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

James G. Field
West Virginia University

Frank A. Bosco
Virginia Commonwealth University

SMA PDI | November 3, 2021
New Orleans, LA

All submission materials (e.g., original submission, PowerPoint slides, sample data sets) can be found at:
https://jamiefield.github.io/research/sma2021

Agenda

- What is meta-analysis and why is so important?
- The current research environment
- Threats to our cumulative scientific knowledge: outliers and publication bias
- What is sensitivity analysis?
- Live demonstration of Meta-Sen
- Recommendations for minimizing the impact of outliers and/or PB
- Discussion/questions/comments from the audience
- Additional slides
  - Review of two outlier assessment methods
  - Review of five publication bias assessment methods
  - Results that illustrate the combined effect of outliers and PB on recently published meta-analytic datasets

What is Meta-Analysis?

- Meta-analysis is a statistical technique by which information from independent studies is assimilated
  - Field, A. P. (2011)

- Meta-analysis is a quantitative method used to combine the quantitative outcomes (effect sizes) of primary research studies.
  - Combines the results from two or more studies
  - Estimates an ‘average’ effect between two constructs

- Meta-analysis is the statistical or data analytic part of a systematic review of a research topic.

What is Meta-Analysis?

- There are two common approaches to meta-analysis
  - The Hunter and Schmidt (2004; 2015) approach, which is most common in organizational research
  - The Hedges and Olkin (1985) approach

- For a description of both approaches, and their differences, please refer to Kepes et al. (2013)

- Important note:
  - We use the Hedges and Olkin (1985) approach as most sensitivity analysis techniques have not been developed for psychometrically-adjusted effect sizes

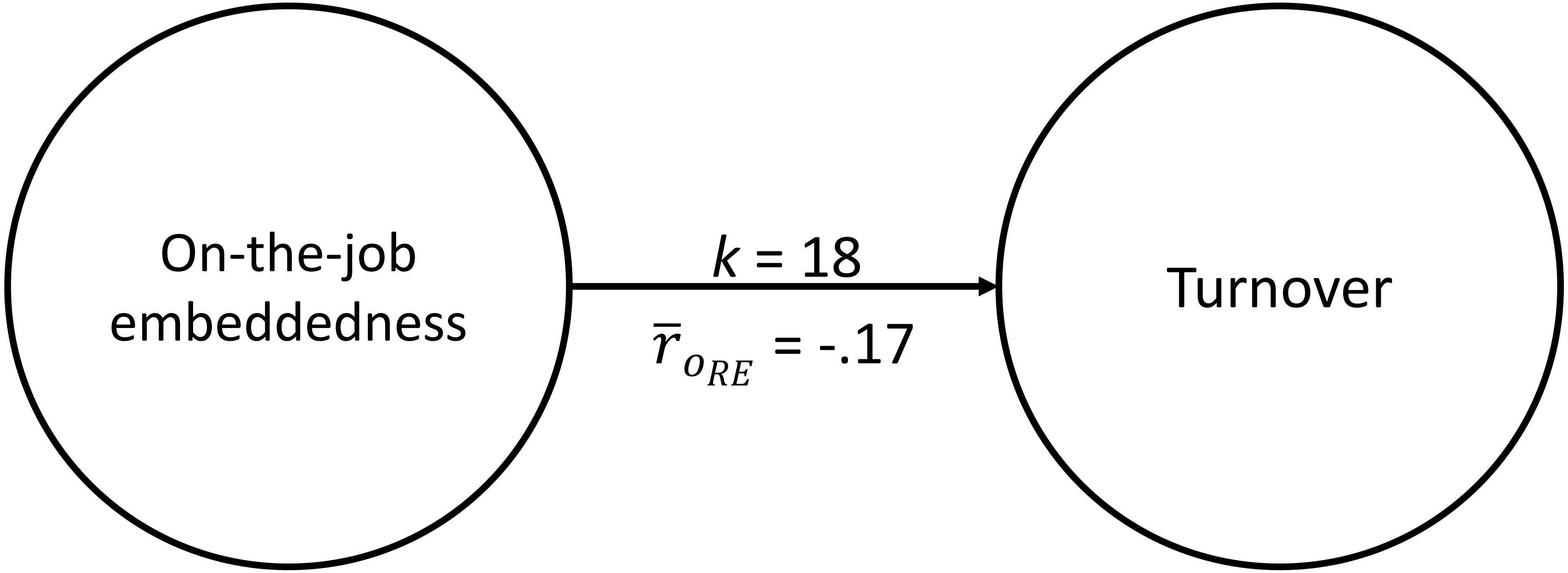
What is Meta-Analysis?

- An example from the published literature

Study ID	Reference	Year	IV	DV	*n*	*r*
1	Mitchell, Holtom, Lee, & Erez	2001	On-the-job embeddedness	Turnover	177	-0.24
2	Crossley, Bennett, Jex, & Burnfield	2007	On-the-job embeddedness	Turnover	306	-0.08
3	Mitchell, Holtom, Lee, & Erez	2001	On-the-job embeddedness	Turnover	208	-0.21
4	Ramesh & Gelfand	2010	On-the-job embeddedness	Turnover	474	-0.13
5	Giosan, Holtom, & Watson	2005	On-the-job embeddedness	Turnover	122	-0.30
6	Mallol, Holtom, & Lee	2007	On-the-job embeddedness	Turnover	164	-0.16
7	Lee, Mitchell, & Holtom	2004	On-the-job embeddedness	Turnover	809	-0.11
8	Kraimer, Shaffer, Harrison, & Ren	2012	On-the-job embeddedness	Turnover	112	-0.17
9	Ramesh & Gelfand	2010	On-the-job embeddedness	Turnover	323	-0.14
10	Harris, Wheeler, & Kacmar	2011	On-the-job embeddedness	Turnover	205	-0.19
11	Allen	2006	On-the-job embeddedness	Turnover	222	-0.23
12	Wheeler, Halbesleben, & Sablynski	2011	On-the-job embeddedness	Turnover	142	-0.26
13	Tanova & Holtom	2008	On-the-job embeddedness	Turnover	9277	-0.08
14	Wheeler, Halbesleben, & Sablynski	2011	On-the-job embeddedness	Turnover	134	-0.19
15	Mallol, Holtom, & Lee	2007	On-the-job embeddedness	Turnover	164	-0.13
16	Tharenou & Caulfield	2010	On-the-job embeddedness	Turnover	546	-0.18
17	Smith, Holtom, & Mitchell	2011	On-the-job embeddedness	Turnover	750	-0.25
18	Smith, Holtom, & Mitchell	2011	On-the-job embeddedness	Turnover	1,089	-0.19

Study ID	Reference	Year	IV	DV	*n*	*r*	sei
1	Mitchell, Holtom, Lee, & Erez	2001	On-the-job embeddedness	Turnover	177	-0.24	0.0758
2	Crossley, Bennett, Jex, & Burnfield	2007	On-the-job embeddedness	Turnover	306	-0.08	0.0574
3	Mitchell, Holtom, Lee, & Erez	2001	On-the-job embeddedness	Turnover	208	-0.21	0.0698
4	Ramesh & Gelfand	2010	On-the-job embeddedness	Turnover	474	-0.13	0.0461
5	Giosan, Holtom, & Watson	2005	On-the-job embeddedness	Turnover	122	-0.3	0.0917
6	Mallol, Holtom, & Lee	2007	On-the-job embeddedness	Turnover	164	-0.16	0.0788
7	Lee, Mitchell, & Holtom	2004	On-the-job embeddedness	Turnover	809	-0.11	0.0352
8	Kraimer, Shaffer, Harrison, & Ren	2012	On-the-job embeddedness	Turnover	112	-0.17	0.0958
9	Ramesh & Gelfand	2010	On-the-job embeddedness	Turnover	323	-0.14	0.0559
10	Harris, Wheeler, & Kacmar	2011	On-the-job embeddedness	Turnover	205	-0.19	0.0704
11	Allen	2006	On-the-job embeddedness	Turnover	222	-0.23	0.0676
12	Wheeler, Halbesleben, & Sablynski	2011	On-the-job embeddedness	Turnover	142	-0.26	0.0848
13	Tanova & Holtom	2008	On-the-job embeddedness	Turnover	9277	-0.08	0.0104
14	Wheeler, Halbesleben, & Sablynski	2011	On-the-job embeddedness	Turnover	134	-0.19	0.0874
15	Mallol, Holtom, & Lee	2007	On-the-job embeddedness	Turnover	164	-0.13	0.0788
16	Tharenou & Caulfield	2010	On-the-job embeddedness	Turnover	546	-0.18	0.0429
17	Smith, Holtom, & Mitchell	2011	On-the-job embeddedness	Turnover	750	-0.25	0.0366
18	Smith, Holtom, & Mitchell	2011	On-the-job embeddedness	Turnover	1,089	-0.19	0.0303

What is Meta-Analysis?



Why are Meta-Analyses so Important?

- Meta-analytic reviews are a primary way to summarize, integrate, and synthesize areas of research
  - Schmidt & Hunter (2015)

- Allows fields to build a cumulative scientific knowledge
  - Kepes & McDaniel (2015)

- Meta-analytic results serve as input for other analytic techniques that allow researchers to test theory
  - E.g., relative importance analysis; meta-analytic structural equation modeling

- Meta-analytic results often are used to inform evidence-based management
  - Banks et al. (2011); Kepes et al. (2014)

Why are Meta-Analyses so Important?

- Flexibility

Why are Meta-Analyses so Important?

- Flexibility

- Example #1:
    - We used meta-analytic techniques to examine the extent to which six popular cross-cultural models explain variance in research findings.

Journal of International Business Studies (2021)

www.jibs.net

More alike than different? A comparison of variance explained by cross-cultural models

James G. Field[1],
Frank A. Bosco[2],
David Kraichy[3],
Krista L. Uggerslev[4] and
Mingang K. Geiger[1]

[1] Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA; [2] Department of Management and Entrepreneurship, School of Business, Virginia Commonwealth University, Richmond, VA 23284, USA; [3] Department of Human Resources and Organizational Behaviour, Edwards School of Business, University of Saskatchewan, Saskatoon, SK S7N 5A7, Canada; [4] Applied Research Chair in Leadership and Talent, Office of Research and Innovation, Northern Alberta Institute of Technology, Edmonton, AB T5G 2R1, Canada

Correspondence:
JG Field, Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA
e-mail: james.field2@mail.wvu.edu

Abstract
Relatively little is known about the extent to which culture moderates findings in applied psychology research. To address this gap, we leverage the metaBUS database of over 1,000,000 published findings to examine the extent to which six popular cross-cultural models explain variance in findings across 136 bivariate relationships and 56 individual cultural dimensions. We compare moderating effects attributable to Hofstede's dimensions, GLOBE's practices, GLOBE's values, Schwartz's Value Survey, Ronen and Shenkar's cultural clusters, and the United Nations' M49 standard. Results from 25,296 multilevel meta-analyses indicate that, after accounting for statistical artifacts, cross-cultural models explain approximately 5–7% of the variance in findings. The variance explained did not vary substantially across models. A similar set of analyses on observed effect sizes reveal differences of $|r| = .05$–$.07$ attributable to culture. Variance among the 136 bivariate relationships was explained primarily by sampling error, indicating that cross-cultural moderation assessments require atypically large sample sizes. Our results provide important information for understanding the overall level of explanatory power attributable to cross-cultural models, their relative performance, and their sensitivity to variance in the topic of study. In addition, our findings may be used to inform power analyses for future research. We discuss implications for research and practice.
Journal of International Business Studies (2021).
https://doi.org/10.1057/s41267-021-00428-z

Keywords: meta-analysis; big data; open science; cross-cultural research/measurement issues

Why are Meta-Analyses so Important?

- Flexibility

- Example #1:
  - We compared moderating effects attributable to Hofstede's dimensions, GLOBE's practices, GLOBE's values, Schwartz's Value Survey, Ronen and Shenkar's cultural clusters, and the United Nations' M49 standard.

Journal of International Business Studies (2021)

www.jibs.net

More alike than different? A comparison of variance explained by cross-cultural models

James G. Field[1],
Frank A. Bosco[2],
David Kraichy[3],
Krista L. Uggerslev[4] and
Mingang K. Geiger[1]

[1] Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA; [2] Department of Management and Entrepreneurship, School of Business, Virginia Commonwealth University, Richmond, VA 23284, USA; [3] Department of Human Resources and Organizational Behaviour, Edwards School of Business, University of Saskatchewan, Saskatoon, SK S7N 5A7, Canada; [4] Applied Research Chair in Leadership and Talent, Office of Research and Innovation, Northern Alberta Institute of Technology, Edmonton, AB T5G 2R1, Canada

Correspondence:
JG Field, Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA
e-mail: james.field2@mail.wvu.edu

Abstract
Relatively little is known about the extent to which culture moderates findings in applied psychology research. To address this gap, we leverage the metaBUS database of over 1,000,000 published findings to examine the extent to which six popular cross-cultural models explain variance in findings across 136 bivariate relationships and 56 individual cultural dimensions. We compare moderating effects attributable to Hofstede's dimensions, GLOBE's practices, GLOBE's values, Schwartz's Value Survey, Ronen and Shenkar's cultural clusters, and the United Nations' M49 standard. Results from 25,296 multilevel meta-analyses indicate that, after accounting for statistical artifacts, cross-cultural models explain approximately 5–7% of the variance in findings. The variance explained did not vary substantially across models. A similar set of analyses on observed effect sizes reveal differences of $|r| = .05–.07$ attributable to culture. Variance among the 136 bivariate relationships was explained primarily by sampling error, indicating that cross-cultural moderation assessments require atypically large sample sizes. Our results provide important information for understanding the overall level of explanatory power attributable to cross-cultural models, their relative performance, and their sensitivity to variance in the topic of study. In addition, our findings may be used to inform power analyses for future research. We discuss implications for research and practice.
Journal of International Business Studies (2021).
https://doi.org/10.1057/s41267-021-00428-z

Keywords: meta-analysis; big data; open science; cross-cultural research/measurement issues

Why are Meta-Analyses so Important?

- Flexibility

- Example #1:
  - Results from 25,296 multilevel meta-analyses indicated that, after accounting for statistical artifacts, cross-cultural models explain approximately 5–7% of the variance in findings – roughly the same amount as a theoretically relevant moderator.

Journal of International Business Studies (2021)

www.jibs.net

More alike than different? A comparison of variance explained by cross-cultural models

James G. Field[1],
Frank A. Bosco[2],
David Kraichy[3],
Krista L. Uggerslev[4] and
Mingang K. Geiger[1]

[1]Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA; [2]Department of Management and Entrepreneurship, School of Business, Virginia Commonwealth University, Richmond, VA 23284, USA; [3]Department of Human Resources and Organizational Behaviour, Edwards School of Business, University of Saskatchewan, Saskatoon, SK S7N 5A7, Canada; [4]Applied Research Chair in Leadership and Talent, Office of Research and Innovation, Northern Alberta Institute of Technology, Edmonton, AB T5G 2R1, Canada

Correspondence:
JG Field, Department of Management, John Chambers College of Business and Economics, West Virginia University, Morgantown, WV 26506, USA
e-mail: james.field2@mail.wvu.edu

Abstract
Relatively little is known about the extent to which culture moderates findings in applied psychology research. To address this gap, we leverage the metaBUS database of over 1,000,000 published findings to examine the extent to which six popular cross-cultural models explain variance in findings across 136 bivariate relationships and 56 individual cultural dimensions. We compare moderating effects attributable to Hofstede's dimensions, GLOBE's practices, GLOBE's values, Schwartz's Value Survey, Ronen and Shenkar's cultural clusters, and the United Nations' M49 standard. Results from 25,296 multilevel meta-analyses indicate that, after accounting for statistical artifacts, cross-cultural models explain approximately 5–7% of the variance in findings. The variance explained did not vary substantially across models. A similar set of analyses on observed effect sizes reveal differences of $|r| = .05$–$.07$ attributable to culture. Variance among the 136 bivariate relationships was explained primarily by sampling error, indicating that cross-cultural moderation assessments require atypically large sample sizes. Our results provide important information for understanding the overall level of explanatory power attributable to cross-cultural models, their relative performance, and their sensitivity to variance in the topic of study. In addition, our findings may be used to inform power analyses for future research. We discuss implications for research and practice.
Journal of International Business Studies (2021).
https://doi.org/10.1057/s41267-021-00428-z

Keywords: meta-analysis; big data; open science; cross-cultural research/measurement issues

Why are Meta-Analyses so Important?

- Flexibility

- Example #2:
  - We used meta-analytic techniques to examine whether or not Cohen's (1988) effect size benchmarks generalize to the field of applied psychology.

Journal of Applied Psychology
2015, Vol. 100, No. 2, 431–449

0021-9010/15/$12.00 http://dx.doi.org/10.1037/a0038047

Correlational Effect Size Benchmarks

Frank A. Bosco
Virginia Commonwealth University

Herman Aguinis
Indiana University

Kulraj Singh
South Dakota State University

James G. Field
Virginia Commonwealth University

Charles A. Pierce
University of Memphis

Effect size information is essential for the scientific enterprise and plays an increasingly central role in the scientific process. We extracted 147,328 correlations and developed a hierarchical taxonomy of variables reported in *Journal of Applied Psychology* and *Personnel Psychology* from 1980 to 2010 to produce empirical effect size benchmarks at the omnibus level, for 20 common research domains, and for an even finer grained level of generality. Results indicate that the usual interpretation and classification of effect sizes as small, medium, and large bear almost no resemblance to findings in the field, because distributions of effect sizes exhibit tertile partitions at values approximately one-half to one-third those intuited by Cohen (1988). Our results offer information that can be used for research planning and design purposes, such as producing better informed non-nil hypotheses and estimating statistical power and planning sample size accordingly. We also offer information useful for understanding the relative importance of the effect sizes found in a particular study in relationship to others and which research domains have advanced more or less, given that larger effect sizes indicate a better understanding of a phenomenon. Also, our study offers information about research domains for which the investigation of moderating effects may be more fruitful and provide information that is likely to facilitate the implementation of Bayesian analysis. Finally, our study offers information that practitioners can use to evaluate the relative effectiveness of various types of interventions.

Keywords: effect size, statistical analysis, null hypothesis testing, big data

Why are Meta-Analyses so Important?

- Flexibility

- Example #2:
  - Our Results indicated that the usual interpretation and classification of effect sizes as small, medium, and large bear almost no resemblance to findings in the field.

Journal of Applied Psychology
2015, Vol. 100, No. 2, 431–449

0021-9010/15/$12.00 http://dx.doi.org/10.1037/a0038047

Correlational Effect Size Benchmarks

Frank A. Bosco
Virginia Commonwealth University

Herman Aguinis
Indiana University

Kulraj Singh
South Dakota State University

James G. Field
Virginia Commonwealth University

Charles A. Pierce
University of Memphis

Effect size information is essential for the scientific enterprise and plays an increasingly central role in the scientific process. We extracted 147,328 correlations and developed a hierarchical taxonomy of variables reported in *Journal of Applied Psychology* and *Personnel Psychology* from 1980 to 2010 to produce empirical effect size benchmarks at the omnibus level, for 20 common research domains, and for an even finer grained level of generality. Results indicate that the usual interpretation and classification of effect sizes as small, medium, and large bear almost no resemblance to findings in the field, because distributions of effect sizes exhibit tertile partitions at values approximately one-half to one-third those intuited by Cohen (1988). Our results offer information that can be used for research planning and design purposes, such as producing better informed non-nil hypotheses and estimating statistical power and planning sample size accordingly. We also offer information useful for understanding the relative importance of the effect sizes found in a particular study in relationship to others and which research domains have advanced more or less, given that larger effect sizes indicate a better understanding of a phenomenon. Also, our study offers information about research domains for which the investigation of moderating effects may be more fruitful and provide information that is likely to facilitate the implementation of Bayesian analysis. Finally, our study offers information that practitioners can use to evaluate the relative effectiveness of various types of interventions.

Keywords: effect size, statistical analysis, null hypothesis testing, big data

Why are Meta-Analyses so Important?

- Flexibility

- Example #3:
  - We used meta-analytic techniques to examine the potential downstream effects of a questionable research practice

PERSONNEL PSYCHOLOGY

THE STUDY OF PEOPLE AT WORK

PERSONNEL PSYCHOLOGY
2016, 69, 709–750

HARKING'S THREAT TO ORGANIZATIONAL RESEARCH: EVIDENCE FROM PRIMARY AND META-ANALYTIC SOURCES

FRANK A. BOSCO
Virginia Commonwealth University

HERMAN AGUINIS
Indiana University

JAMES G. FIELD
Virginia Commonwealth University

CHARLES A. PIERCE
University of Memphis

DAN R. DALTON
Indiana University

We assessed presumed consequences of hypothesizing after results are known (HARKing) by contrasting hypothesized versus nonhypothesized effect sizes among 10 common relations in organizational behavior, human resource management, and industrial and organizational psychology research. In Study 1, we analyzed 247 correlations representing 9 relations with individual performance in 136 articles published in *Journal of Applied Psychology* and *Personnel Psychology* and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In Study 2, we analyzed 281 effect sizes from a meta-analysis on the job satisfaction–job performance relation and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In addition, in Study 2, we documented that hypothesized variable pairs are more likely to be mentioned in article titles or abstracts. We also ruled out 13 alternative explanations to the presumed HARKing effect pertaining to methodological (e.g., unreliability, publication year, research setting, research design, measure contextualization, publication source) and substantive (e.g., predictor–performance pair, performance measure, satisfaction measure,

Why are Meta-Analyses so Important?

- Flexibility

- Example #3:
  - We observed that correlations are significantly larger when hypothesized compared to nonhypothesized

PERSONNEL PSYCHOLOGY

THE STUDY OF PEOPLE AT WORK

PERSONNEL PSYCHOLOGY
2016, 69, 709–750

HARKING'S THREAT TO ORGANIZATIONAL RESEARCH: EVIDENCE FROM PRIMARY AND META-ANALYTIC SOURCES

FRANK A. BOSCO
Virginia Commonwealth University

HERMAN AGUINIS
Indiana University

JAMES G. FIELD
Virginia Commonwealth University

CHARLES A. PIERCE
University of Memphis

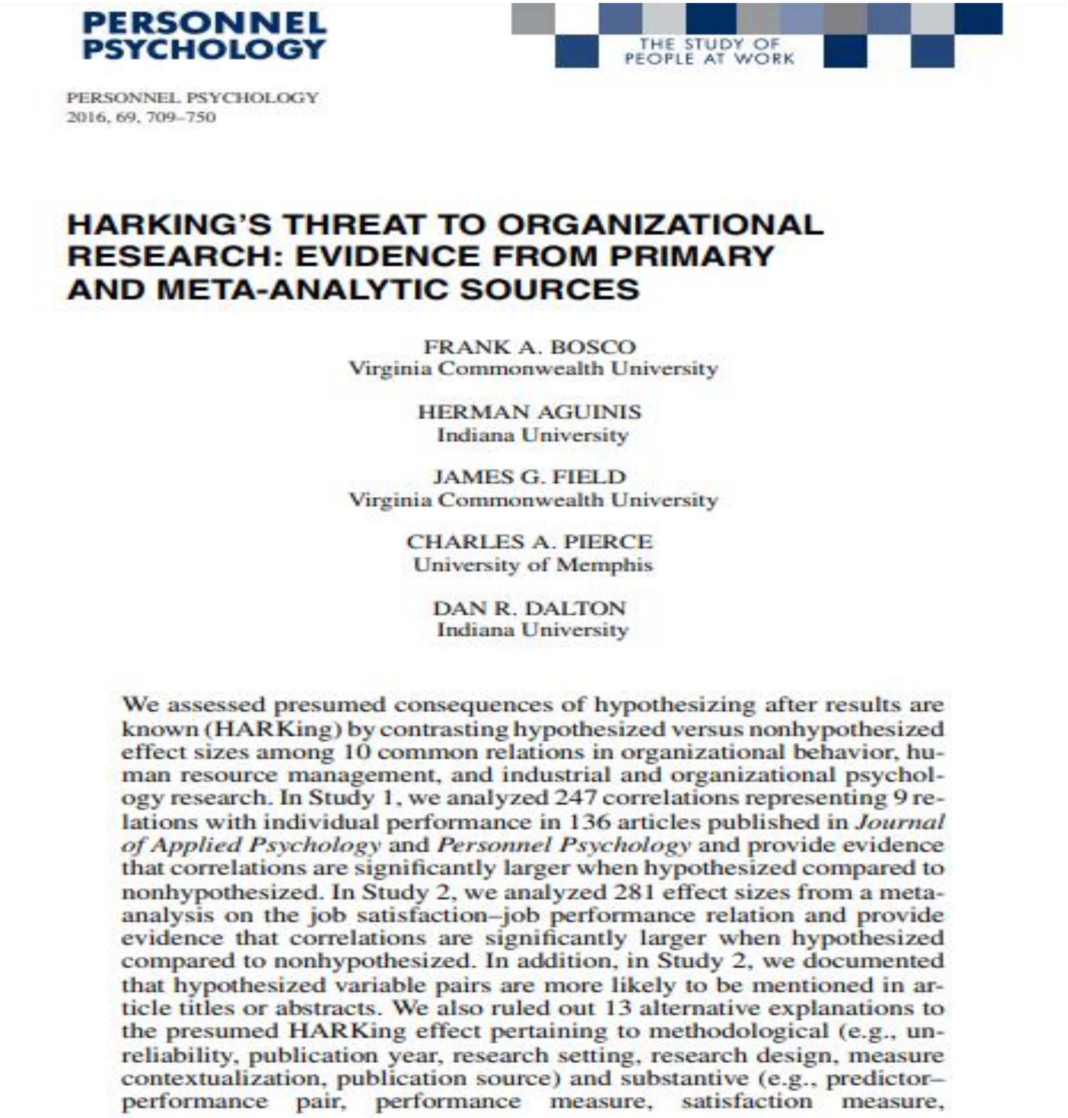
DAN R. DALTON
Indiana University

We assessed presumed consequences of hypothesizing after results are known (HARKing) by contrasting hypothesized versus nonhypothesized effect sizes among 10 common relations in organizational behavior, human resource management, and industrial and organizational psychology research. In Study 1, we analyzed 247 correlations representing 9 relations with individual performance in 136 articles published in *Journal of Applied Psychology* and *Personnel Psychology* and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In Study 2, we analyzed 281 effect sizes from a meta-analysis on the job satisfaction–job performance relation and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In addition, in Study 2, we documented that hypothesized variable pairs are more likely to be mentioned in article titles or abstracts. We also ruled out 13 alternative explanations to the presumed HARKing effect pertaining to methodological (e.g., unreliability, publication year, research setting, research design, measure contextualization, publication source) and substantive (e.g., predictor–performance pair, performance measure, satisfaction measure,

Why are Meta-Analyses so Important?

- Flexibility

- Example #3:
    - Suggests that HARKing (hypothesizing after the results are known) may pose a threat to research results, substantive conclusions, and practical applications



PERSONNEL PSYCHOLOGY

THE STUDY OF PEOPLE AT WORK

PERSONNEL PSYCHOLOGY
2016, 69, 709–750

HARKING'S THREAT TO ORGANIZATIONAL RESEARCH: EVIDENCE FROM PRIMARY AND META-ANALYTIC SOURCES

FRANK A. BOSCO
Virginia Commonwealth University

HERMAN AGUINIS
Indiana University

JAMES G. FIELD
Virginia Commonwealth University

CHARLES A. PIERCE
University of Memphis

DAN R. DALTON
Indiana University

We assessed presumed consequences of hypothesizing after results are known (HARKing) by contrasting hypothesized versus nonhypothesized effect sizes among 10 common relations in organizational behavior, human resource management, and industrial and organizational psychology research. In Study 1, we analyzed 247 correlations representing 9 relations with individual performance in 136 articles published in *Journal of Applied Psychology* and *Personnel Psychology* and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In Study 2, we analyzed 281 effect sizes from a meta-analysis on the job satisfaction–job performance relation and provide evidence that correlations are significantly larger when hypothesized compared to nonhypothesized. In addition, in Study 2, we documented that hypothesized variable pairs are more likely to be mentioned in article titles or abstracts. We also ruled out 13 alternative explanations to the presumed HARKing effect pertaining to methodological (e.g., unreliability, publication year, research setting, research design, measure contextualization, publication source) and substantive (e.g., predictor–performance pair, performance measure, satisfaction measure,

The Future of Meta-Analysis

- Will likely increase in importance
  - Global scientific publication output may grow by up to 400% in the next 50 years

Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References

Lutz Bornmann
Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Munich, Germany. E-mail: bornmann@gv.mpg.de

Rüdiger Mutz
Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Mühlegasse 21 8001, Zuerich, Switzerland. E-mail: ruediger.mutz@gess.ethz.ch

Many studies (in information science) have looked at the growth of science. In this study, we reexamine the question of the growth of science. To do this we (a) use current data up to publication year 2012 and (b) analyze the data across all disciplines and also separately for the natural sciences and for the medical and health sciences. Furthermore, the data were analyzed with an advanced statistical technique—segmented regression analysis—which can identify specific segments with similar growth rates in the history of science. The study is based on two different sets of bibliometric data: (a) the number of publications held as source items in the Web of Science (WoS, Thomson Reuters) per publication year and (b) the number of cited references in the publications of the source items per cited reference year. We looked at the rate at which science has grown since the mid-1600s. In our analysis of cited references we identified three essential growth phases in the development of science, which each led to growth rates tripling in comparison with the previous phase: from less than 1% up to the middle of the 18th century, to 2 to 3% up to the period between the two world wars, and 8 to 9% to 2010.

is to follow the published literature and infer from the growth of the literature the movement of ideas and associations between scientists" (p. 249). Price (1961, 1951, 1965) can undoubtedly be seen as a pioneering researcher on literature dynamics (de Bellis, 2009). Price analyzed the references listed in the 1961 edition of the *Science Citation Index* (*SCI*, Thomson Reuters) and the papers collected in the *Philosophical Transactions of the Royal Society of London*. His results show that science is growing exponentially (in a certain period by a certain percentage rate) and doubles in size every 10 to 15 years. The exponential growth in science established by Price has become today a generally accepted thesis which has also been confirmed by other studies (Tabah, 1999).

In this study, we reexamine the question of the growth of science. To do this we (a) use current data up to publication year 2012 and (b) analyze the data across all disciplines and also separately for the natural sciences and for the medical and health sciences. Furthermore, the data are analyzed with

The Future of Meta-Analysis

- Will likely increase in importance
  - Global scientific publication output may grow by up to 400% in the next 50 years

- The need for curation of findings is becoming clear

Lefebvre *et al. Systematic Reviews* 2013, **2**:78
http://www.systematicreviewsjournal.com/content/2/1/78

SYSTEMATIC REVIEWS

COMMENTARY **Open Access**

Methodological developments in searching for studies for systematic reviews: past, present and future?

Carol Lefebvre[1*], Julie Glanville[2], L Susan Wieland[3], Bernadette Coles[4] and Alison L Weightman[4]

Abstract

The Cochrane Collaboration was established in 1993, following the opening of the UK Cochrane Centre in 1992, at a time when searching for studies for inclusion in systematic reviews was not well-developed. Review authors largely conducted their own searches or depended on medical librarians, who often possessed limited awareness and experience of systematic reviews. Guidance on the conduct and reporting of searches was limited. When work began to identify reports of randomized controlled trials (RCTs) for inclusion in Cochrane Reviews in 1992, there were only approximately 20,000 reports indexed as RCTs in MEDLINE and none indexed as RCTs in Embase. No search filters had been developed with the aim of identifying all RCTs in MEDLINE or other major databases. This presented The Cochrane Collaboration with a considerable challenge in identifying relevant studies.

Over time, the number of studies indexed as RCTs in the major databases has grown considerably and the Cochrane Central Register of Controlled Trials (CENTRAL) has become the best single source of published controlled trials, with approximately 700,000 records, including records identified by the Collaboration from Embase and

The Future of Meta-Analysis

- Will likely increase in importance
  - Global scientific publication output may grow by up to 400% in the next 50 years

- The need for curation of findings is becoming clear

Ip *et al. Systematic Reviews* 2012, **1**:15
http://www.systematicreviewsjournal.com/content/1/1/15

SYSTEMATIC REVIEWS

COMMENTARY **Open Access**

A Web-based archive of systematic review data

Stanley Ip, Nira Hadar, Sarah Keefe, Christopher Parkin, Ramon Iovin, Ethan M Balk and Joseph Lau*

Abstract

Systematic reviews have become increasingly critical to informing healthcare policy; however, they remain a time-consuming and labor-intensive activity. The extraction of data from constituent studies comprises a significant portion of this effort, an activity which is often needlessly duplicated, such as when attempting to update a previously conducted review or in reviews of overlapping topics.

In order to address these inefficiencies, and to improve the speed and quality of healthcare policy- and decision-making, we have initiated the development of the Systematic Review Data Repository, an open collaborative Web-based repository of systematic review data. As envisioned, this resource would serve as both a central archive and data extraction tool, shared among and freely accessible to organizations producing systematic reviews worldwide. A suite of easy-to-use software tools with a Web frontend would enable researchers to seamlessly search for and incorporate previously deposited data into their own reviews, as well as contribute their own.

In developing this resource, we identified a number of technical and non-technical challenges, as well as devised a number of potential solutions, including proposals for systems and software tools to assure data quality, stratify and control user access effectively and flexibly accommodate all manner of study data, as well as means by which to govern and foster adoption of this new resource.

Herein we provide an account of the rationale and development of the Systematic Review Data Repository thus far, as well as outline its future trajectory.

Keywords: Archive, data repository, extraction, systematic review

The Future of Meta-Analysis

- Will likely increase in importance
  - Global scientific publication output may grow by up to 400% in the next 50 years

- The need for curation of findings is becoming clear

OPEN ACCESS Freely available online

PLOS | MEDICINE

Policy Forum

Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap

Julian H. Elliott[1,2]*, Tari Turner[2,3], Ornella Clavisi[4], James Thomas[5], Julian P. T. Higgins[6,7], Chris Mavergames[8], Russell L. Gruen[4,9]

1 Department of Infectious Diseases, Alfred Hospital and Monash University, Melbourne, Australia, 2 School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia, 3 World Vision Australia, Melbourne, Australia, 4 National Trauma Research Institute, Alfred Hospital, Melbourne, Australia, 5 EPPI-Centre, Institute of Education, University of London, London, England, 6 School of Social and Community Medicine, University of Bristol, Bristol, England, 7 Centre for Reviews and Dissemination, University of York, York, England, 8 Informatics and Knowledge Management Department, The Cochrane Collaboration, Freiburg, Germany, 9 Department of Surgery, Monash University, Melbourne, Australia

The Bridge from Evidence to Practice

Health research promises societal benefit by making better health possible. However, there has always been a gap between research findings (what is known) and health care practice (what is done), described as the "evidence-practice" or "know-do" gap [1]. The reasons for this gap are complex [2], but it is clear that synthesising the complex, incomplete, and at times conflicting findings of biomedical research into forms that can readily inform health decision making is an essential component of the bridge from "knowing" to "doing."

Summary

- The current difficulties in keeping systematic reviews up to date leads to considerable inaccuracy, hampering the translation of knowledge into action.
- Incremental advances in conventional review updating are unlikely to lead to substantial improvements in review currency. A new approach is needed.
- We propose living systematic review as a contribution to evidence synthesis that combines currency with rigour to enhance the accuracy and utility of health evidence.
- Living systematic reviews are high quality, up-to-date online summaries of health research, updated as new research becomes available, and enabled by improved production efficiency and adherence to the norms of scholarly communication.
- Together with innovations in primary research reporting and the creation and use of evidence in health systems, living systematic review contributes to an emerging evidence ecosystem.

The Current Environment

RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration*,†

+ See all authors and affiliations

Industrial and Organizational Psychology
PERSPECTIVES ON SCIENCE AND PRACTICE

Industrial and Organizational Psychology, **6** (2013), 252–268.

FOCAL ARTICLE

How Trustworthy Is the Scientific Literature in Industrial and Organizational Psychology?

SCIENCE

Psychology's Replication Crisis Can't Be Wished Away

It has a real and heartbreaking cost.

The Washington Post
Democracy Dies in Darkness

Monkey Cage

Does social science have a replication crisis?

The Current Environment

J Bus Psychol (2011) 26:105–121
DOI 10.1007/s10869-010-9185-2

Meta-analytic Decisions and Reliability: A Serendipitous Case of Three Independent Telecommuting Meta-analyses

Levi R. G. Nieminen · Jessica M. Nicklin · Tara K. McClure · Madhura Chakrabarti

Science Contents News Careers Journals

Meta-analyses were supposed to end scientific debates. Often, they only cause more controversy

By Jop de Vrieze | Sep. 18, 2018 , 4:15 PM

Original Investigation

The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses

JOHN P.A. IOANNIDIS

Meta-Analytic Choices and Judgment Calls: Implications for Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact

Herman Aguinis
Dan R. Dalton
Indiana University
Frank A. Bosco
Charles A. Pierce
University of Memphis
Catherine M. Dalton
Indiana University

Excerpt taken from Engber, D. (March, 2016). Everything is crumbling. *Slate.*

The Current Environment

- Meta-analysis is not immune from scrutiny
  - "All the old methods are in doubt. Even meta-analyses, which once were thought to yield a gold standard for evaluating bodies of research now seem somewhat worthless. "Meta-analyses are f*cked," Inzlicht warned me. If you analyze 200 lousy studies, you'll get a lousy answer in the end. It's garbage in, garbage out."
    - From:


SLATE

Health and Science has moved! You can find new stories here.

COVER STORY READ THIS FIRST. MARCH 6 2016 8:02 PM

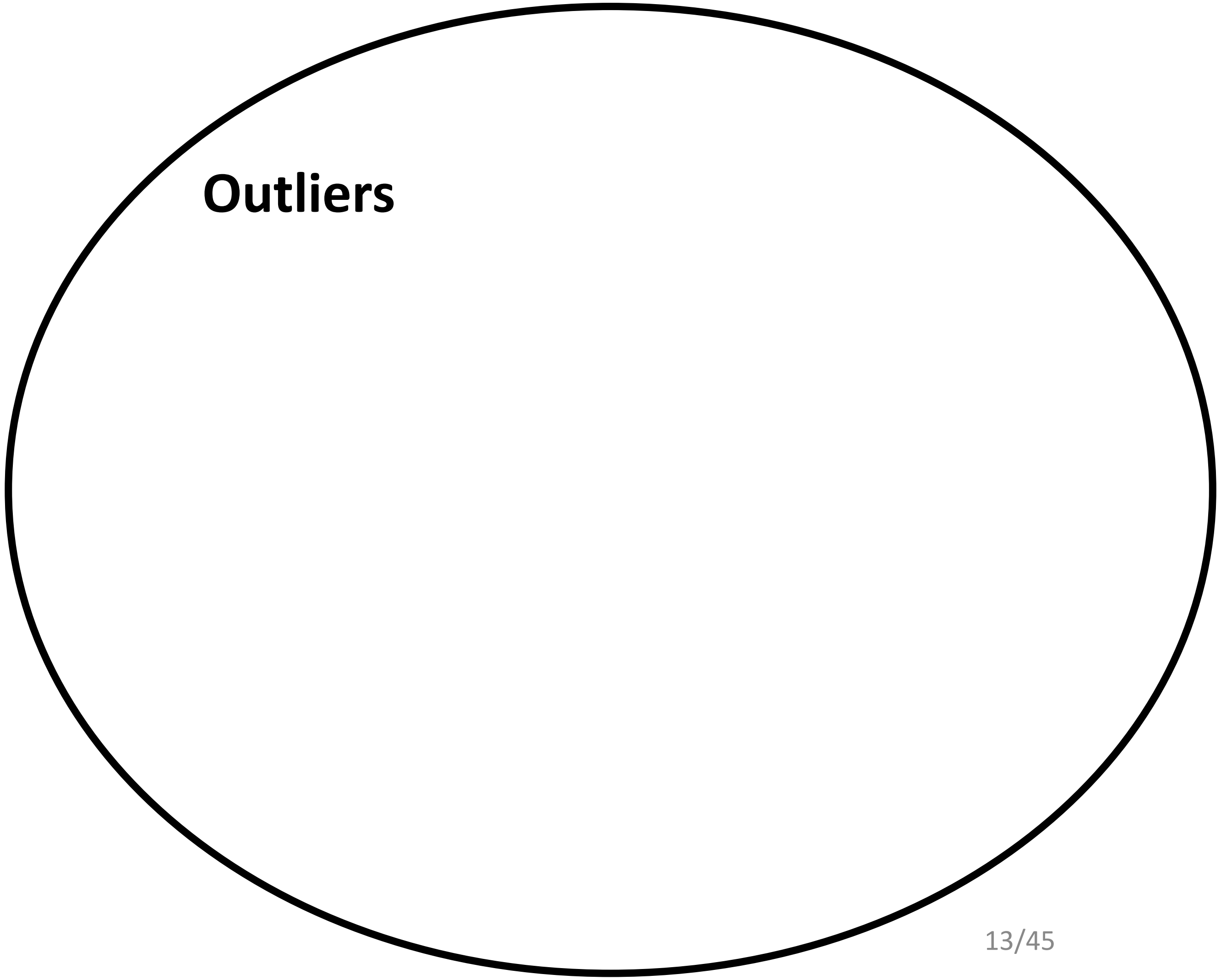
Everything Is Crumbling

An influential psychological theory, borne out in hundreds of experiments, may have just been debunked. How can so many scientists have been so wrong?

By Daniel Engber

- What could be driving opinions like these?

Threats to our Cumulative Knowledge



Threats to our Cumulative Knowledge

Outliers

An observation that appears "to deviate markedly from other members of the sample in which it occurs" (Grubbs, 1969, p. 1)

Threats to our Cumulative Knowledge

Outliers

Outcome-level causes
(e.g., effect size magnitude,
p-value)

Threats to our Cumulative Knowledge

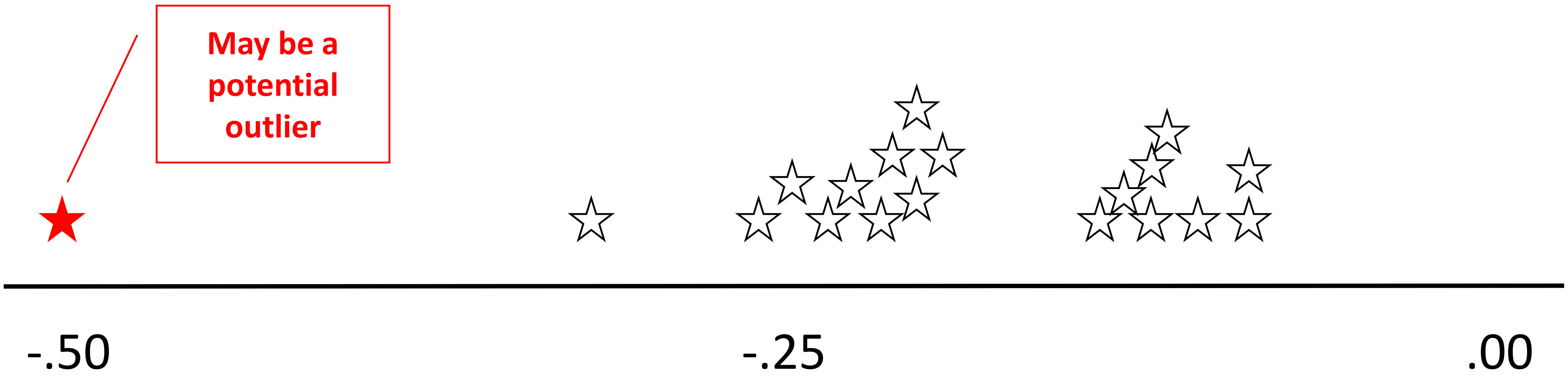
- Outcome-level causes of outliers: Effect size magnitude
  - Samples that have an effect size that diverges from all other samples in the dataset may need to be removed before performing a meta-analysis as they could introduce residual heterogeneity that may threaten its results and conclusions.

Threats to our Cumulative Knowledge

- Outcome-level causes of outliers: Effect size magnitude
  - Each ☆ represents an effect size in the Jiang et al. (2012) dataset (our running example)

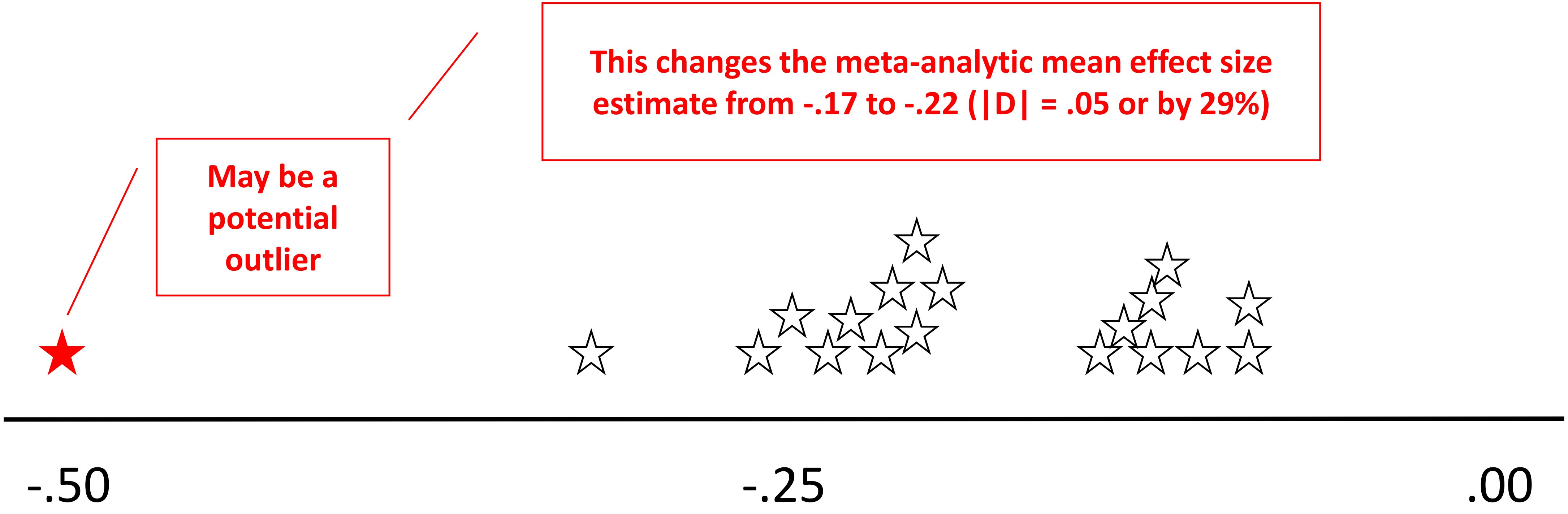
Threats to our Cumulative Knowledge

- Outcome-level causes of outliers: Effect size magnitude
  - Each ☆ represents an effect size in the Jiang et al. (2012) dataset (our running example)



Threats to our Cumulative Knowledge

- Outcome-level causes of outliers: Effect size magnitude
  - Each ☆ represents an effect size in the Jiang et al. (2012) dataset (our running example)



Threats to our Cumulative Knowledge

Outliers

Outcome-level causes
(e.g., effect size magnitude,
p-value)

Sample-level causes
(e.g., sample size,
sample type)

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample size
    - Given that 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, relatively large samples can have an undue influence on the meta-analytic mean.

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample size
  - The sample sizes included in Jiang et al.'s (2012) meta-analytic dataset range from 122 → 1,089

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample size
  - The sample sizes included in Jiang et al.'s (2012) meta-analytic dataset range from 122 → 1,089

Study ID	Sample size
1	177
2	306
3	208
4	474
5	122
6	164
7	809
8	112
9	323
10	205
11	222
12	142
13	9277
14	134
15	164
16	546
17	750
18	1,089

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample size
  - The sample sizes included in Jiang et al.'s (2012) meta-analytic dataset range from 122 → 1,089
  - Imagine adding an additional effect size that had a corresponding sample size of 50,000
  - Given that meta-analyses weight by precision, this addition would likely have a noticeable effect on the meta-analytic mean effect size estimate

Study ID	Sample size
1	177
2	306
3	208
4	474
5	122
6	164
7	809
8	112
9	323
10	205
11	222
12	142
13	9277
14	134
15	164
16	546
17	750
18	1,089

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: 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.

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample type
  - The sample types included in Jiang et al.'s (2012) meta-analytic dataset range are *fairly* similar (i.e., all are adults located in western countries)

Study ID	Sample type
1	Financial service employees (US)
2	Health care employees (US)
3	Business students (US)
4	Automotive employees (US)
5	Employed adults (US, UK)
6	Financial service employees (US)
7	Financial service employees (US)
8	Financial service employees (US)
9	Grocery store employees (US)
10	Public hospital employees (US)
11	Call center employees (US)
12	Call center employees (India)
13	Militaty (US)
14	Militaty (US)
15	Employed adults (Europe)
16	For profits organizations (Austrailia)
17	Employed adults (US)
18	Employed adults (US)

Threats to our Cumulative Knowledge

- Sample-level causes of outliers: Sample type
  - The sample types included in Jiang et al.'s (2012) meta-analytic dataset range are *fairly* similar (i.e., all are adults located in western countries)
  - Imagine adding an additional effect size from an *unusual* sample:
    - High school students located in Taiwan

Study ID	**Sample type**
1	Financial service employees (US)
2	Health care employees (US)
3	Business students (US)
4	Automotive employees (US)
5	Employed adults (US, UK)
6	Financial service employees (US)
7	Financial service employees (US)
8	Financial service employees (US)
9	Grocery store employees (US)
10	Public hospital employees (US)
11	Call center employees (US)
12	Call center employees (India)
13	Militaty (US)
14	Militaty (US)
15	Employed adults (Europe)
16	For profits organizations (Austrailia)
17	Employed adults (US)
18	Employed adults (US)

Threats to our Cumulative Knowledge

Outliers

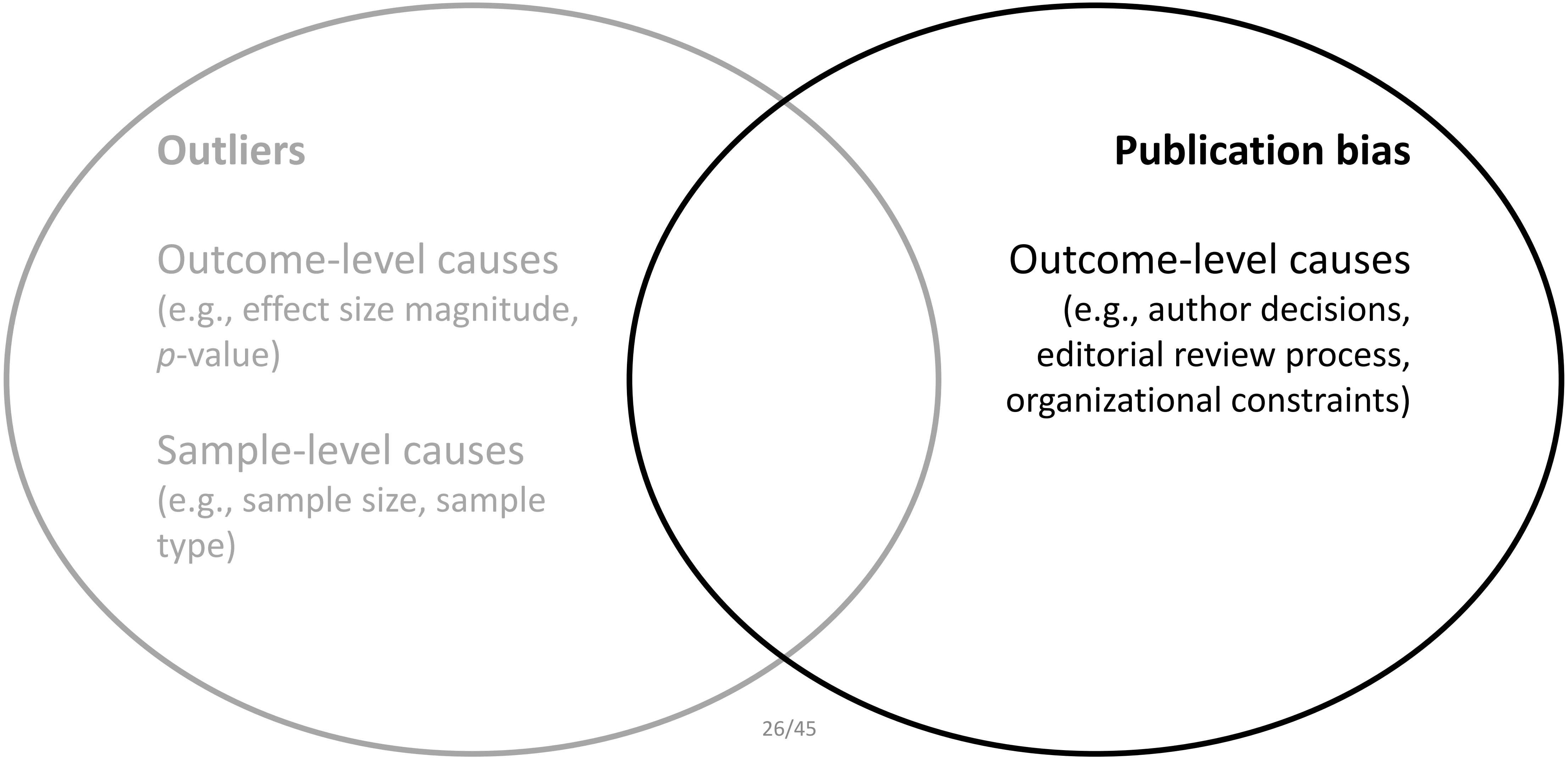
Outcome-level causes
(e.g., effect size magnitude, *p*-value)

Sample-level causes
(e.g., sample size, sample type)

Publication bias

A systematic suppression of research findings, which causes the available literature to be unrepresentative of all completed research on a relation of interest

Threats to our Cumulative Knowledge



Threats to our Cumulative Knowledge

Outliers

Outcome-level causes
(e.g., effect size magnitude,
p-value)

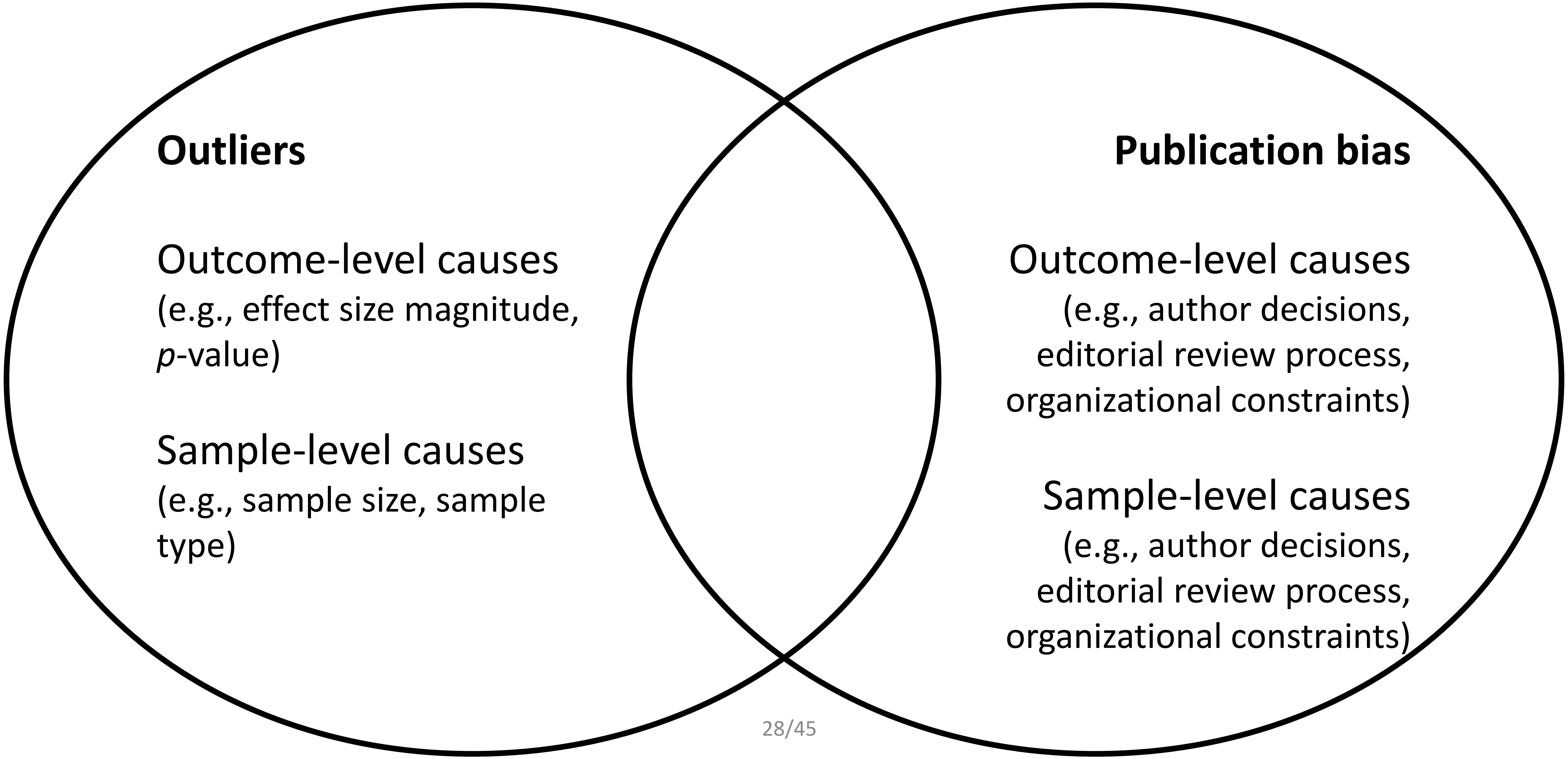
Sample-level causes
(e.g., sample size, sample
type)

Publication bias

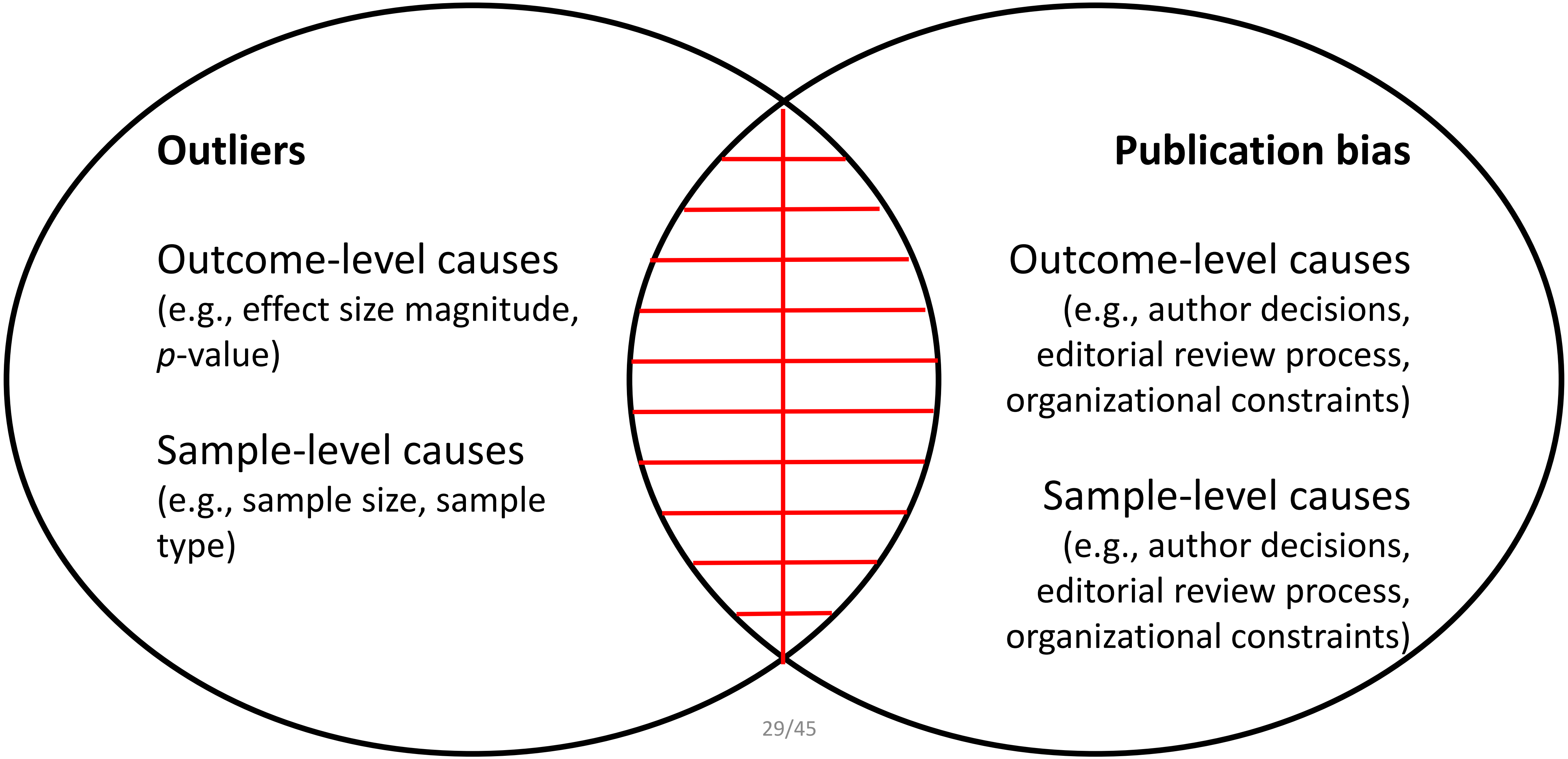
Outcome-level causes
(e.g., author decisions,
editorial review process,
organizational constraints)

Sample-level causes
(e.g., author decisions,
editorial review process,
organizational constraints)

Threats to our Cumulative Knowledge



Threats to our Cumulative Knowledge

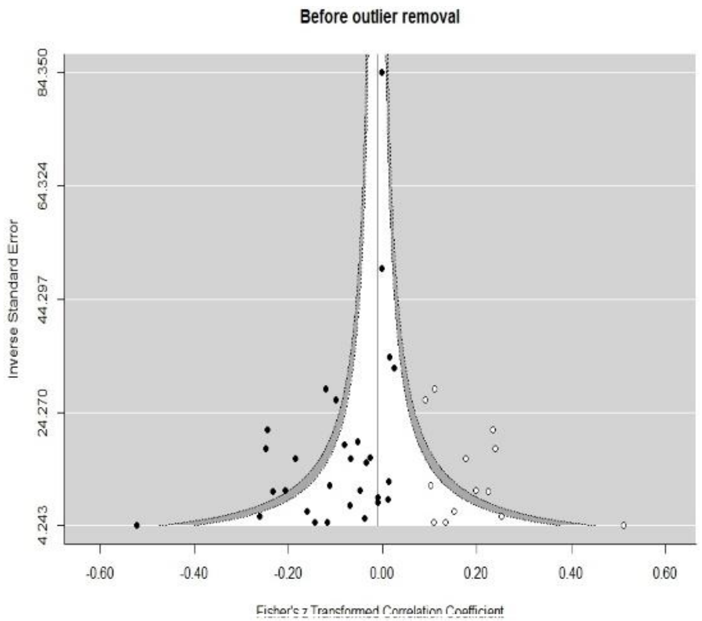


Threats to our Cumulative Knowledge

- Combined outlier *and* publication bias effect
  - Rarely tested!
  - However, outlier-induced heterogeneity may limit the efficacy of publication bias detection methods (Kepes & McDaniel, 2015; Peters et al., 2007; Terrin et al., 2003).
    - Some scholars have started to examine the possibility of a combined effect
      - (Field, Bosco, & Kepes, 2021, Kepes & McDaniel, 2015; Kepes et al., 2017; and Kepes & Thomas, 2018).

Threats to our Cumulative Knowledge

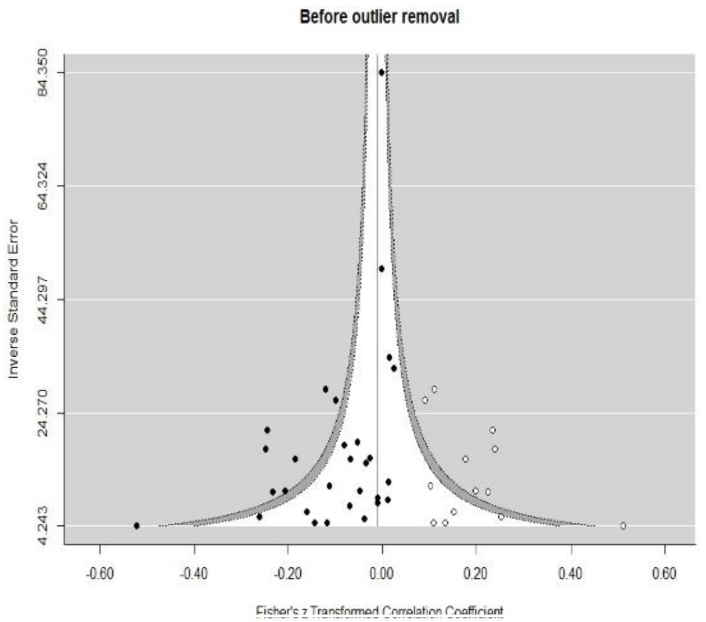
- Combined outlier *and* publication bias effect
  - k = 29
  - $\bar{r}_{o_{RE}}$ = -.08



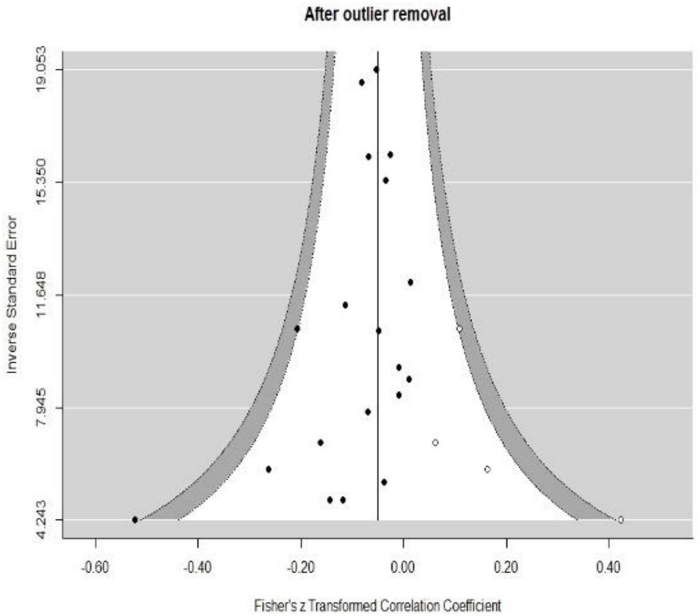
k = 29; ik = 13; t&f$_{FE}\bar{r}_o$ = -.01

Threats to our Cumulative Knowledge

- Combined outlier *and* publication bias effect
  - k = 29
  - $\bar{r}_{o_{RE}}$ = -.08



k = 29; ik = 13; t&f$_{FE}\bar{r}_o$ = -.01



k = 19; ik = 4; t&f$_{FE}\bar{r}_o$ = -.05

- Suggests that the publication bias detection result overestimates the distorting effect of publication bias!

What is Sensitivity Analysis?

- A sensitivity analysis examines the extent to which results and conclusions are altered as a result of changes in the data or analysis approach
  - Greenhouse & Iyengar (2009)

- If the conclusions do not change as a result of the sensitivity analysis, one can state that the conclusions are robust and one can have greater confidence in the conclusions.

What is Sensitivity Analysis?

- “Sensitivity analysis is the most powerful tool we have for assessing the influence of the specific choices made by the researchers”
  - Aytug, Rothstein, Zhou, & Kern (2012, p. 118)

What is Sensitivity Analysis?

- Sensitivity analyses are rarely conducted in meta-analyses in the organizational sciences
  - Kepes, McDaniel, Brannick, & Banks (2013)
- Because meta-analyses have a strong impact on our literatures, sensitivity analyses need to become much more common (and reported) in meta-analyses.

Live Meta-Sen Demonstration

- https://metasen.shinyapps.io/gen1/
- You can find some additional sample data files here: https://jamiefield.github.io/research/sma2021
- Citation for Meta-Sen app:

  Field, J. G., Bosco F. A., Kepes, S., (2021). How trustworthy is our cumulative knowledge on turnover? *Journal of Business and Psychology, 36*,. doi: 10.1007/s10869-020-09687-3.

How to minimize the impact of outliers & PB

- Ultimately, it is always best to report the range of results

- The effect of PB can be reduced by
  - Conducting extremely thorough literature reviews
  - Using research registries
  - Changing the journal review process
  - Altering author and organization norms
  - Obsessing less about theoretical contributions
  - Supporting data repositories like metaBUS

How to minimize the impact of outliers & PB

Step 1 → Step 2 → Step 3 → Step 4 → Step 5 → Step 6 → Step 7

- Step 1: Conduct a meta-analysis on original dataset

How to minimize the impact of outliers & PB

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses

How to minimize the impact of outliers & PB

Step 1 → Step 2 → Step 3 → Step 4 → Step 5 → Step 6 → Step 7

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses
- Step 3: Perform Viechtbauer and Cheung's (2010) influence diagnostics

How to minimize the impact of outliers & PB

Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses
- Step 3: Perform Viechtbauer and Cheung's (2010) influence diagnostics
- Step 4: If detected, remove outliers and repeat Steps 1 and 2. If outliers are not detected in Step 3, proceed directly to Step 5

How to minimize the impact of outliers & PB

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses
- Step 3: Perform Viechtbauer and Cheung's (2010) influence diagnostics
- Step 4: If detected, remove outliers and repeat Steps 1 and 2. If outliers are not detected in Step 3, proceed directly to Step 5
- Step 5: Report BRE and MRE (see Kepes et al., 2012)**

How to minimize the impact of outliers & PB

Step 1 → Step 2 → Step 3 → Step 4 → Step 5 → Step 6 → Step 7

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses
- Step 3: Perform Viechtbauer and Cheung’s (2010) influence diagnostics
- Step 4: If detected, remove outliers and repeat Steps 1 and 2. If outliers are not detected in Step 3, proceed directly to Step 5
- Step 5: Report BRE and MRE (see Kepes et al., 2012)**
- Step 6: Visually inspect the range of results before and after outlier removal

How to minimize the impact of outliers & PB

Step 1 › Step 2 › Step 3 › Step 4 › Step 5 › Step 6 › Step 7

- Step 1: Conduct a meta-analysis on original dataset
- Step 2: Perform osr and recommended PB analyses
- Step 3: Perform Viechtbauer and Cheung's (2010) influence diagnostics
- Step 4: If detected, remove outliers and repeat Steps 1 and 2. If outliers are not detected in Step 3, proceed directly to Step 5
- Step 5: Report BRE and MRE (see Kepes et al., 2012)**
- Step 6: Visually inspect the range of results before and after outlier removal
- Step 7: Assess the robustness of recommendations for practice

Thank you for attending today!

Remember…

You can find this presentation and some other potentially helpful resources at:

https://jamiefield.github.io/research/sma2021

Questions?
Comments?
Complaints?

Feel free to follow up with me…

james.field2@mail.wvu.edu

[@fieldjamie](https://twitter.com/fieldjamie)

jamiefield.github.io

In the following slides…

- I review two outlier and five publication bias assessment methods used by the Meta-Sen app
- I present results that illustrate an outlier and PB effect, as well as a combined effect of these phenomena, on meta-analytic findings on employee turnover

Review of Two Outlier Assessment Methods

- One form of sensitivity analysis is to conduct meta-analyses with and without outliers
- Only 3% of meta-analyses conduct outlier analyses (Aguinis et al., 2011)
  - Effect size outlier (large or small)
    - Graphical methods and statistical tests for outliers (e.g., SAMD statistic; Beal, Corey, & Dunlap, 2002)
  - Sample size outlier (large)
    - Sample sizes influence effect size weights in meta-analyses.

Review of Two Outlier Assessment Methods

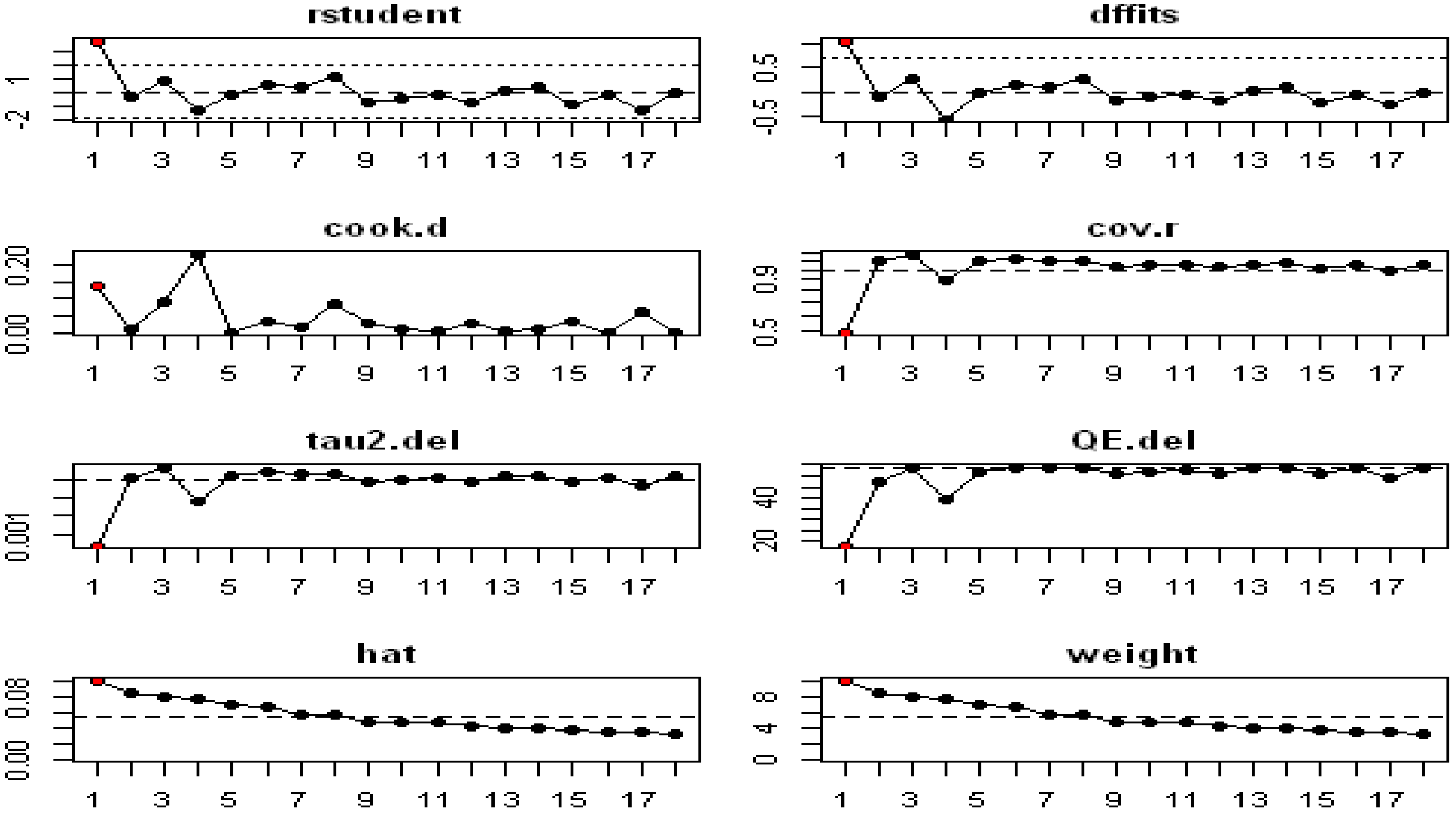
- One sample removed analysis:
  - Individual samples are removed one-by-one from the dataset and the point estimate is recalculated after each removal.
  - Thus, a one-sample removed analysis, yields *k*-1 meta-analytic mean estimates.
  - Given the Jiang, et al. (2012) dataset included 18 effect sizes, the one-sample removed analysis will produce 17 estimates of the meta-analytic mean
  - Important questions to ask:
    - How much does the distribution mean change when a given sample is excluded from the analysis?
    - Are the results due to a small number of influential samples?

Review of Two Outlier Assessment Methods

- Viechtbauer and Chueng's (2010; Viechtbauer [2015]) multivariate, multidimensional influence diagnostics:
  - A framework that calculates leave-one-out analyses for
    - externally standardized residuals
    - DFFITS value,
    - Cook's distance,
    - covariance ratio,
    - the leave-one-out amount of heterogeneity,
    - the leave-one-out test statistic for the test of heterogeneity, and
    - DFBETAS values.
  - In addition, an inspection of the hat matrix is examined for highly influential observations.

Review of Two Outlier Assessment Methods

- Viechtbauer and Chueng's (2010; Viechtbauer [2015]) multivariate, multidimensional influence diagnostics:



!!IMPORTANT!!

This is an iterative process that must be performed until *all* identified outliers are removed

Review of Five PB Assessment Methods

- Symmetry-based methods

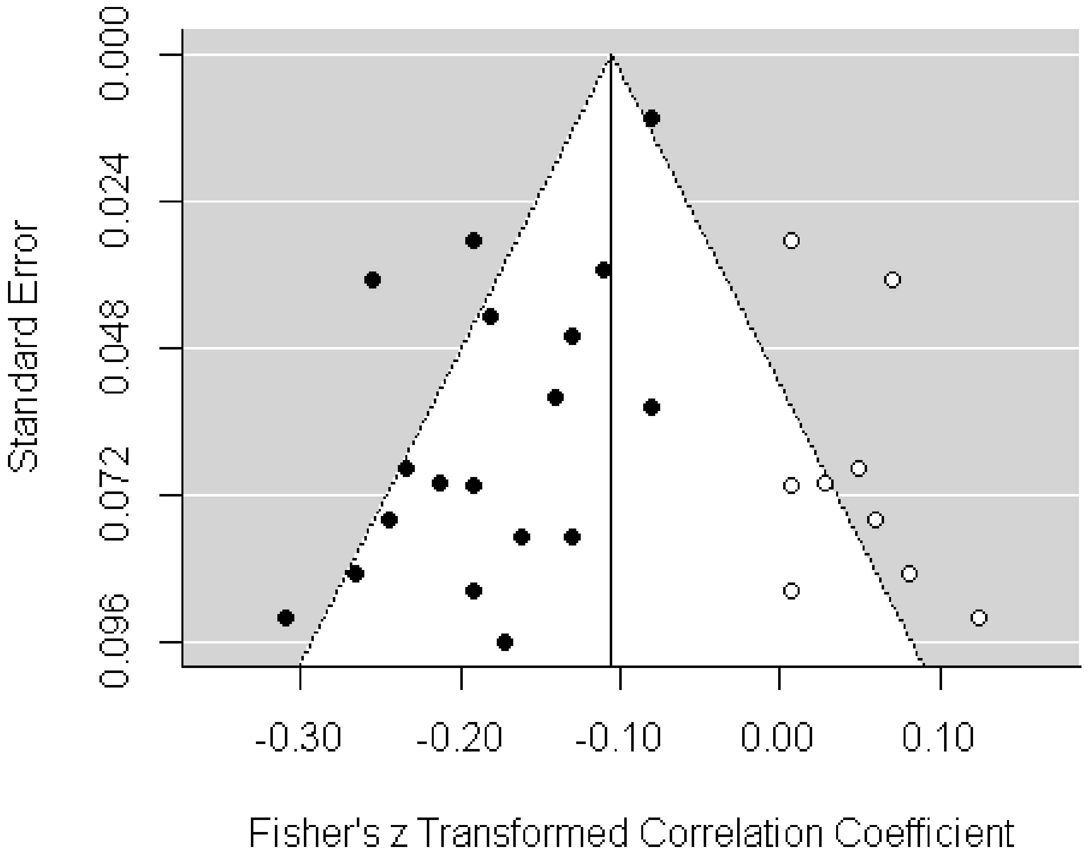
  - When sampling error is the sole source of variance, and the sampling distribution is symmetrical, then a funnel plot can be examined for symmetry.

  - A funnel plot is a plot of effect sizes by precision (1/standard error).

  - Examples of symmetry-based methods include (1) trim and fill models and (2) contour-enhanced funnel plot

Review of Five PB Assessment Methods

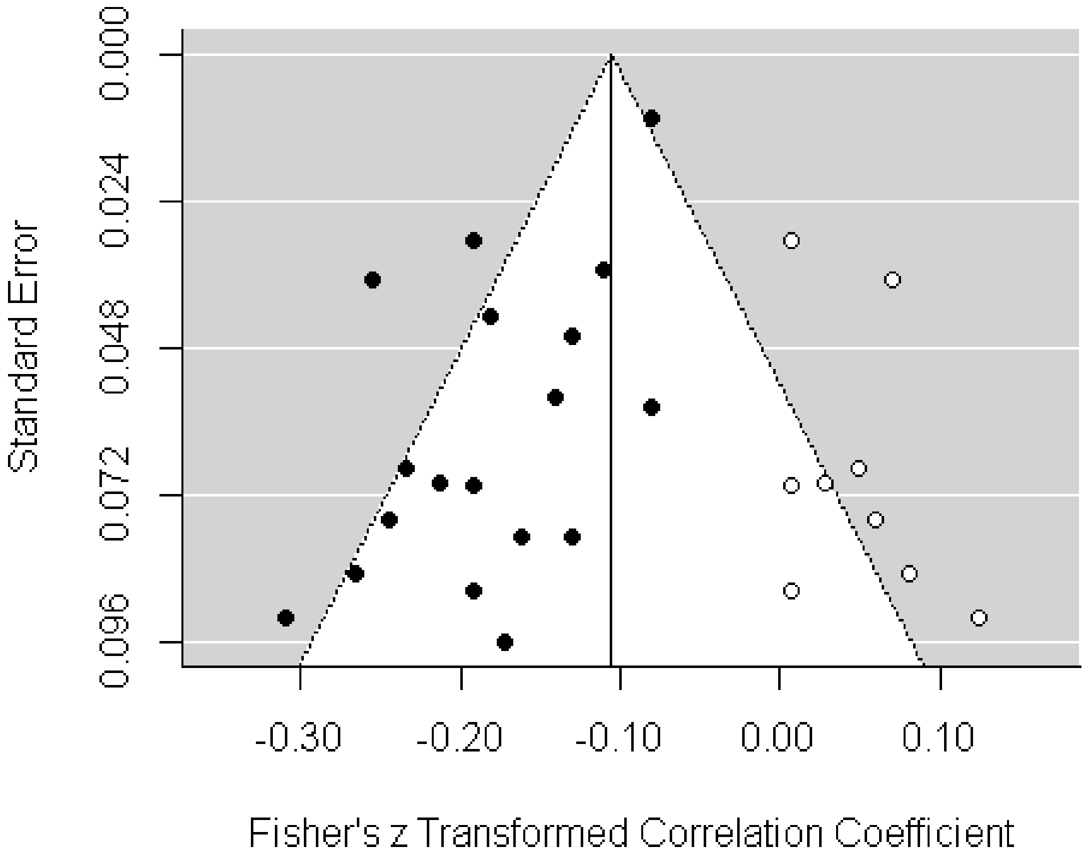
- Trim and fill models
  - The trim and fill method is probably the most useful symmetry based method in that it estimates what the population distribution would be if the missing studies were located
  - Analyses are re-conducted on the distribution containing both the observed data and the imputed data



FE trim and fill model of Jiang et al.'s meta-analytic distribution

Review of Five PB Assessment Methods

- Trim and fill models
  - It is unwise to consider this distribution of observed and imputed data as the "true" distribution
  - More reasonable to compare the observed mean with the trim and fill adjusted mean
  - If the mean drops from .45 to .15, one should worry about publication bias
  - But, one should not assume that .15 is the best estimate of the population mean



FE trim and fill model of Jiang et al.'s meta-analytic distribution

Review of Five PB Assessment Methods

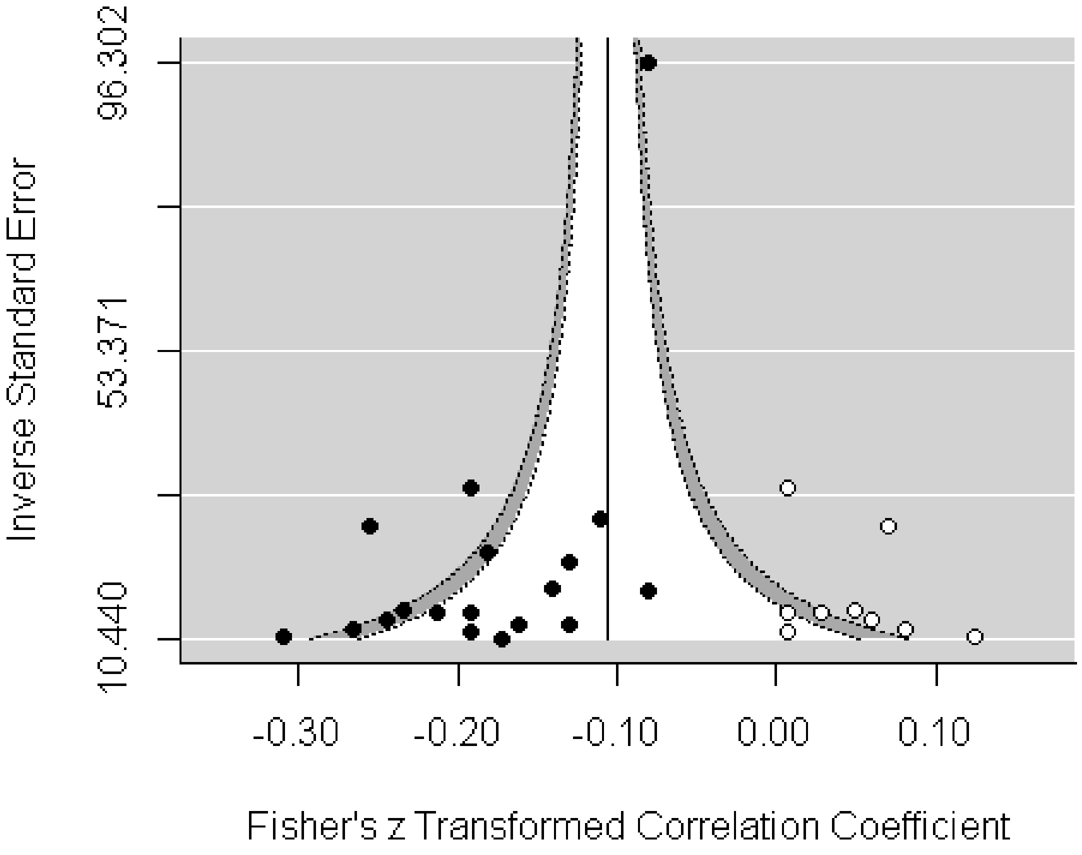
- Some asymmetry is not due to publication bias but to “small sample effects.”

  - A medicine may work best with the sickest (small N) patients and work less well with moderately sick (larger N) patients.

  - Small sample studies may yield larger effects due to better measures that are more difficult to collect in larger samples.

Review of Five PB Assessment Methods

- Contour-enhanced funnel plots
  - Related to the funnel plot and trim and fill i the contour-enhanced funnel plot, which displays graphically whether the imputed samples are a function of statistical significance (Peters et al., 2008).
    - Helps separate publication bias effects from "small sample effects."



Contour enhanced funnel plot of Jiang et al.'s meta-analytic distribution

Review of Five PB Assessment Methods

- A priori selection models

  - Selection models, also called weight-function models, originated in econometrics to estimate missing data at the item level.

  - Hedges and Vevea introduced the method to the publication bias literature
    - Hedges (1992)
    - Vevea and Hedges (1995)

  - Relatively robust to heterogeneity
    - Vevea and Woods (2005)

Review of Five PB Assessment Methods

- A priori selection models

  - As with trim and fill, selection models estimate what the population distribution would be if the missing studies were located and included in the meta-analytic distribution

  - When one is conducting a meta-analysis without regard to suppressed studies, one is implicitly assuming that one has 100% of the completed studies
    - This assumption is unlikely to be valid
      - Vevea and Woods (2005)

- Selection models permit you to make other assumptions

Review of Five PB Assessment Methods

- A priori selection models
  - Selection models assume that the probability that an effect size is included in a distribution is a function of a characteristic of that effect size
    - This characteristic is usually the level of statistical significance
  - Consider an *a priori* assumed selection model

Review of Five PB Assessment Methods

- A priori selection models
  - Selection models assume that the probability that an effect size is included in a distribution is a function of a characteristic of that effect size
    - This characteristic is usually the level of statistical significance
  - Consider an *a priori* assumed selection model

Significance level	Probability of being in the distribution
$p \leq .001$	100%
$.001 < p \leq .05$	90%
$.005 < p \leq .10$	70%
$p > .10$	30%

Review of Five PB Assessment Methods

- Cumulative meta-analysis by precision
  - Sort samples by sample size or precision
  - Conduct a meta-analysis starting with one effect size (the most precise effect) and add an additional effect size (with increasingly less precision) with each iteration of the meta-analysis
  - Inspect the meta-analytic means for drift

Review of Five PB Assessment Methods

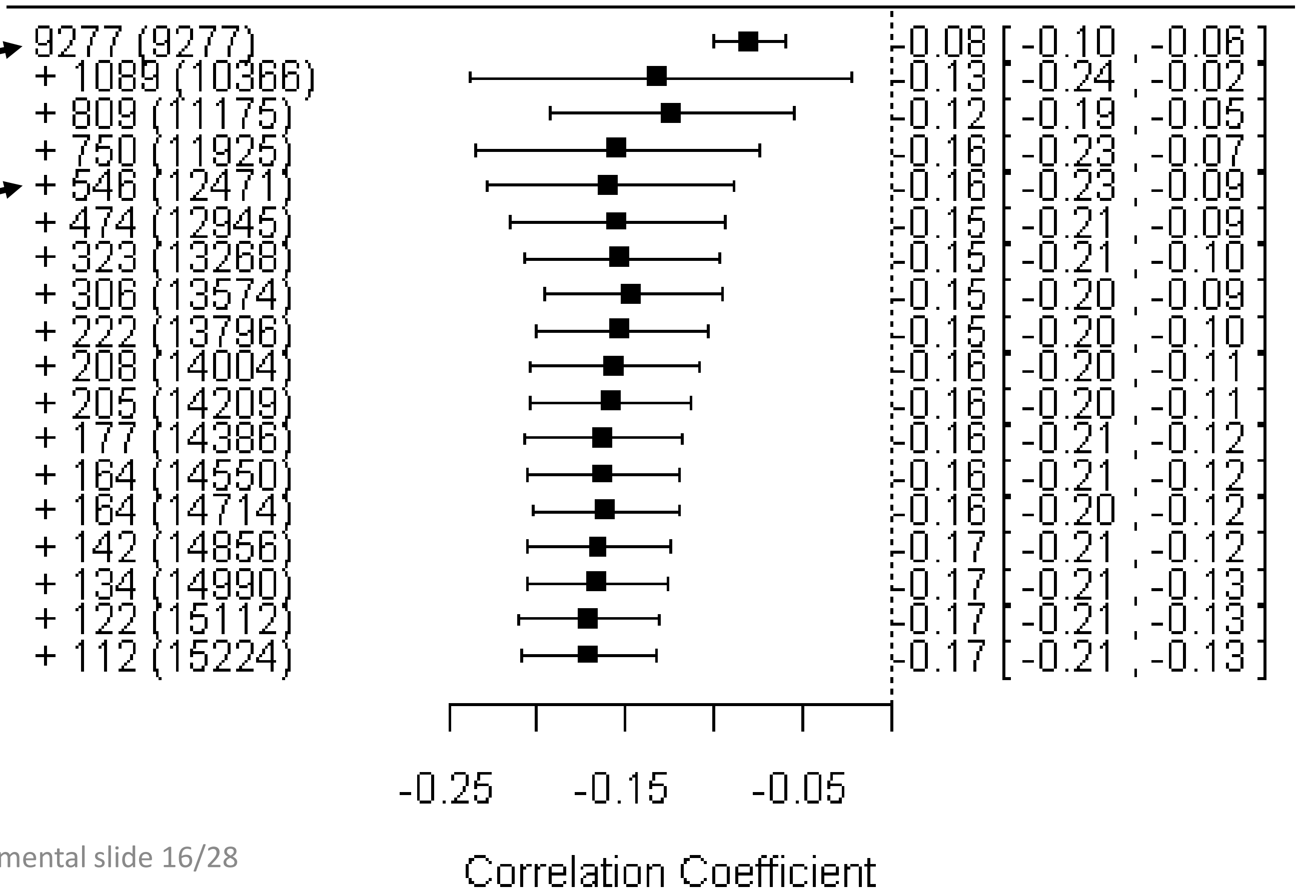
- Cumulative meta-analysis by precision

CMA by precision of Jiang et al.'s meta-analytic distribution

The most precise sample (N = 9,277), has an effect size of -.08.

With five studies, the cumulative sample size is 12,471 and the mean effect size is -.16

By the time one gets to 18 studies (N = 15,224), the mean effect size is -.17



Review of Five PB Assessment Methods

- Cumulative meta-analysis by precision
  - Gives similar results to that obtained in symmetry based methods
  - When symmetry analyses suggest small effects are suppressed, cumulative meta-analysis will show a drift toward larger effects
  - When symmetry analyses suggest larger effects are suppressed, cumulative meta-analysis will show a drift toward smaller effects.

Review of Five PB Assessment Methods

- Precision-effect test-precision effect estimate with standard error analysis (PET-PEESE)
  - A relatively new PB detection technique (Stanley and Doucouliagos, 2014)
  - This method is a combination of two regression models (PET and PEESE)
    - Conditional decision rule that determines which of the two models should be used
  - PET → Observed effect sizes are regressed on their corresponding standard errors using meta-regression techniques
  - PEESE → Observed effect sizes are regressed on their corresponding *squared* SE

Evidence of Combined Outlier and PB Effect

- Comprehensive sensitivity analyses were conducted on 201 recently published meta-analytic distributions on employee turnover
  - Examined the trustworthiness of these distributions
  - Does a greater threat to the trustworthiness arise from outliers or publication bias?
  - Assessed if meta-analytic and PB results changed after outlier removal
  - Examined whether or not recommendations for practice were robust to outliers and/or PB

Evidence of Combined Outlier and PB Effect

- **How trustworthy is our cumulative scientific knowledge on turnover?**
  - 95% (190/201) of the turnover distributions were misestimated by a “noticeable” amount (i.e., > 20%; Kepes et al., 2012)

Evidence of Combined Outlier and PB Effect

- **Does a greater threat to the trustworthiness arise from outliers or publication bias?**
  - PB was the source of non-robustness in 95% (190/201) of the turnover meta-analytic mean effect size estimates
  - Outliers was the source of the non-robustness in 69% (139/201) of the turnover meta-analytic mean effect size estimates
  - Therefore, PB > outliers

  - A combined outlier *and* PB effect was observed in 69% (138/201) of the turnover distributions

Evidence of Combined Outlier and PB Effect

- **Do outliers distort meta-analytic results?**
  - 59% (121/201) of the turnover distributions contained at least one outlier
    - 11 distributions had $k < 10$ and, thus, could not be reanalyzed after outlier removal
  - For the 110 distriubutions that could be compared, our results suggest that 88% (97/110) of the meta-analytic mean effect size estimates changed after outlier removal
    - 45% (49/110) were misestimated by more than 20%

Evidence of Combined Outlier and PB Effect

- **Do outliers distort publication bias results?**
  - For the 110 turnover distributions that could be compared::
    - t&f$_{FE}$ $\bar{r}_o$ remained unchanged following outlier removal in only 31% (34/110) of the cases
    - t&f$_{RE}$ $\bar{r}_o$ remained unchanged following outlier removal in only 15% (16/110) of the cases
    - *pr* $\bar{r}_o$remained unchanged following outlier removal in only 66% (73/110) of the cases
    - *pp* $\bar{r}_o$ remained unchanged following outlier removal in only 19% (21/110) of the cases
    - sm$_{\mathrm{m}}$ $\bar{r}_o$ remained unchanged following outlier removal in only 5% (6110) of the cases

Evidence of Combined Outlier and PB Effect

- **Do recommendations for practice change after accounting for outliers and PB?**
  - Hancock et al. (2013) recommended that most organizations should increase their investments in reducing turnover
  - Estimated that a one SD decrease in turnover would be associated with a $352 million increase in profits for *Fortune* 1,000 companies
  - However, out results suggest that this may be dramatically overestimated
  - Our FE trim and fill mean estimate following outlier removal (*k* = 46, t&f$_{FE}\,\bar{r}_o$ = -.02 suggests that a one SD decrease in turnover would be associated with a $101 million increase in profits
    - Suggests that the originally estimated financial benefit of a reduction in turnover may be overestimated by $251 million (or 249%)

Evidence of Combined Outlier and PB Effect

- **Do recommendations for practice change after accounting for outliers and PB?**
  - We found that 33% (14/43) of the recommendations for practice were *not* robust to outliers and publication bias

  - Specifically, at least one of the following three occurred after taking into account the effect of outliers and/or PB
    - The direction of the meta-analytic mean used to justify the recommendation changed
    - The magnitude of he meta-analytic mean used to justify the recommendation changed by at least 20%
    - A moderating effect used to justify the recommendation disappeared

Convergence of PB Detection Methods

Convergence Rates Regarding Practical Differences Before and After Outlier Removal for 110 Turnover Distributions

	Before outlier removal								After outlier removal		
PB method	Negligible	Moderate	Severe	1.	2.	3.	4.	5.	Negligible	Moderate	Severe
1. t&f$_{FE}\,\bar{r}_o$	28 (23%)	17 (14%)	76 (63%)	-	56 (46%)	50 (41%)	76 (63%)	94 (78%)	35 (29%)	24 (20%)	62 (51%)
2. t&f$_{RE}\,\bar{r}_o$	68 (56%)	20 (17%)	33 (27%)	54 (45%)	-	75 (62%)	55 (45%)	43 (36%)	74 (61%)	29 (24%)	18 (15%)
3. sm$_m\,\bar{r}_o$	49 (40%)	28 (23%)	44 (36%)	60 (50%)	63 (52%)	-	59 (49%)	45 (37%)	73 (60%)	28 (23%)	20 (17%)
4. *pr* $\bar{r}_o$	35 (29%)	23 (19%)	63 (52%)	68 (56%)	48 (40%)	65 (54%)	-	74 (61%)	36 (30%)	34 (28%)	51 (42%)
5. *pp* $\bar{r}_o$	21 (17%)	18 (15%)	82 (68%)	94 (78%)	53 (44%)	54 (45%)	64 (53%)	-	23 (19%)	24 (20%)	74 (61%)

Convergence of PB Detection Methods

Convergence Rates Regarding Practical Differences Before and After Outlier Removal for 110 Turnover Distributions

	Before outlier removal								After outlier removal		
PB method	Negligible	Moderate	Severe	1.	2.	3.	4.	5.	Negligible	Moderate	Severe
1. t&f$_{FE}\,\overline{r}_o$	28 (23%)	17 (14%)	**76 (63%)**	-	56 (46%)	50 (41%)	76 (63%)	94 (78%)	35 (29%)	24 (20%)	**62 (51%)**
2. t&f$_{RE}\,\overline{r}_o$	68 (56%)	20 (17%)	**33 (27%)**	54 (45%)	-	75 (62%)	55 (45%)	43 (36%)	74 (61%)	29 (24%)	**18 (15%)**
3. sm$_{\mathrm{m}}\,\overline{r}_o$	49 (40%)	28 (23%)	**44 (36%)**	60 (50%)	63 (52%)	-	59 (49%)	45 (37%)	73 (60%)	28 (23%)	**20 (17%)**
4. *pr* $\overline{r}_o$	35 (29%)	23 (19%)	**63 (52%)**	68 (56%)	48 (40%)	65 (54%)	-	74 (61%)	36 (30%)	34 (28%)	**51 (42%)**
5. *pp* $\overline{r}_o$	21 (17%)	18 (15%)	**82 (68%)**	94 (78%)	53 (44%)	54 (45%)	64 (53%)	-	23 (19%)	24 (20%)	**74 (61%)**

Convergence of PB Detection Methods

- Based on our turnover and metaBUS results we recommend that future meta-analysts use the following to triangulate the potentially most robust estimate of the "true" meta-analytic effect size
  - FE trim and fill model
  - CMA by precision
  - PET-PEESE analysis

- For outlier detection, we recommend Viechtbauer and Cheung's (2010; Viechtbauer 2015) influence diagnostics procedure due to its statistical rigor