Cumulative Meta-Analysis as a Publication Bias Method

Michael A. McDaniel
Virginia Commonwealth University

Paper presented at the 24th Annual Meeting of the Society for Industrial and Organizational Psychology. New Orleans. April. 2009.

Dr. Michael Borenstein (www.meta-analysis.com) suggested to me the use of cumulative meta-analysis as a publication bias tool. Without his insight, this paper would not have been written. The paper builds on work by Jeffery Pollack, Brian Tate, and Ernest O'Boyle. Correspondence on this paper can be directed to the author by e-mail: mamcdani@vcu.edu

This paper reviews the use of cumulative meta-analysis as a publication bias tool. The rationale of the method is explained and three applications of the method are offered. Publication bias exists when primary study results available to a reviewer systematically differ from all primary study results (McDaniel, Rothstein, & Whetzel, 2006). Thus, the phenomena might be better called "availability bias" but the term publication bias has taken root and will be used in this paper. The existence of publication bias threatens the conclusions that can be drawn from meta-analytic reviews.

Typically, studies that fail to find significant results are "suppressed" in favor of studies that report significant results (Dickersin, 2005). The preference for the publication of statistically significant results is often due to editorial preferences and actions of authors. For example, editors and authors often consider statistically significant results to be more interesting than results that do not reach statistical significance. Further, journals have limited space and may give preference to the more interesting articles. Finally, authors may tailor their publications to editorial preferences.

Publication bias can also be a function of intentional distortion. Some studies (McDaniel, et al., 2006; Pollack & McDaniel, 2008) presented evidence consistent with the conclusion that some test publishers may intentionally distort their validity coefficients, such that small magnitude validity coefficients are suppressed (e.g., do not appear in technical manuals). To the extent that such data suppression occurs, it makes the test vendors' products look more useful than they are.

Although examination of potential publication bias in management and I/O psychology is rare, publication bias analyses are routinely conducted in medical research and are given high visibility in news coverage. For example, the New York Times reported:

> "The drug maker Pfizer earlier this decade manipulated the publication of scientific studies to bolster the use of its epilepsy drug Neurontin for other disorders, while suppressing research that did not support those uses according to experts who reviewed thousands of company documents for plaintiffs in a lawsuit against the company. …. Dr. Dickersin, the Johns Hopkins expert, said that of 21 studies she reviewed, five were positive and 16 negative, meaning they did not prove the drug was effective. Of the five positive studies, four were published in full journal articles, yet only six of the negative studies were published and, of those, two were published in abbreviated form." (S. Saul, October 8, 2008. *The New York Times*).

The goal of this paper is to introduce cumulative meta-analysis as a publication bias method to the I/O and management research community. In a cumulative meta-analysis, studies are sorted by a variable of interest, often time. One then conducts iterative meta-analyses adding one additional effect size (e.g., correlation or mean difference) for each meta-analysis. The first mean reported is the effect size from the first study. The second mean is the mean from the meta-analysis of the first and second study. The third mean is the mean of the meta-analysis of the first three studies, and so on. Historically, cumulative meta-analysis has been used to determine the time point at which a result stabilizes.

One of the most prominent examples of cumulative meta-analysis involved the streptokinase (a blood thinner) treatment of myocardial infarction (Lau, Schmid, & Chalmers, 1995). In that analysis, the studies were sorted by time of publication and the meta-analysis was iteratively conducted each time adding in one effect size. Lau et al. (1995) found that although randomized clinical trials continued until 1989, the streptokinase treatment could have been deemed an effective treatment as early as 1973. The sixteen year gap between when the drug should have been implemented as a standard therapy and when the last clinical trial was completed likely caused unnecessary deaths.

When cumulative meta-analysis is used as a publication bias method, studies are sorted by standard error from low to high (or alternatively by sample size from high to low). Low standard error studies are those with the largest sample size. For correlations and standardized mean differences, these two indices are very highly correlated (e.g., often around .99). Effect sizes (e.g., correlations and standardized mean differences) with large samples sizes (and lower standard errors) are more precise because the effects have narrow confidence intervals. Typically in meta-analysis, such studies are given greater weight because of their precision and likely have greater information value.

Once the effect sizes have been sorted from high to low on precision, one then conducts iterative meta-analyses adding one additional effect size for each iteration of the meta-analysis. This results in a series of cumulative mean estimates, each based on one more effect size than the previous mean. The cumulative means can be examined and plotted for evidence of drift as more studies are added to the meta-analysis. The meta-analytic means from the studies entered into the iterative meta-analyses early are the estimates of the population mean from the larger samples. The meta-analytic means added in later stages of the iterative meta-analysis are from the addition of the smaller samples to a distribution that already contains the larger samples. If small sample size studies with small effects are being suppressed (a common publication bias scenario), the cumulative means will drift in a more positive direction as the smaller sample size studies are added to the cumulative meta-analysis. This occurs because the small sample size studies available to the analyst will tend to have larger magnitude effects than the large sample studies available to the analyst. Figure 1 shows an illustrative cumulative meta-analysis where effect sizes have been sorted by their standard error from low to high. Note that the small standard error effect sizes (i.e., the effect sizes from large sample size studies) yield the cumulative means at the top of the graph and have a magnitude of about .10. However, as small sample studies are added, the cumulative mean shifts closer to .20. This suggests that the smaller samples have larger effect sizes than the large *N* samples, consistent with a conclusion of publication bias.

Method

Cumulative meta-analysis was applied to three data sets related to conscientiousness to illustrate the use of the method and to evaluate its consistency with another publication bias method called trim and fill (Duval, 2005). The first analysis is based on data analyzed by Pollack and McDaniel (2008) that had examined potential publication bias in the PreVisor™ Employment Inventory. In the author's judgment, this test is primarily an assessment of conscientiousness. The second analysis uses data for the Personal Characteristics Inventory (PCI) conscientiousness scale (Wonderlic, 2002). The third analysis is based on data from Tate and McDaniel (2008) and examines Black-White mean differences in conscientiousness. The analyses were conducted using the Comprehensive Meta-Analysis (CMA) 2.0 software (Borenstein, Hedges, Higgins, & Rothstein, 2005).

Results

Example 1:Validity coefficients for the PreVisor™ Employment Inventory

Pollack and McDaniel (2008) examined the validity data in the Employment Inventory technical manual (Paajanen, Hansen & McLellan, 1993) for potential publication bias. Figure 2a (adapted from Pollack and McDaniel, 2008, dependability rating criterion) is a funnel plot showing that the validity data for the Employment Inventory are not symmetrical. In the current analysis, the Y axis is precision and the X axis is the Fisher z transformation of the validity coefficients (See McDaniel et al. [2006], for an overview of funnel plots and trim and fill). Specifically, there are very few small sample studies with low magnitude validity coefficients (the lower side of the funnel plot). This plot would be consistent with an inference that such studies were suppressed, although other inferences are possible. Figure 2b shows the funnel after the trim and fill procedure imputed 22 studies that were needed to bring the distribution into symmetry.

Figure 3 shows a cumulative meta-analysis of the Employment Inventory data in which studies are sorted by precision (low to high standard errors). The cumulation of the largest four studies, with a cumulative *N* over 5,000, yielded a mean validity of .216. Many would think one could get a fairly accurate estimate of the validity of a test based on 5,000 cases. However, as less precise studies are added, the validity shows a slight drift higher such that by the time the cumulative *N* reaches 10,000 (the 14 largest sample size studies), the mean validity is .233. At 41 studies included with a cumulative sample size of 15,000, the mean is at .242. By the time one includes all the data (70 studies, *N* = 16,941), the mean drifts to .249. Although the drift is not dramatic, these results are consistent with the conclusion of the Pollack and McDaniel's (2008) trim and fill analysis (see Figure 2 in the present paper). Both analyses are consistent with the inference that small sample, small magnitude validity studies may not have been included in the technical manual. If this were true, the technical manual data would be overstating the actual validity of the test. As noted by Pollack and McDaniel (2008), other inferences are possible. For example, the validities may be influenced by one or more moderators that co-vary with sample size.

Example 2: Validity coefficients for the Wonderlic PCI conscientiousness scale

The manual for the Wonderlic PCI lists 14 validity coefficients for the conscientiousness scale. The right graph in Figure 4 displays the trim and fill results. Although the trim and fill analysis imputed four additional studies, the mean shifted only from .23 to .22 suggesting no or minimal publication bias. The cumulative meta-analysis, the left graph in Figure 4, shows very little drift in the means as lower precision studies are added. Thus, both publication bias analyses of the PCI conscientiousness validity data reach the same conclusion that there is no to minimal evidence of publication bias in these data.

Example 3:Black-White standardized mean differences in conscientiousness

Tate and McDaniel (2008) presented evidence of potential publication bias in published studies of mean racial differences in personality. To explain the logic of the Tate and McDaniel analysis, we first offer a discussion of some simulated data. Figure 5a displays the simulated results of multiple samples of Black-White standardized mean differences drawn from a population with a mean of zero. The black circles are effect sizes indicating that Blacks score more favorably on conscientiousness than Whites. The white circles are effect sizes indicating that Whites score more favorably on conscientiousness than Blacks. The population mean is invariant at zero and the sample effect sizes diverge from zero due to random sampling error.

Figure 5b shows the subset of observed validity coefficients from Figure 5a that would appear in published studies if all authors reported results that favor Blacks but did report results that favor Whites. The result of this decision rule is that the published studies would report a meta-analytic mean racial difference indicating that Blacks show greater conscientiousness than Whites. Publication bias analyses of such data would suggest that effect sizes favoring Whites are missing from the distribution.

Figure 5c shows the effect sizes favoring Whites that could have been presented in published studies but were not. If one wrote the authors of the published studies and obtained the mean racial difference results not reported in their publication, the results for these obtained but never published data would look like the data reported in Figure 5c. The meta-analytic mean of these data would show that the mean racial differences in conscientiousness favor Whites. Publication bias analyses on the effect sizes in Figure 5c would suggest that effect sizes favoring Blacks are missing from the distribution.

Tate and McDaniel (2008) applied this reasoning to potential publication bias in published studies. One set of data were those drawn from published studies that reported mean racial differences in conscientiousness. The second set of data was obtained from authors of published studies in which the mean racial differences in conscientiousness were not reported in the published study. Figure 6a displays graphics from the analysis of the published studies in which the mean racial differences were reported in the published article. The observed mean *d* of -.07 indicated that Blacks, on average, were slightly more conscientious than Whites. The trim and fill analysis imputed three studies (favoring Whites) to bring the distribution into symmetry and moved the mean in the direction of being more favorable to Whites (-.07 to .04). This would be consistent with the hypothesis that mean racial differences are more likely to be reported when the mean differences favor Blacks. The cumulative meta-analysis shows that the larger sample, more precise studies show effect sizes that have positive *d* values (favoring Whites) and that as small samples are added to the analysis the cumulative means drift to the left such that the cumulative mean for all studies is negative (favoring) Blacks. Thus, both sets of analyses suggest that there is a publication bias in journals that favors the publication of mean racial differences in conscientiousness when the results indicate Blacks have more conscientiousness than Whites. Although the evidence is consistent with an inference of publication bias, one would still likely conclude that the Black-White mean difference in conscientiousness is small.

Figure 6b displays graphics from the analysis of the published studies where the mean racial differences were <u>not</u> reported in the published article. That is, these data were obtained from the authors of the published articles who did report these results in their published articles. The trim and fill analysis imputed two studies (favoring Blacks) to bring the distribution into symmetry with the *d* moving from .00 to -.14 (favoring Blacks). This would be consistent with the hypothesis that mean racial differences favoring Whites in conscientiousness are less likely to be reported in published studies. The cumulative meta-analysis shows that the larger sample, more precise studies show negative effect sizes (favoring Blacks) and that as smaller samples are added to the analysis the cumulative means drift to the right (in the direction more favorable to Whites) such that the cumulative mean for all studies, although not favoring Whites, does reach zero. Thus, both graphics in Figure 6b suggest that there is publication bias in this distribution of data obtained from journal authors who did not include their results in their published papers. These mean differences that were not presented in the published articles tended to show mean differences that favor Whites.

Thus, the analyses shown in Figures 6a and 6b are consistent with the inference that mean racial differences in conscientiousness are more likely to be presented in published studies when they favor Blacks than when they favor Whites. Despite this consistent publication bias effect, the mean racial differences are relatively small no matter how one cuts the data and the best conclusion is that the mean difference between Blacks and Whites on conscientiousness is small.

Discussion

This paper has introduced cumulative meta-analysis as a publication bias method to the I/O and management research community. It was demonstrated in three data sets that the results yield similar conclusions to those drawn from trim and fill publication bias analyses (Duval, 2005). Typically, conclusions that are consistent across two analysis methods are given greater credibility than results based on only one analysis method. In brief, both publication bias analyses yield evidence consistent with an inference of publication bias in the PreVisor™ Employment Inventory. In contrast, for the Wonderlic PCI conscientiousness scale, the two publication bias methods yielded results consistent with the inference of no publication bias. Finally, for journal publications from data sets containing information on Black-White differences in conscientiousness, both publication bias methods suggest that there is a tendency for published journal articles to suppress data on differences that favor Whites. However, the publication bias analyses do not alter the conclusion that mean racial differences in conscientiousness are small.

This paper also makes two other methodological contributions to the publication bias literature. First, the paper highlights the Tate and McDaniel (2008) insight that inferences concerning publication bias can be informed by comparing results from published studies with results based on data obtained from the authors of published studies where the results of interest were not presented in the published studies. Second, the paper describes two measures that, in the author's opinion, assess conscientiousness and the comparison of the differing publication bias conclusions has import for alternative explanations for the findings. The analysis of the PreVisor™ Employment Inventory data yielded results consistent with an inference of publication bias, but the analysis of the Wonderlic PCI conscientiousness measure does not. An alternative explanation for the Employment Inventory results would be the existence of a moderator that co-varies with sample size. For example, for the Employment Inventory data, one could speculate that smaller sample studies have larger validities than the larger sample studies because the former have greater standardization (e.g., training of raters who provide criterion ratings; more control over the conditions in which the test is administered). The credibility of this or other moderator analyses would be enhanced if the moderator operates consistently in both the Employment Inventory data and in the PCI conscientiousness data, given that both primarily tap conscientiousness. However, a moderator in the PCI conscientiousness data is unlikely to be found because the *Q* statistic associated with the variance homogeneity test indicates that all the variance in the PCI validity distribution is due to sampling error. This does not necessarily rule out finding a common moderator for both data sets but it is not a favorable situation. However, other perspectives and analyses are possible.

Research results consistent with publication bias raises concerns about the extent to which publication bias has distorted the science in I/O psychology and management. Studies addressing this potential bias appeared as early as 1993. Examining the validity data for the General Aptitude Test Battery (GATB), Vevea, Clements, and Hedges (1993) found no evidence of publication bias that would alter conclusions about the GATB tests' validity. However, Russell et. al (1994) reported that studies whose authors were in private industry yielded mean employment test validities of .32 but the mean validities were substantially lower (.24) for those studies whose authors were employed in academia. Also, studies conducted for organizational need, resulted in higher mean validities (.32) than those conducted due to researcher interest (.24). Thus, for example, studies conducted for EEO compliance reasons yielded higher mean validities (.33) then those studies conducted for the purposes of theory testing and development. Russell et al. (1994) offered several possible causes of the effects which could be interpreted as motivations for publication bias, although Russell and colleagues did not use the phrase publication bias. I argue that the results of the Russell et al. research are consistent with a financial motive for data suppression.

Currently, the extent to which publication bias distorts the science in our field is not known. This is primarily because few look for publication bias. I suggest that publication bias has the largest probability of occurring in three types of literature. The first type is the "we all know that it is true" literature. An example would be the employment interview literature where it has been concluded in multiple meta-analyses that structured interviews are more valid than unstructured interviews. The wide acceptance of this conclusion makes it unlikely that results contrary to the conclusion will be submitted for publication, or if submitted, will be published. The second type is the "It is money that I love" literatures. These are literatures where the results of studies influence the amount of money that a person or organization can make or lose. Test vendors, not unlike pharmaceutical companies, may benefit financially if a study shows favorable results for their product. Likewise, they may financially suffer if a study is disfavorable to their product. The third type is the "You can't talk about this" literature. Included in this literature is research on socially unpleasant topics such as race and sex differences.

The only sure way of evaluating the extent to which our science and practice has been damaged by publication bias is to look for it. It is recommended that publication bias analyses should be included in all meta-analyses. Likewise, test vendors should report publication bias analyses in their technical manuals and consumers of test vendor products should consider whether products they seek to purchase have validity results consistent with inferences of no publication bias. Thus, for example, one could rely on this paper to conclude that the Wonderlic PCI conscientious scale validities reported in the manual have not been distorted due to data suppression.

The many news reports concerning data suppression conducted (or allegedly conducted) by pharmaceutical companies has severely damaged the credibility of these companies. The bias-free research conducted or funded by pharmaceutical companies is now tainted and questioned based on concerns about publication bias in other areas of research. Many have commented on the gap between human resource practices and human resource research and the difficulties involved with having organizations use evidenced-based management (Highouse, 2008; Rynes, Brown, & Colbert, 2002; Rynes, Brown, & Colbert, 2002). It would be very unfortunate for our field if this situation was exacerbated by evidence reported widely in news outlets (e.g., "test vendors accused of biased reporting") suggested that our science is wrong. Thus, our field would benefit from increased attention to publication bias, and when it is found, for increased attention addressed to fixing our science.

References

Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. R. (2005). *Comprehensive meta-analysis. Version 2*. Englewood, NJ: Biostat.

Dickerson, K. (2005). Publication bias: Recognizing the problem, understandings its origins and scope, and preventing harm. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta analysis: Prevention, assessment, and adjustments* (pp. 11-34). Chichester, UK: Wiley.

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

Highhouse, S. (2008). Stubborn reliance on intuition and subjective judgment in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 333-342.

Lau, J., Schmid, C. H., & Chalmers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology, 48,* 45-57.

McDaniel, M.A., Rothstein, H.R. & Whetzel, D.L. (2006). Publication bias: A case study of four test vendors. *Personnel Psychology, 59,* 927-953.

Ones DL, Viswesvaran C, Schmidt FL. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78,* 679–703.

Paajanen, G. E., Hansen, T. L. & McLellan, R. A. (1993). *PDI Employment Inventory and PDI Customer Service Inventory manual*, Minneapolis: Personnel Decisions, Inc.

Pollack, J.M. & McDaniel, M.A. (2008, April). *An examination of the PreVisor™ Employment Inventory for publication bias.* Paper presented at the 23rd Annual Conference of the Society for Industrial and Organizational Psychology. San Francisco.

Russell, C.J., Settoon, R.P., McGrath, R.N., Blanton, A.E., Kidwell, R.E., Lohrke, F.T., Scifres, E.L., & Danforth, G.W. (1994). Investigator characteristics as moderators of personnel selection research: A meta-analysis. *Journal of Applied Psychology, 79*, 163-170.

Rynes, S.L., Brown, K.G., & Colbert, A.E. (2002). Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Executive, 16*, 92-103.

Rynes, S.L., Colbert, A.E., & Brown, K.G. (2002). Hr professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management, 41*, 149–174.

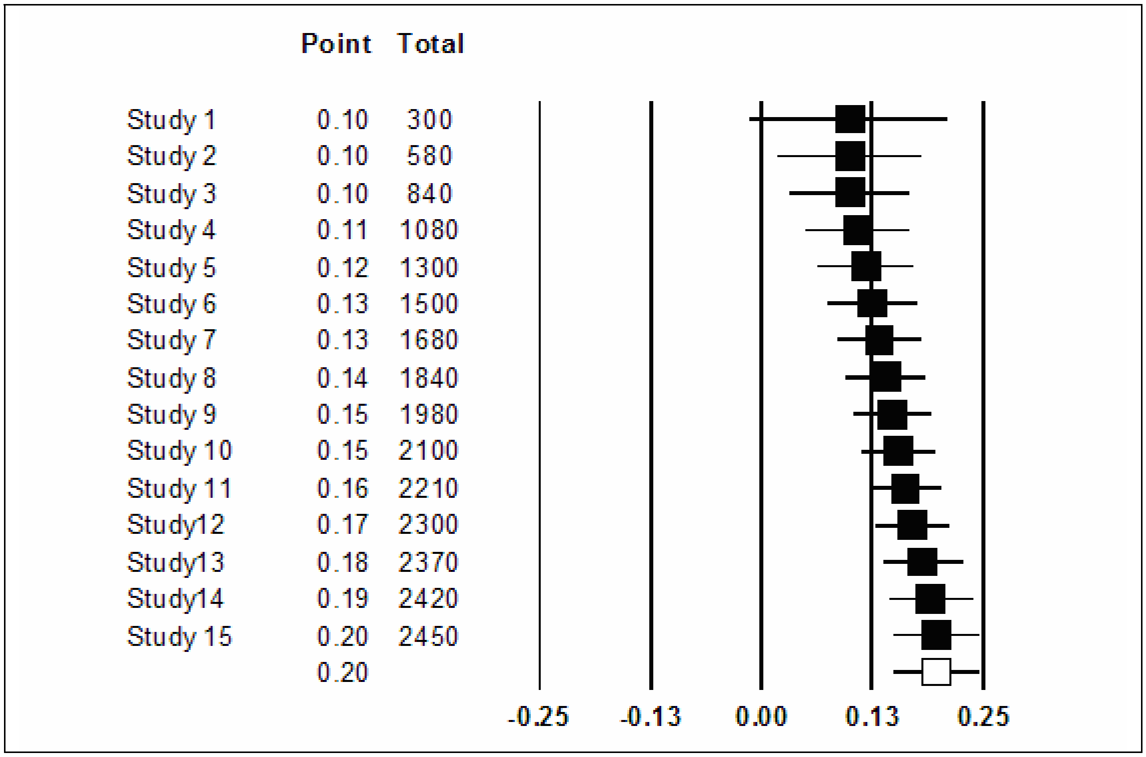
Saul, S. (2008). *Experts conclude Pfizer manipulated studies*. Article reported in New York Times October 8, 2008. Accessed on March 18, 2009 from http://www.nytimes.com/2008/10/08/health/research/08drug.html.

Tate, B.W. & McDaniel, M.A. (2008, August). *Race differences in personality: an evaluation of moderators and publication bias*. Paper presented at the Annual meeting of the Academy of Management, Anaheim CA.

Vevea J.L, Clements N.C, Hedges L.V. (1993). Assessing the effects of selection bias on validity data for the General Aptitude Test Battery. *Journal of Applied Psychology, 78*, 981–987.

Wonderlic, Inc. (2002). *Personal Characteristics Inventory*. Libertyville, IL: Author.

Figure 1. A cumulative means graph



	Point	Total
Study 1	0.10	300
Study 2	0.10	580
Study 3	0.10	840
Study 4	0.11	1080
Study 5	0.12	1300
Study 6	0.13	1500
Study 7	0.13	1680
Study 8	0.14	1840
Study 9	0.15	1980
Study 10	0.15	2100
Study 11	0.16	2210
Study12	0.17	2300
Study13	0.18	2370
Study14	0.19	2420
Study 15	0.20	2450
	0.20	

Figure 2. Observed and imputed validity distributions for PreVisor Employment Inventory. Adapted from Pollack & McDaniel (2008)

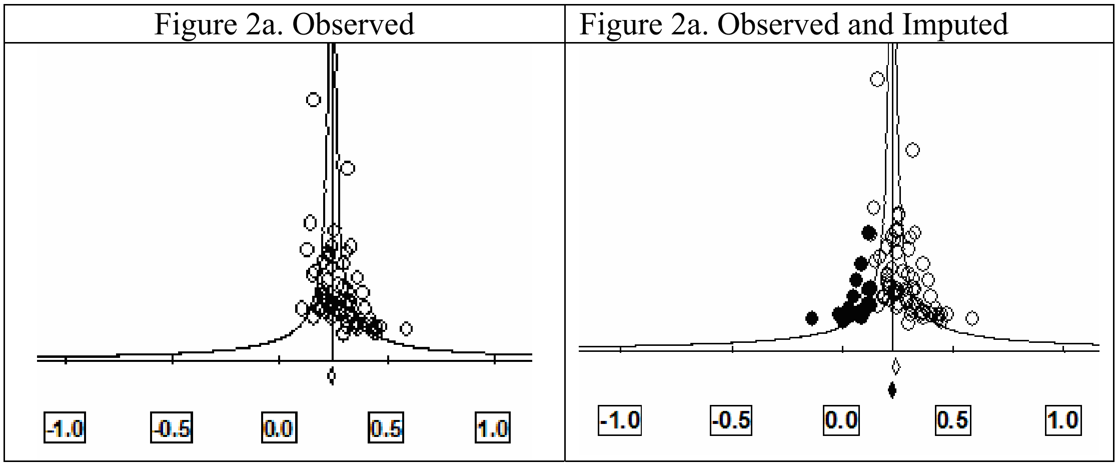
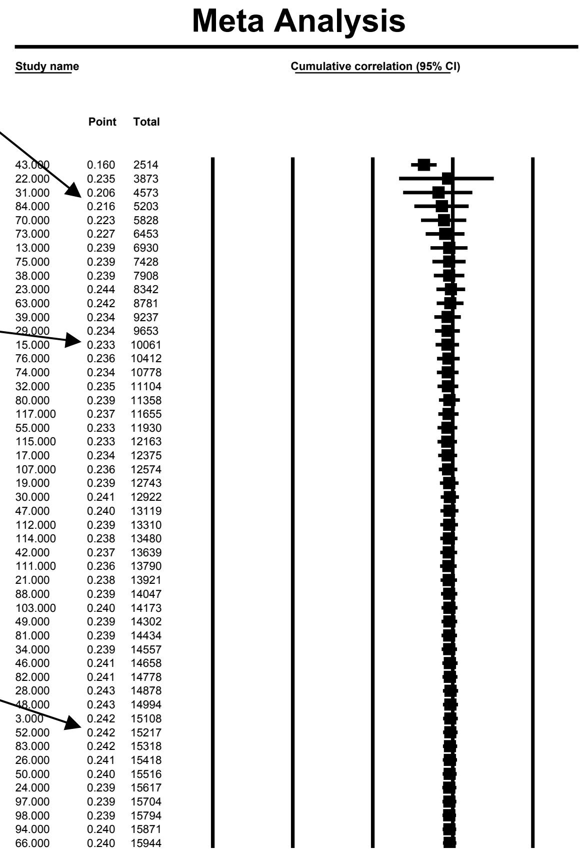


Figure 2a. Observed

Figure 2a. Observed and Imputed

Figure 3. Cumulative meta-analysis of the PreVisor data.



Study name	Point	Total
43.000	0.160	2514
22.000	0.235	3873
31.000	0.206	4573
84.000	0.216	5203
70.000	0.223	5828
73.000	0.227	6453
13.000	0.239	6930
75.000	0.239	7428
38.000	0.239	7908
23.000	0.244	8342
63.000	0.242	8781
39.000	0.234	9237
29.000	0.234	9653
15.000	0.233	10061
76.000	0.236	10412
74.000	0.234	10778
32.000	0.235	11104
80.000	0.239	11358
117.000	0.237	11655
55.000	0.233	11930
115.000	0.233	12163
17.000	0.234	12375
107.000	0.236	12574
19.000	0.239	12743
30.000	0.241	12922
47.000	0.240	13119
112.000	0.239	13310
114.000	0.238	13480
42.000	0.237	13639
111.000	0.236	13790
21.000	0.238	13921
88.000	0.239	14047
103.000	0.240	14173
49.000	0.239	14302
81.000	0.239	14434
34.000	0.239	14557
46.000	0.241	14658
82.000	0.241	14778
28.000	0.243	14878
48.000	0.243	14994
3.000	0.242	15108
52.000	0.242	15217
83.000	0.242	15318
26.000	0.241	15418
50.000	0.240	15516
24.000	0.239	15617
97.000	0.239	15704
98.000	0.239	15794
94.000	0.240	15871
66.000	0.240	15944

With 4 studies needed to bring the N to over 5,000, the mean validity is .216.

With 14 studies needed to bring the N to over 10,000, the mean validity is .233.

With 41 studies needed to bring the N to over 15,000, the mean validity is .242.

Not shown, but after adding all 70 studies (N = 16,941), the mean is .249.

Figure 4. Cumulative meta-analysis and trim and fill analysis of the Wonderlic PCI Conscientiousness data.

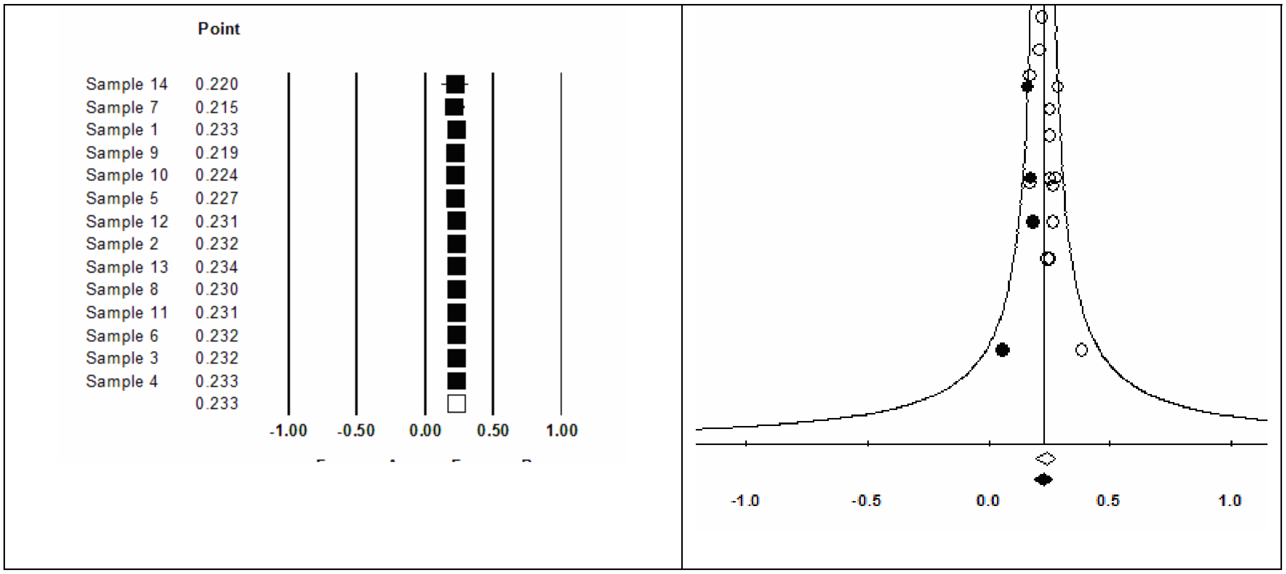


Figure 5. Hypothetical data illustrating how data suppression may lead to incorrect conclusions concerning Black-White mean differences in conscientiousness.

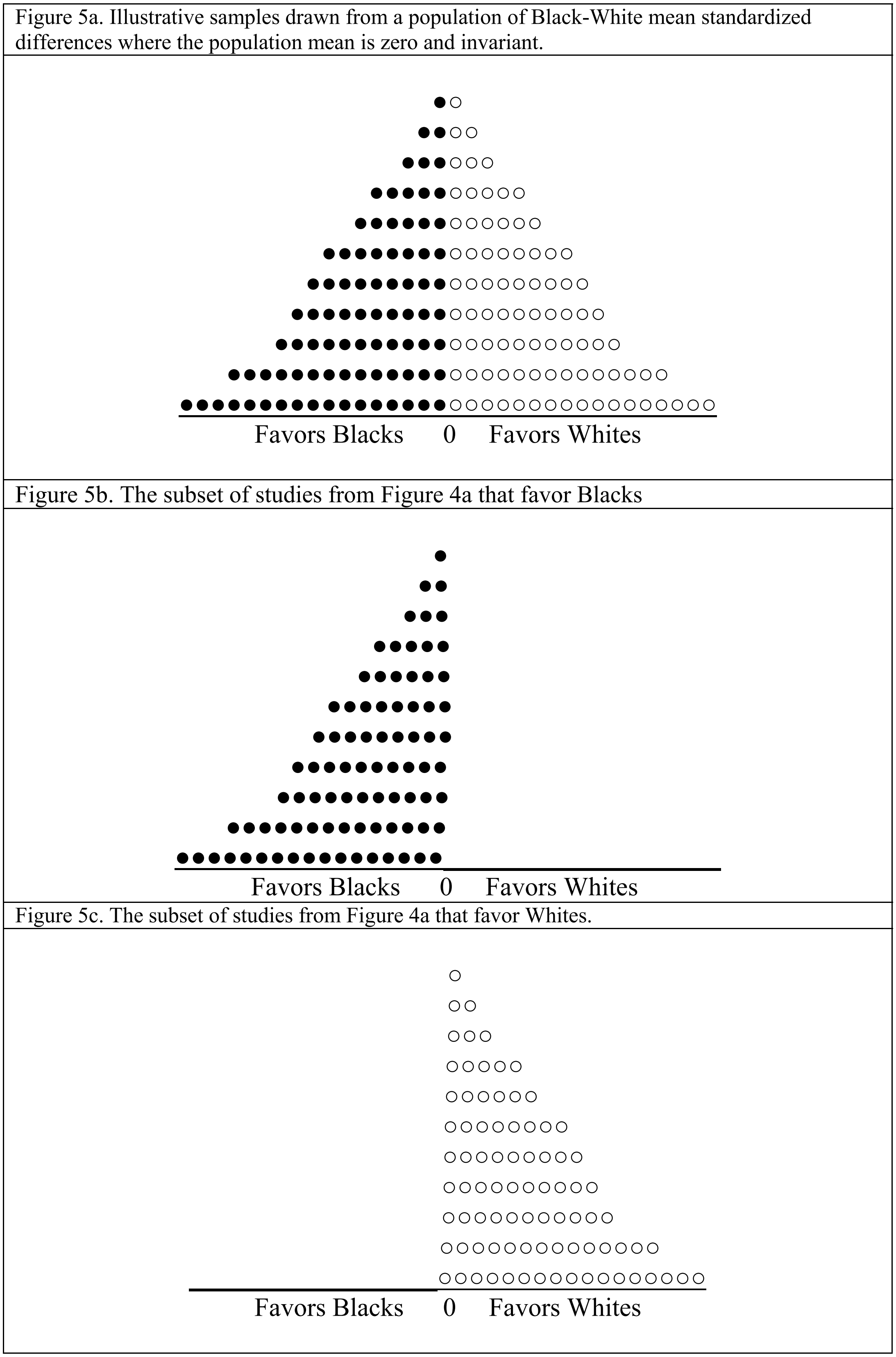


Figure 5a. Illustrative samples drawn from a population of Black-White mean standardized differences where the population mean is zero and invariant.

Figure 5b. The subset of studies from Figure 4a that favor Blacks

Figure 5c. The subset of studies from Figure 4a that favor Whites.

Figure 6. Cumulative meta-analysis and trim and fill publication bias graphics for conscientiousness effect sizes from journals. Adapted from Tate and McDaniel (2008).

Figure 6a. Cumulative meta analysis and trim and fill analysis for data from published studies where the mean racial differences were reported in the journal article.

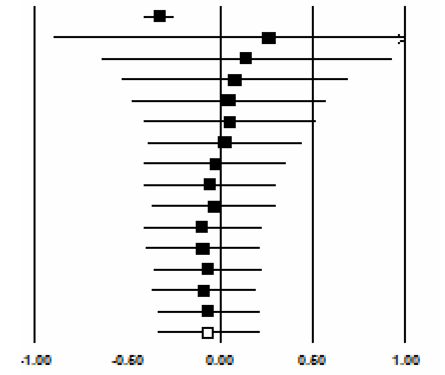
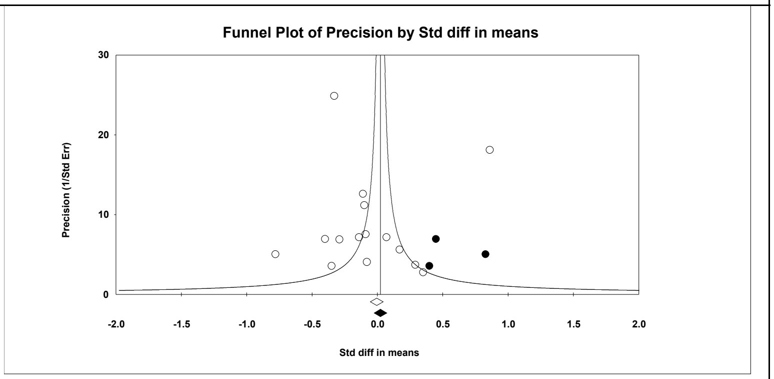
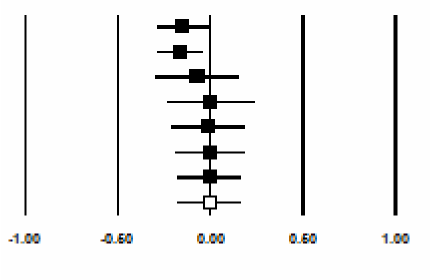
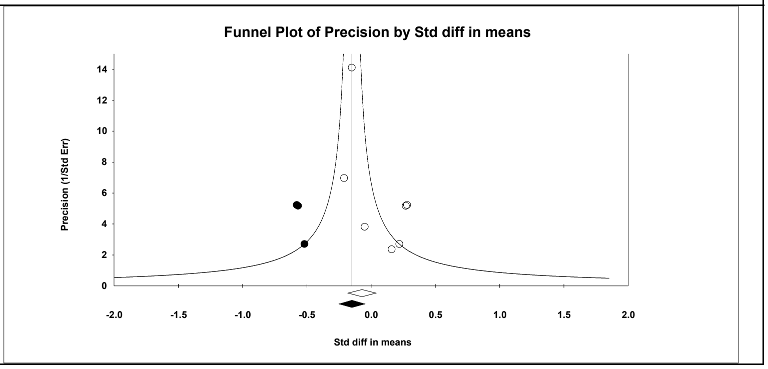
Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
	

Figure 6b. Cumulative meta analysis and trim and fill analysis for data from published studies where the mean racial differences were NOT reported in the journal article but the data were obtained from the author.

Cumulative Meta-Analysis. Studies sorted low to high by standard error.	Trim and Fill Plot
 -1.00 -0.50 0.00 0.50 1.00	 Funnel Plot of Precision by Std diff in means Precision (1/Std Err) 14 12 10 8 6 4 2 0 -2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 Std diff in means