ROLAND T. RUST, DONALD R. LEHMANN, and JOHN U. FARLEY*

A central assumption of meta-analysis is that the sample of studies fairly represents all work done in the field, published and unpublished. However, if studies with "poor" results are less likely to be published, a potential publication bias is present. The authors propose a maximum likelihood approach to estimating publication bias for the situation in which censorship based on effect size may occur. An explicit hypothesis test is provided for testing whether or not censorship is present. The method also simultaneously estimates the proportion of studies censored, the threshold past which censorship is avoided, and the probability of censorship if a potential observation is under the censorship threshold. Two published meta-analyses are examined and some publication bias is found in each, but no publication bias is detected in a meta-analysis of proprietary research data.

Estimating Publication Bias in Meta-Analysis

In recent years, meta-analysis has become popular as a method of generalizing the findings of a cross-section of marketing studies (Farley and Lehmann 1986). The quality of generalizations available from a meta-analysis depends on how representative the available studies are of both the present research base and a reasonable range of research environments.

As a field matures, "publication bias" may be an increasing problem—the tendency of journals to accept only strong effects or statistically significant findings may lead to an upward bias in magnitude of reported effects. As evidence accumulates, results that depart from those of past studies are looked at more suspiciously by reviewers, who are often authors of previous studies. Also, only discriminating studies may be published because of extensive replication. Authors conditioned by the refereeing process may suppress or at least cull out results that show relatively small or insignificant effects. Finally, "better" journals may impose what appear to be more rigorous standards, which can lead to further suppression of "weak" results. This practice has been referred to as the "file drawer problem" (Rosenthal 1979) and it has been demonstrated empirically by surveys (Chase and Chase 1976; Greenwald 1975). A good review by Begg and Berlin (1988) documents the seriousness of publication bias.

*Roland T. Rust is Professor and area head of Marketing, Owen Graduate School of Management, Vanderbilt University. Donald R. Lehmann is George E. Warren Professor of Business and John U. Farley is R. C. Kopf Professor of International Business, Columbia University.

The authors thank Robert A. Peterson for generously providing data used in the study.

Journal of Marketing Research
Vol. XXVII (May 1990), 220–226

Prior research on publication bias has followed two main approaches. First, in the "fail-safe sample size" approach (Cooper 1979), the number of unpublished studies needed to offset a published conclusion is computed. Fail-safe sample sizes for censorship bias based on significance (Rosenthal 1979) and effect size (Orwin 1983) have been explored.

Second, in the "maximum likelihood" approach (Hedges and Olkin 1985), a likelihood function describing the censorship process is specified and maximized. Hedges and Olkin assume that the observed test statistic is distributed according to a noncentral t distribution and that all nonsignificant results are censored out of the sample—an unrealistic assumption because nonsignificant results sometimes are published. Iyengar and Greenhouse (1988) relax the assumption of total censorship of nonsignificant results, but they also assume that all significant studies pass through uncensored, which again seems unrealistic.

We develop a maximum likelihood, weighting function approach to estimating publication bias based on effect size. Rather than simply assuming a censorship process, we provide an explicit statistical test for whether censorship is present. Further, we leave the censorship threshold as a parameter to be estimated, rather than assuming *a priori* that the censorship threshold is a particular critical value. We also provide the flexibility of assuming alternative parametric forms of the underlying density. We use two published and one unpublished meta-analysis datasets to test whether censorship is detected.

In the next section we describe the method for estimating publication bias based on effect size. We then apply the method to our three meta-analysis datasets. Finally, we provide discussion and conclusions.

A METHOD FOR ASSESSING PUBLICATION BIAS

Estimating publication bias involves estimating a complete distribution from a partially censored one. Mathematically, the problem is one of the TOBIT type, with the added complication that the small values are truncated rather than being replaced by zero values.

We assume that the variable of interest (advertising elasticities, coefficients from various classes of buyer behavior models, correlations, goodness of fit measures, etc.) is distributed according to a density, $f(x)$. We assume that the underlying distributional form that generates the data is known or at least can be approximated satisfactorily. Parameters of this distribution are unknown and must be estimated. Subsequently we provide statistical criteria for deciding whether one functional form is to be preferred over another. Testing of alternative distributional assumptions enables us to relax the assumption of a known distribution.

Censorship Based on Effect Size

We assume that publication bias involves a fixed censorship threshold C, beyond which no censorship occurs. Let X be the effect size observed in a study and let $f(x)$ be the underlying probability density of effect sizes (prior to censorship) in the population. The random variable is assumed to be defined on the positive real line or to be transformed appropriately so that it meets this condition.

If $x \leq C$, we assume there is a fixed positive probability ϕ that the value will be censored—that is, not published. The censorship density, $g(x)$, generates the published data. We assume that all values are correct, so $g(x)$ is the probability density of effect sizes *after* censorship. If there is no censorship, $g(x)$ is the same as $f(x)$. If q is the probability that an observation X will be uncensored and p is the probability that X exceeds C, $g(x)$ can be written as

$$(1) \qquad g(x) = \begin{cases} f(x)/q & x > C \\ (1-\phi)f(x)/q & x \leq C \end{cases}$$

where:

$$(2) \qquad q = p + (1-p)(1-\phi) = 1 - \phi(1-p).$$

Equation 1 diminishes the likelihood of observing values below the censorship threshold.

The likelihood function for the uncensored observation $X_1, \ldots, X_n$ is

$$(3) \qquad L = \prod_i g(x_i) = \prod_{x \geq C} \{f(x_i)/[1-\phi(1-p)]\} \cdot \prod_{x<C} \{1-\phi)f(x_i)/[1-\phi(1-p)]\} = [\prod_i f(x_i)] \cdot (1-\phi)^y/[1-\phi(1-p)]^N$$

where y is the number of published observations less than C and N is the total number of published observations. When $\phi = 0$ (no censorship), C is also zero and

$$L = \prod_i f(x_i).$$

Testing for the presence of censorship involves a constrained version of the general censorship model, with ϕ and C constrained to be zero, in a standard likelihood ratio test. Let $\hat{f}_0(x)$ be the estimated density without censorship and let $\hat{f}(x)$ be the estimated density from the general censorship model. The likelihood ratio statistic Λ then is given by

$$(4) \qquad \Lambda = \{\prod_i [\hat{f}_0(x_i)/\hat{f}(x_i)]\} \cdot [1-\phi(1-p)]^N/(1-\phi)^y.$$

The statistic $-2 \cdot \ln\Lambda$ will be distributed asymptotically chi square with two degrees of freedom corresponding to ϕ and C.

For example, if $f(x)$ is exponential ($f(x) = (1/\theta)\, e^{-x/\theta}$), the uncensored likelihood is

$$(5) \qquad L(\hat{\omega}) = \bar{x}^{-N}\exp(-N)$$

and the censored likelihood is

$$(6) \qquad L(\hat{\Omega}) = \hat{\theta}^{-N}[\exp(-N\bar{x}/\hat{\theta})](1-\hat{\phi})^{\hat{y}}/[1-\hat{\phi}(1-\hat{p})^N].$$

The likelihood ratio is

$$(7) \qquad \Lambda = \{\exp[N((\bar{x}/\hat{\theta})-1)]\}\{\hat{\theta}[1-\hat{\phi}(1-\hat{p})]/\bar{x}\}^N(1-\hat{\phi})^{-\hat{y}}$$

and the estimated bias is $\bar{x} - \hat{\theta}$ if the censored model is accepted.

Alternative Assumptions About Underlying Distributions

Our model can test different underlying assumptions about the parametric form of the underlying density. We examine two general shapes that might describe a meta-analysis and that we use subsequently: (1) $f(x)$ starting from its maximum at zero and declining at a decreasing rate, which might describe an experimental literature with many small effects and a few large ones and (2) $f(x)$ having a mode at some point to the right of the origin, which might describe the distribution of parameters that has a mode greater than zero. These two general shapes can be modeled by the exponential distribution and Erlang 2 distribution, respectively. We anticipate that these two distributions approximate the densities found most often in practical situations encountered by researchers. However, the method is fully general, and the testing of more complicated or differently shaped distributions can be undertaken at the researcher's discretion.

Alternative assumptions about the underlying distribution can be compared by using Akaike's criterion (Akaike 1974; Rust and Schmittlein 1985), which compares models according to the log of the maximum likelihood, with correction for the number of parameters to be estimated. For example, the uncensored exponential and the uncensored Erlang 2 distributions both have one parameter. The censored model adds two parameters, one for censorship threshold and another for probability of censorship. Suppose the exponential assumption results in censorship being identified and produces a maximum log-likelihood L_{EX}, whereas the Erlang 2 assumption leads to an inference of no censorship and a maximum log-likelihood L_{ER}. If A_{EX} and A_{ER} denote the respective Akaike criteria, the exponential model is chosen over the uncensored Erlang 2 model if

$$(8) \qquad (A_{EX} = L_{EX} - 3) > (A_{ER} = L_{ER} - 1).$$

Simulated Results

A simulation was used to investigate performance of the model for two distributions. Assuming an underlying standardized exponential distribution (unit mean and variance) or a standard Erlang 2 distribution (mean $\sqrt{2}$, variance 1), we varied the censorship thresholds, probabilities of censorship when under the threshold, and sample sizes. The design is shown in Table 1.

If the test for censorship was significant, all parameters of the censorship model were estimated. Otherwise, a null "uncensored" model was used to produce the parameter estimates. Table 2 shows the accuracy of the method in correctly identifying censorship, summarized for all experimental conditions. The method generally performs better in the absence of censorship than in its presence. When the actual distribution is uncensored exponential, our method generally finds no censorship. When the underlying distribution is censored exponential, our method identifies censorship in the majority of cases. When an Erlang 2 distribution was assumed, the model was correct 94% of the time when censorship was not present. However, the model detected censorship only 43% of the time when it was present.

Table 3 shows how the model performs on several criteria under various experimental conditions. "Bias in estimated mean" indicates the extent to which the model tends systematically to overestimate or underestimate the mean. "Proportion of cases in which censorship found" indicates how often censorship is discovered. The "mean of absolute error in estimated mean" is the average magnitude of the error in estimating the mean. Similarly, "mean of absolute error in estimated probability" is the average magnitude of the error in estimating the probability of censorship when a potential observation is below the censorship threshold and "mean of absolute error in estimated threshold" is the average magnitude of the error in estimating the censorship threshold. Part A of Table 3 shows that the accuracy of estimates deteriorates for threshold values of two or more standard deviations, but accuracy deteriorates even more for the null model. In general, high threshold values result in poorer accuracy on all error criteria. The censorship model outperforms the null model on all but two criteria, both under the Erlang assumption.

Part B of Table 3 shows that the censorship model estimates improve with sample size, though the improvement is not rapid. (Sample sizes of 100 to 300 studies

Table 1
DESIGN OF THE SIMULATION

Variable	*Levels*
Probability of censorship (6 levels)	0 (none), .2, .4, .6, .8, 1.0 (complete)
Censorship threshold (6 levels)	$.1\mu$, $.2\mu$, $.5\mu$, μ, $\mu + \sigma$, $\mu + 2\sigma$
Underlying distribution assumed (2 levels)	Standardized exponential (μ = 1, σ = 1), standardized Erlang 2 ($\mu\sqrt{2}$, σ = 1)
Sample size (3 levels)	50, 200, 500
Replication (10 per distribution)	

Table 2
ACCURACY IN IDENTIFYING CENSORSHIP IN SIMULATED DATA

		Identified as		
		Uncensored	*Censored*	*% correct*
Actual distribution = exponential				
Actual	Uncensored	152	28	84
	Censored	362	538	60
Actual distribution = Erlang 2				
Actual	Uncensored	170	10	94
	Censored	512	388	43

Table 3
PERFORMANCE OF MODEL UNDER CENSORSHIP

	Experimental factors									
	Bias in est. mean		*Proportion of cases in which censorship found*		*Mean (S.D.) of absolute error in est. mean*		*Mean (S.D.) of absolute error in est. probability*		*Mean (S.D.) of absolute error in est. threshold*	
	Exp.	*Erl.*	*Exp.*	*Erl.*	*Exp.*	*Erl.*	*Exp.*	*Erl.*	*Exp.*	*Erl.*
A. Censorship threshold										
$\mu/10$	.00	.02	.67	.07	.08 (.07)	.08[a] (.08)	.21 (.21)	.59 (.29)	.06 (.19)	.17 (.14)
$\mu/5$	−.01	−.02	.78	.25	.08 (.07)	.09[a] (.08)	.17 (.21)	.42 (.30)	.10 (.12)	.25 (.13)
$\mu/2$	.00	−.01	.87	.51	.08 (.08)	.10 (.10)	.14 (.19)	.21 (.24)	.13 (.18)	.40 (.26)
μ	.03	.00	.67	.59	.10 (.10)	.10 (.10)	.15 (.18)	.14 (.17)	.39 (.47)	.71 (.48)
$\mu + \sigma$	.19	.04	.33	.52	.22 (.20)	.14 (.15)	.35 (.26)	.16 (.19)	1.46 (.84)	1.38 (.82)
$\mu + 2\sigma$	.16	.18	.27	.28	.19 (.19)	.25 (.29)	.39 (.29)	.33 (.26)	2.39 (1.98)	2.67 (1.09)
B. Sample size										
50	.07	.04	.51	.24	.17 (.16)	.18 (.16)	.30 (.26)	.40 (.30)	.80 (1.07)	1.05 (1.09)
200	.06	.04	.60	.39	.11 (.12)	.11 (.18)	.23 (.24)	.29 (.29)	.75 (1.08)	.91 (1.07)
500	.06	.02	.68	.48	.10 (.13)	.08 (.13)	.18 (.23)	.23 (.26)	.71 (1.10)	.03 (1.03)
C. Probability of censorship										
.2	.03	.03	.28	.06	.09 (.08)	.09 (.08)	.20[a] (.19)	.21[a] (.05)	1.10 (1.06)	1.36 (1.19)
.4	.06	.05	.41	.28	.11 (.10)	.11 (.10)	.32 (.16)	.33 (.14)	1.01 (1.12)	1.18 (1.21)
.6	.08	.07	.59	.45	.15 (.13)	.14 (.14)	.32 (.25)	.31 (.16)	.91 (1.17)	.98 (1.19)
.8	.16	.04	.72	.66	.22 (.23)	.13 (.16)	.28 (.34)	.32 (.35)	.74 (1.16)	.62 (.76)
1.0	−.01	−.01	1.00	.71	.06 (.07)	.16 (.26)	.00 (.00)	.31 (.45)	.01 (.01)	.50 (.51)

[a]Null model performed better.

tend to characterize marketing meta-analyses.) In all cases, the censorship model outperforms the null model.

The probability of censorship also affects the accuracy of the estimates (Table 3, part C). Unsurprisingly, the higher the probability of censorship, the more likely our method is to detect censorship. As the estimate of the probability of censorship becomes more accurate, so does the estimate of where the threshold is. In only one situation does the null model outperform the censorship model—in this one case under both distributional assumptions. The Erlang model is slower to find censorship, probably because the left tail of the Erlang is small in relation to the exponential, and hence more evidence is needed to verify undersampling of that part of the distribution.

Though inappropriate application of the censorship model should be avoided, Table 4 shows that the quality of the estimates under the two models is comparable in the absence of censorship.

In general, the simulations indicate that the proposed method is conservative, tending to err more on the side of not detecting censorship. When censorship is identified, the incidence of "false positives" is low. Importantly, even when there is no censorship, use of the censorship model produces mean estimates that are almost as good as those obtained from the (correct) no-censorship model. The ability of the model to produce useful results is very poor only under exceptionally difficult circumstances (e.g., very large or very small censorship threshold, small sample size, or very low proportion of studies censored).

APPLICATIONS OF THE METHOD

We used the method to estimate publication bias in two published meta-analyses—a study of effect sizes in consumer behavior experiments (Peterson, Albaum, and Beltramini 1985) and a meta-analysis of parameters of econometric models of advertising (Assmus, Farley, and Lehmann 1984). For comparison, we also examined a proprietary set of econometric advertising models. As in

Table 4
PERFORMANCE OF ESTIMATION METHODS IN ABSENCE OF CENSORSHIP

	Exponential		*Erlang 2*	
	Null model	*Censorship model*	*Null model*	*Censorship model*
Bias in the estimated mean	.01	.01	.00	−.01
Proportion of cases in which censorship was found	NA	.16	NA	.06
Mean (S.D.) of absolute error in estimated mean	.07 (.07)	.08 (.08)	.07 (.07)	.08 (.08)

the preceding section, we compare results from assumptions of underlying exponential and Erlang distributions of parameters. We used two specially developed computer programs, BIASEX for the exponential assumption and BIASER for the Erlang 2 assumption, which are available from the authors. The use of these programs involves the following steps for implementation.

1. Input the effect sizes into a dataset, one row per study.
2. Run BIASEX and BIASER on the data. Outputs of each program include the log-likelihoods (censored model and uncensored model), parameter estimates, estimated censorship threshold, estimated probability of censorship, estimated mean, estimated bias, and the results of the likelihood ratio chi square test of whether or not censorship is present.
3. Compare the best model (censored or uncensored) from BIASEX with the best model from BIASER using Akaike's criterion (e.g., equation 8 shows the comparison of censorship exponential vs. uncensored Erlang 2).
4. Use the parameter estimates of the best model from step 3.

The empirical results based on the three meta-analyses studied are reported in Table 5.

Effect Sizes in Consumer Experiments

Peterson, Albaum, and Beltramini (1985) examined 311 articles from journals and proceedings published between 1970 and 1982 that reported the results of experimental manipulations. Their dependent variable was ω^2—the fraction of variability in the dependent variable attributable to a particular experimental effect. Aside from the inherent general potential for publication bias, only 115 of the articles reported actual values for ω^2 or provided the raw material needed to calculate it.

Applying our method to their data, we found censorship under the assumption of exponential distribution and none under the assumption of Erlang distribution of ω^2 (Table 5). The article indicates a preponderance of small effects, which might be more consistent with the exponential assumption. This is consistent with the Akaike model comparison, which prefers the censored exponential assumption to the uncensored Erlang. On the basis of the censored exponential distribution, the estimated publication bias is relatively small—about 5% of the mean effect.

Parameters of Econometric Advertising Models

Assmus, Farley, and Lehmann (1984) examined 127 econometric advertising models from 28 different studies. Our method detected no censorship in short-term elasticities of R^2 under either the exponential or Erlang distribution. Advertising carryover, however, appears to be censored under both exponential and Erlang assumptions. The Erlang assumption may be more reasonable

Table 5
ESTIMATED PUBLICATION BIASES IN THREE META-ANALYSES

	Published consumer models[a]	*Published advertising models*[b]			*Proprietary advertising models*[c]	
	Effect sizes in consumer experiments ω^2	*Short-term advertising elasticity*	*Carryover coefficient*	R^2	*Short-term advertising elasticity R*	
Underlying distribution is exponential						
Censorship found?	Yes	No	Yes	No	No	No
Likelihood ratio chi square (critical value is 5.99)	61.04	1.87	57.82	2.54	3.55	.29
Estimated mean under censorship model	.156	—	.34	—	—	—
Sample mean from studies	.163	.28	.45	.78	(proprietary)	(proprietary)
Estimated bias	.007	—	.11	—	—	—
Bias as % of estimated mean	4.5	—	30.0	—	—	—
Estimated % of studies censored	5.3	—	26.2	—	—	—
Censorship threshold	.010	—	.10	—	—	—
Probability of censorship if under threshold	1.00	—	.74	—	—	—
Underlying distribution is Erlang 2						
Censorship found?	No	No	Yes	No	No	No
Estimated mean	—	—	.41	—	—	—
Estimated bias	—	—	.04	—	—	—
Bias as % of estimated mean	—	—	9.3	—	—	—
Estimated % of studies censored	—	—	8.5	—	—	—
Censorship threshold	—	—	.12	—	—	—
Probability of censorship if under threshold	—	—	.89	—	—	—
Erlang 2 preferred in Akaike comparison to exponential?	No	No	Yes	No	No	Yes

[a]Peterson, Albaum, and Beltramini (1985).
[b]Assmus, Farley, and Lehmann (1984).
[c]Obtained from internal company studies, which were not subject to publication review.

in this case because of a preponderance of middle-sized values of carryover coefficients, a conclusion confirmed by the Akaike model comparison. Under the exponential assumptions, 74% of the observations below .104 are censored; under the Erlang assumptions, 89% of observations under .115 are censored. The mean estimated carryover is .45 for the uncensored model, .41 for the Erlang 2, and .34 for the exponential distribution, a reduction of about 10% and of some practical significance as well.

A set of similar proprietary models was made available by private correspondence after publication of the original meta-analysis, and these data prompted our interest in censoring problems in meta-analyses. A comparison is interesting in that, because these proprietary studies had not undergone the publication process, we did not anticipate publication bias. (Our correspondence indicated that the results represent a complete set and were not filtered before we received them.) Reassuringly, no publication bias was found in the proprietary data under either distributional assumption, either for R^2 or short-term elasticity. Unfortunately, carryover measures were not used in the proprietary models, so we could not test for censorship based on large coefficients.

DISCUSSION

We provide a quick test for identifying whether or not a study is seriously contaminated by publication bias. Bias may arise through self-censorship by the author to eliminate unimpressive findings or because editors and reviewers prefer articles with strong and "significant" findings.

Our method provides an explicit statistical test for the presence of censorship based on effect size. Alternative assumptions of the underlying distribution of effect sizes also can be tested.

In our method, a likelihood ratio test compares the censored model with the uncensored model under a particular functional form for generating the dependent variable of the meta-analysis. When censorship is detected, the method provides estimates of the extent of bias, the threshold below which censorship takes place, and the proportion of results below the threshold that are censored.

Simulations under an exponential distribution show that the censorship model usually finds censorship when it is present (given sufficiently high probability of censorship) and correctly does not find censorship when it is not present. Simulation results with an Erlang 2 distribution are similar, though there is less sensitivity in detecting censorship under this distributional assumption. If censorship is present, the censorship model does a better job of estimating the mean of the underlying distribution, particularly when there is a large sample size or a high probability of censorship.

Evidence of censorship is found in two published meta-analyses, whereas no censorship is found in a proprietary dataset corresponding to one of the meta-analyses. These findings are consistent with common sense and our general belief about the forces producing publication biases. In a study of experimental effects in published consumer research, the estimated publication bias in ω^2 was minimal. In a set of econometric advertising models, no publication bias was found in short-term advertising elasticities or in goodness-of-fit measures. Publication bias was found in advertising carryover, suggesting that a meta-analysis examining lagged advertising effects might tend to overstate carryover because of publication bias.

Limitations

Though our study extends the literature in publication bias in several ways, certain limitations should be noted. First, our model assumes censorship based on effect size instead of statistical significance. Possibly *large* effects also are censored, either because they imply obviousness or because they seem "too good to be true." Second, a meta-analysis may miss important results because the reporting was inadequate. For example, Peterson, Albaum, and Beltramini (1985) found that many studies simply reported "n.s." instead of a value when a result was not significant. Third, our method does not adjust for differences in sample size across the studies considered, though weighting on the basis of sample size is easy to incorporate into equation 3 of our model. Fourth, some important, but unknown, method variable may contaminate or otherwise alter the results. Finally, as in any maximum likelihood procedure, the consistency property does not guarantee good results unless a large sample is available. The simulation results we report give some idea of the model's performance for small samples under various circumstances.

Though its limitations should be carefully noted, the proposed method should prove useful to researchers who need a quick test of whether or not a meta-analysis is seriously contaminated by publication bias based on effect size.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transaction on Automatic Control,* 19, 716–23.

Assmus, Gert, John U. Farley, and Donald R. Lehmann (1984), "How Advertising Affects Sales: Meta-Analysis of Econometric Results," *Journal of Marketing Research,* 21 (February), 65–74.

Begg, Colin B. and Jess A. Berlin (1988), "Publication Bias: A Problem in Interpreting Medical Data," *Journal of the Royal Statistical Society,* A, 151 (Part 3), 419–63.

Chase, L. and R. Chase (1976), "Statistical Power Analysis of Applied Psychological Research," *Journal of Applied Psychology,* 61 (2), 234–7.

Cooper, H. (1979), "Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research," *Journal of Personality and Social Psychology,* 37 (1), 131–46.

Farley, John U. and Donald R. Lehmann (1986), *Meta-Analysis in Marketing: Generalization of Response Models.* Lexington, MA: Lexington Books.

Glass, Gene V., Barry McGraw, and Mary Lee Smith (1981), *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications, Inc.

Greenwald, A. (1975), "Consequences of Prejudice Against the Null Hypothesis," *Psychological Bulletin*, 82 (1), 1–20.

Hedges, Larry V. and Ingram Olkin (1985), *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press, Inc.

Hunter, John E., Frank L. Schmidt, and Gregg B. Jackson (1982), *Meta-Analysis: Cumulating Research Findings Across Studies*, Beverly Hills, CA: Sage Publications, Inc.

Iyengar, Satish and Joel B. Greenhouse (1988), "Selection Models and the File Drawer Problem, *Statistical Science*, 3 (1), 109–35.

Orwin, R. (1983), "A Fail-Safe N for Effect Sizes in Meta-Analysis," *Journal of Educational Statistics*, 8 (2), 157–9.

Peterson, Robert A., Gerald Albaum, and Richard P. Beltramini (1985), "A Meta-Analysis of Effect Size in Consumer Behavior Experiments," *Journal of Consumer Research*, 12 (June), 97–103.

Rosenthal, R. (1979), "The 'File Drawer Problem' and Tolerance for Null Results," *Psychological Bulletin*, 86 (3), 638–41.

Rust, Roland T. and David C. Schmittlein (1985), "A Bayesian Cross-Validated Likelihood Method for Comparing Alternative Specifications of Quantitative Models," *Marketing Science*, 4 (Winter), 20–40.

Reprint No. JMR272108