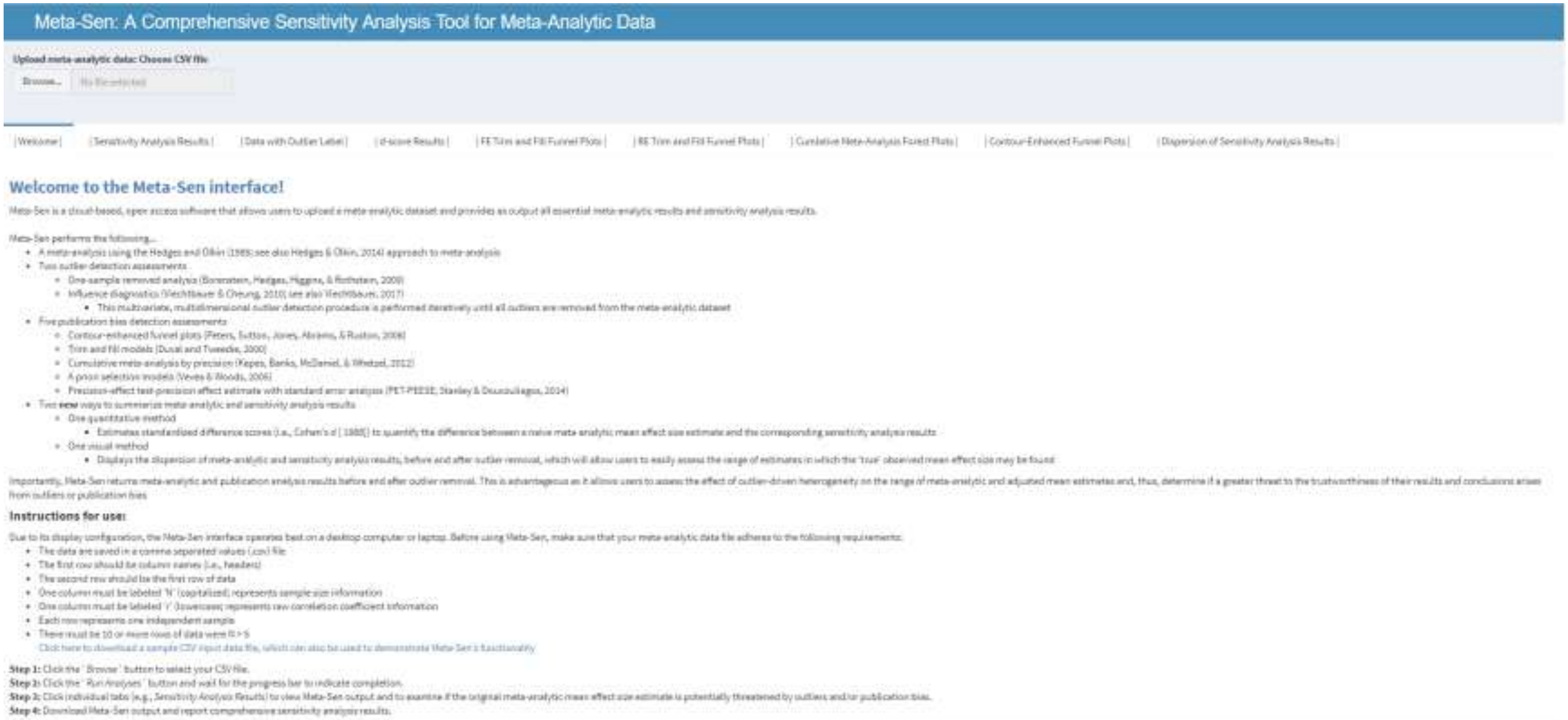
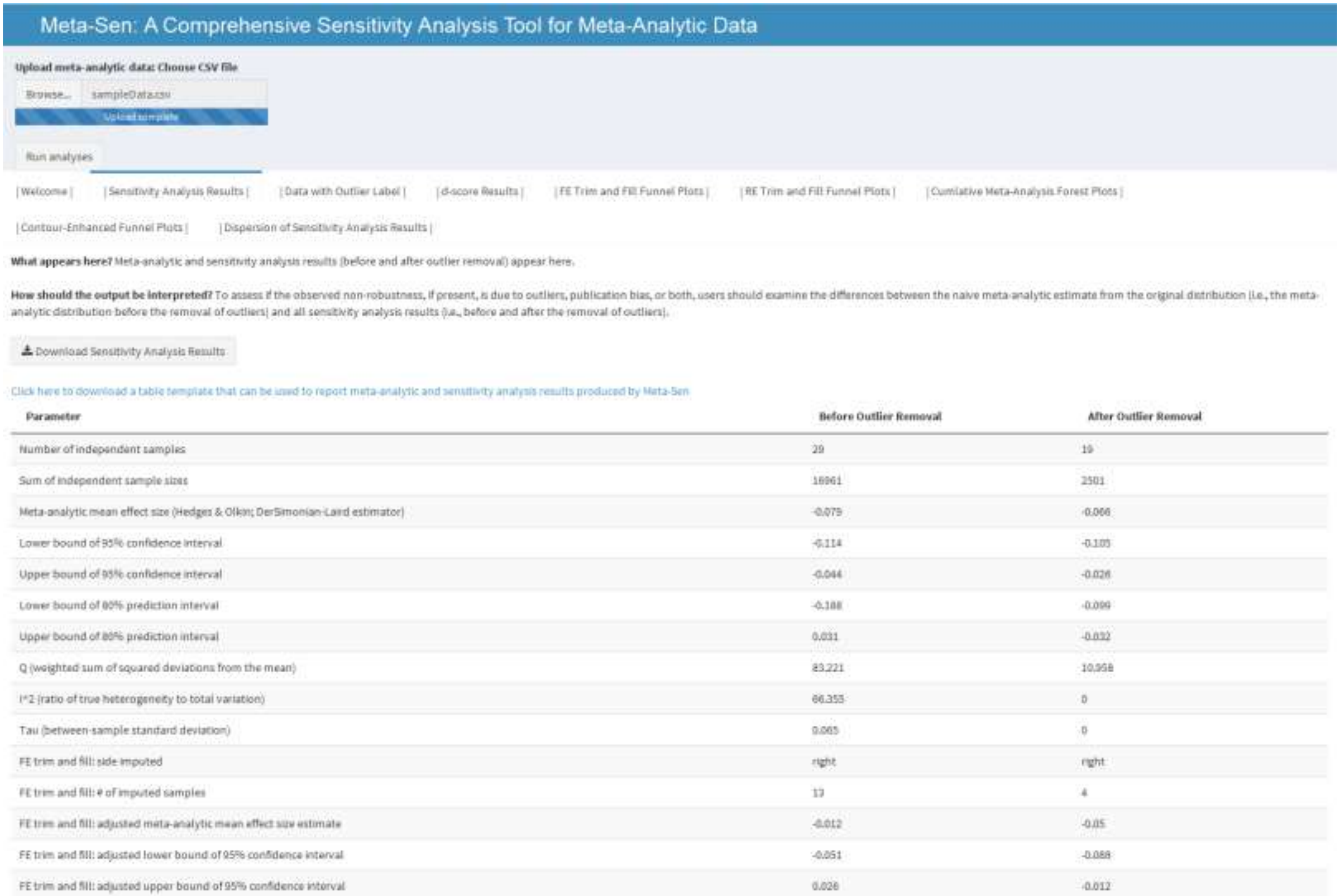


Figure 2

Short view of results tab showing meta-analytic and sensitivity analysis results before and after outlier removal



Parameter	Before Outlier Removal	After Outlier Removal
Number of independent samples	29	19
Sum of independent sample sizes	16961	2501
Meta-analytic mean effect size (Hedges & Olkin; DerSimonian-Laird estimator)	-0.079	-0.066
Lower bound of 95% confidence interval	-0.114	-0.105
Upper bound of 95% confidence interval	-0.044	-0.026
Lower bound of 80% prediction interval	-0.188	-0.099
Upper bound of 80% prediction interval	0.031	-0.032
Q (weighted sum of squared deviations from the mean)	83.221	10.958
I^2 (ratio of true heterogeneity to total variation)	66.355	0
Tau (between-sample standard deviation)	0.065	0
FE trim and fill: side imputed	right	right
FE trim and fill: # of imputed samples	13	4
FE trim and fill: adjusted meta-analytic mean effect size estimate	-0.012	-0.05
FE trim and fill: adjusted lower bound of 95% confidence interval	-0.051	-0.088
FE trim and fill: adjusted upper bound of 95% confidence interval	0.026	-0.012

Figure 3

Short view of data with outlier identification tab showing uploaded meta-analytic dataset and outlier classification

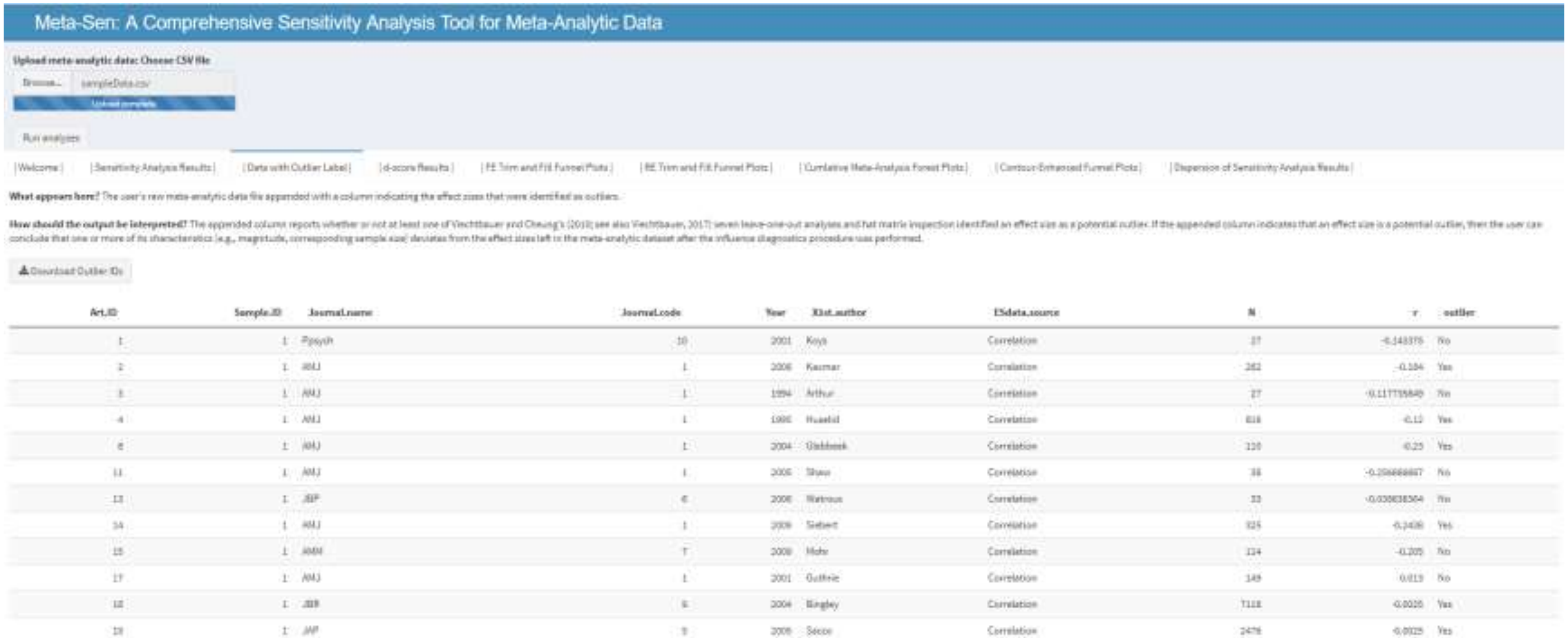
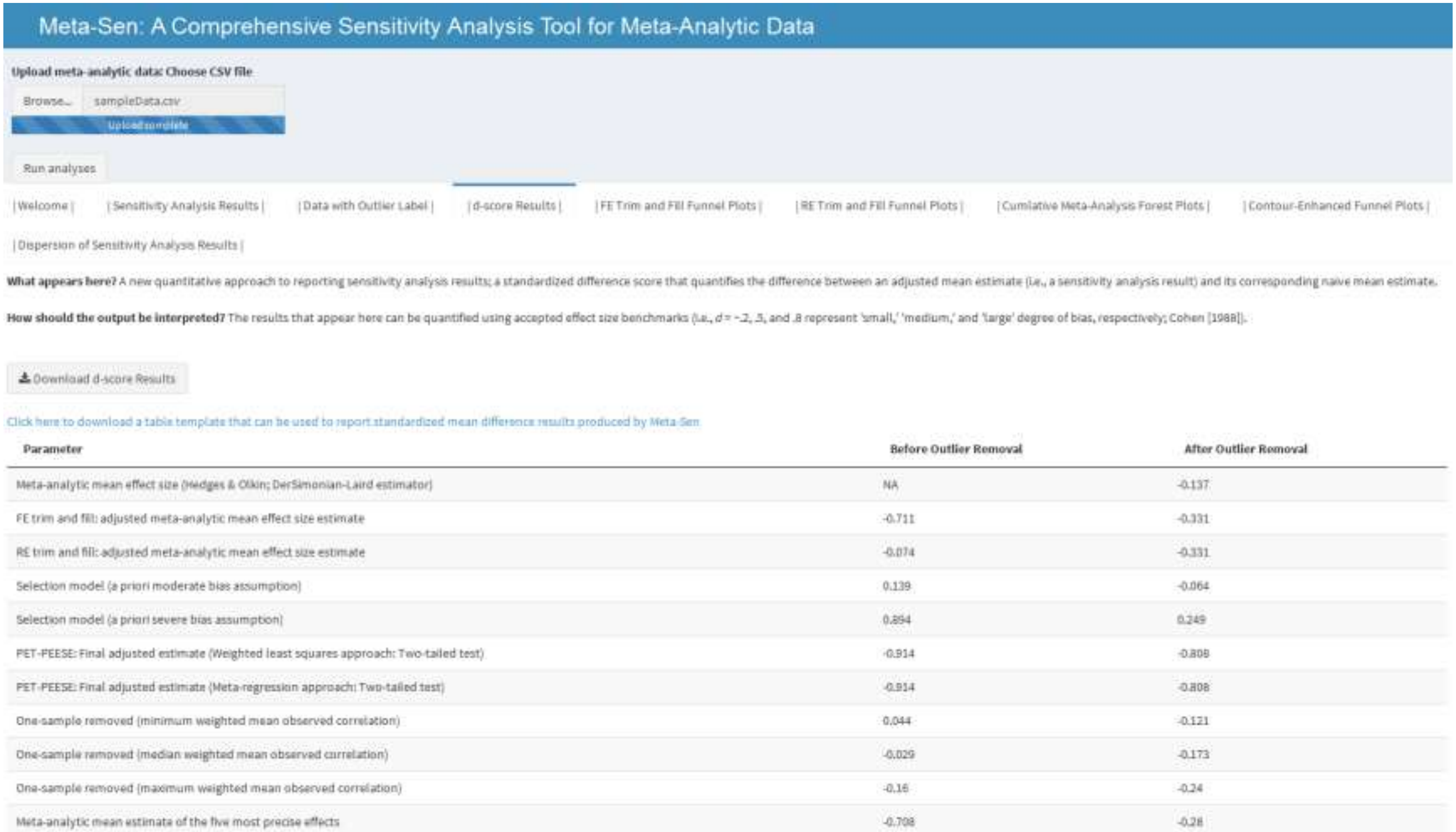


Figure 4

Full view of *d*-score results tab showing standardized mean difference results before and after outlier removal.


Meta-Sen: A Comprehensive Sensitivity Analysis Tool for Meta-Analytic Data

Upload meta-analytic data: Choose CSV file

Browse... sampleData.csv

Upload complete

Run analyses

| Welcome | | Sensitivity Analysis Results | | Data with Outlier Label | | d-score Results | | FE Trim and Fill Funnel Plots | | RE Trim and Fill Funnel Plots | | Cumlative Meta-Analysis Forest Plots | | Contour-Enhanced Funnel Plots | | Dispersion of Sensitivity Analysis Results |

What appears here? A new quantitative approach to reporting sensitivity analysis results; a standardized difference score that quantifies the difference between an adjusted mean estimate (i.e., a sensitivity analysis result) and its corresponding naive mean estimate.

How should the output be interpreted? The results that appear here can be quantified using accepted effect size benchmarks (i.e., *d* = ~.2, .5, and .8 represent 'small,' 'medium,' and 'large' degree of bias, respectively; Cohen [1988]).

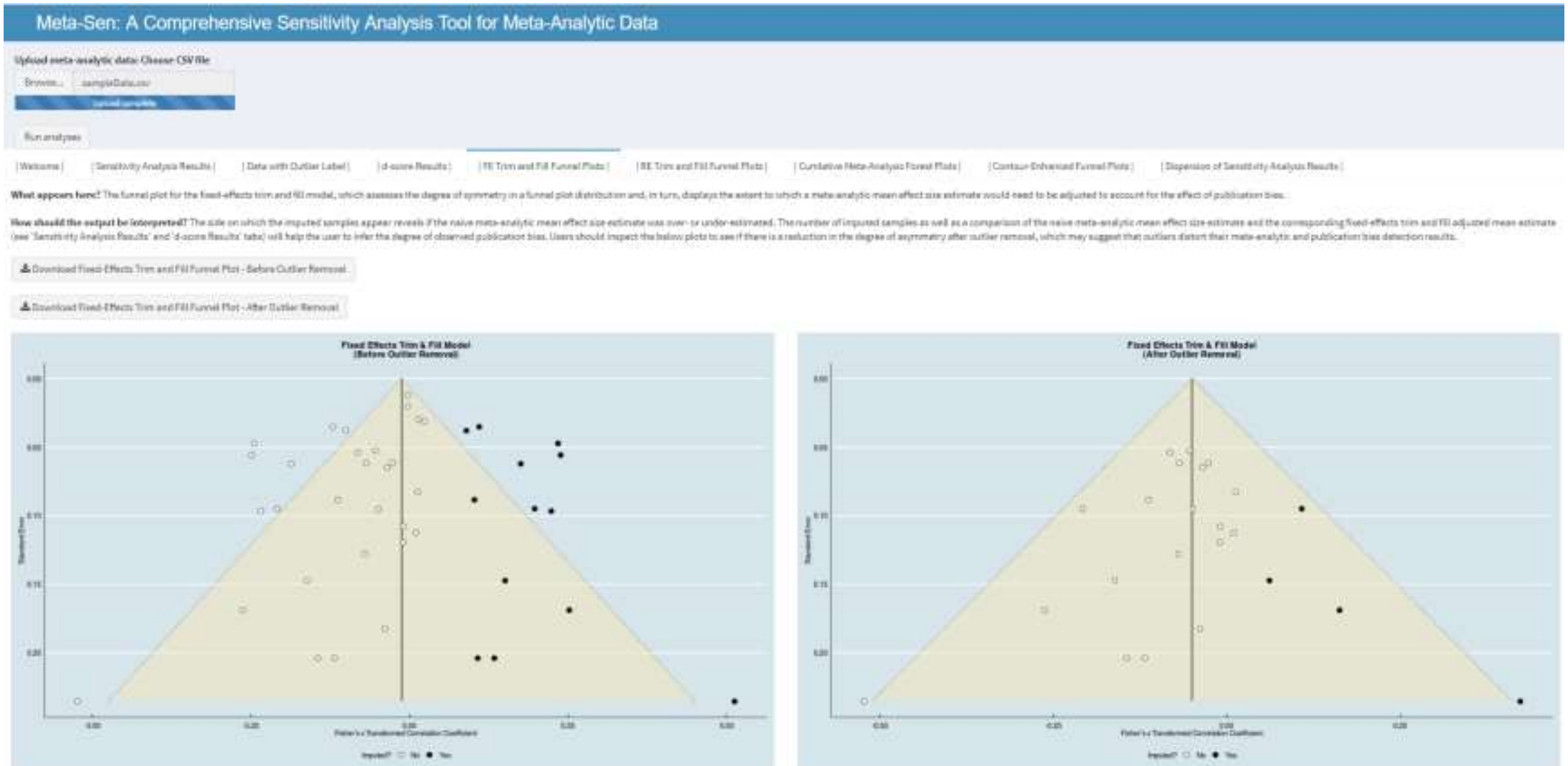
Download d-score Results

Click here to download a table template that can be used to report standardized mean difference results produced by Meta-Sen

Parameter	Before Outlier Removal	After Outlier Removal
Meta-analytic mean effect size (Hedges & Olkin; DerSimonian-Laird estimator)	NA	-0.137
FE trim and fill: adjusted meta-analytic mean effect size estimate	-0.711	-0.331
RE trim and fill: adjusted meta-analytic mean effect size estimate	-0.074	-0.331
Selection model (a priori moderate bias assumption)	0.139	-0.064
Selection model (a priori severe bias assumption)	0.894	0.249
PET-PEESE: Final adjusted estimate (Weighted least squares approach: Two-tailed test)	-0.914	-0.808
PET-PEESE: Final adjusted estimate (Meta-regression approach: Two-tailed test)	-0.914	-0.808
One-sample removed (minimum weighted mean observed correlation)	0.044	-0.121
One-sample removed (median weighted mean observed correlation)	-0.029	-0.173
One-sample removed (maximum weighted mean observed correlation)	-0.16	-0.24
Meta-analytic mean estimate of the five most precise effects	-0.708	-0.28

Figure 5

Full view of fixed-effects trim and fill funnel plots tab showing results before (left panel) and after (right panel) outlier removal.



Note. The clear dots represent observed correlations, the filled black dots represent the trim and fill imputed correlations. The vertical line represents the adjusted meta-analytic mean effect size.

Figure 6

Full view of dispersion of sensitivity analysis results tab

Note. Before = before outlier removal; after = after outlier removal; mean = random-effects weighted mean observed correlation; osrmax = one-sample removed maximum weighted mean observed correlation; osrmed = one-sample removed median weighted mean observed correlation; osrmin = one-sample removed minimum weighted mean observed correlation; pp = precision-effect test-precision effect estimate with standard error adjusted observed mean (meta-regression; two-tailed approach); pr = meta-analytic mean estimate of the five most precise effects; smm = one-tailed moderate selection model's adjusted observed mean; sms = one-tailed severe selection model's adjusted observed mean; tffe = fixed-effects trim and fill adjusted observed mean; tfre = random-effects trim and fill adjusted observed mean; The dashed vertical line represents a mean estimate of zero. The solid vertical line represents the naïve meta-analytic mean effect size.