Psychological Methods
1996, Vol. 1, No. 1, 66–80

1082-989X/96/$3.00

A Procedure for Combining Sample Standardized Mean Differences and Vote Counts to Estimate the Population Standardized Mean Difference in Fixed Effects Models

Brad J. Bushman
Iowa State University

Morgan C. Wang
University of Central Florida

Missing effect-size estimates pose a difficult problem in meta-analysis. Conventional procedures for dealing with this problem include discarding studies with missing estimates and imputing single values for missing estimates (e.g., 0, mean). An alternative procedure, which combines effect-size estimates and vote counts, is proposed for handling missing estimates. The combined estimator has several desirable features: (a) It uses all the information available from studies in a research synthesis, (b) it is consistent, (c) it is more efficient than other estimators, (d) it has known variance, and (e) it gives weight to all studies proportional to the Fisher information they provide. The combined procedure is the method of choice in a research synthesis when some studies do not provide enough information to compute effect-size estimates but do provide information about the direction or statistical significance of results.

Missing data is perhaps the largest problem facing the practicing meta-analyst. Frequently, meta-analytic procedures cannot be applied because researchers fail to report relevant statistics or adequate descriptions of methods. Missing effect-size estimates pose a particularly difficult problem (Pigott, 1994). Without a statistical measure for the results of a study, meta-analytic methods cannot be used at all. Unfortunately, the proportion of studies with missing effect-size estimates in a research synthesis is often quite large. For example, we located 59 articles published in *Psychological Bulletin* between January 1990 and November 1995 that used meta-analytic procedures to combine effect-size estimates. The proportion of studies with missing estimates in these articles ranged from .03 to .86 (M = .23, SD = .19).

Brad J. Bushman, Department of Psychology, Iowa State University; Morgan C. Wang, Department of Statistics, University of Central Florida.

This research was supported in part by National Science Foundation Grant DMS-9305864.

Correspondence concerning this article should be addressed to Brad J. Bushman, Department of Psychology, W112 Lagomarcino Hall, Iowa State University, Ames, Iowa 50011-3180. Electronic mail may be sent via Internet to bushman@iastate.edu.

When a primary research report does not contain enough information to compute an effect-size estimate, it may still provide information about the magnitude of the treatment effect. Often this information is in the form of a report of the decision yielded by the significance test (e.g., a significant positive mean difference) or in the form of a direction of the effect without regard to its statistical significance (e.g., a positive mean difference).[1] Thus, the meta-analyst has access to at least one of four types of data from a primary research report: (a) information that can be used to compute an effect-size estimate, (b) information about whether the hypothesis test found a statistically significant treatment effect and the direction of the effect, (c) information about only the direction of the treatment effect, and (d) no information about the treatment effect. These data are rank ordered, from most to least, in terms of the amount of information they contain (Hedges, 1986). Effect-size estimates are considered missing if the data are of the second, third, or fourth type.

[1] We use the term *positive* to refer to results in the predicted direction.

Currently, the most common solutions to the problem of missing effect-size estimates are (a) to omit from the review those studies with missing effect-size estimates and analyze only complete cases, (b) to set the missing effect-size estimates equal to zero, (c) to set the missing effect-size estimates equal to the mean obtained from studies with effect-size estimates, and (d) to set studies equal to the conditional mean obtained from studies with effect-size estimates (i.e., Buck's [1960] method). None of these current procedures for handling missing effect-size estimates distinguishes among the different types of missing effect-size data, and all of these procedures have serious problems that limit their usefulness.

In this article, we propose an alternative procedure for handling missing effect-size estimates. Our procedure, called the *combined procedure*, combines effect-size estimates and vote counts to estimate the population effect size. The combined procedure distinguishes among the different types of missing data, but it does not use the fourth type of data (i.e., no information about the treatment effect).[2] The particular effect-size index we consider in this article is the standardized mean difference, although the combined procedure can be applied to other effect-size indexes (see Bushman & Wang, 1995, for a similar procedure for the correlation coefficient). First, we describe the problems associated with existing procedures for handling missing effect-size estimates. Second, we describe effect-size and vote-counting procedures for standardized mean differences. Although these procedures are described in detail elsewhere (e.g., Cooper & Hedges, 1994; Hedges & Olkin, 1985), a discussion of the combined procedure we are proposing requires a discussion of effect-size and vote-counting procedures. Third, we describe how to implement our combined procedure. Finally, we describe why our combined procedure is the method of choice for handling missing effect-size estimates.

Existing Methods for Handling Missing Effect-Size Estimates

Omit Studies With Missing Effect-Size Estimates (or Analyzing Only Complete Cases)

One common method for dealing with missing effect-size estimates is to omit studies from the meta-analysis that do not contain enough information to compute an effect-size estimate. In our sample of meta-analyses from *Psychological Bulletin*, studies with missing effect-size estimates were excluded in 26 of the 59 (44%) meta-analyses. This method assumes that the studies with effect-size estimates are a representative subset of the original sample of studies (Pigott, 1994).

There are two major problems with analyzing only complete cases. First, effect-size estimates are likely to be missing for reasons related to the data. Authors are more likely to report details of statistical analyses when the results are significant than when they are nonsignificant. This conditional reporting of statistical results is part of what has been labeled "prejudice against the null hypothesis" (Greenwald, 1975). Studies with effect-size estimates might represent a biased subset of the original sample of studies. Omitting studies with missing effect-size estimates therefore limits the generalizability of the results of a research synthesis. A second problem with this method is that the variance of the effect-size estimates will be unnecessarily large when studies with missing effect-size estimates are omitted. The variance of the effect-size estimates is used in several meta-analytic procedures (e.g., weighted mean estimates, homogeneity tests).

Set Missing Effect-Size Estimates Equal to Zero

A second method for handling missing effect-size estimates is to set the missing estimates equal to zero. In our sample of meta-analyses from *Psychological Bulletin*, missing effect-size estimates were set equal to zero in 9 of the 59 (15%) meta-analyses. This method assumes that all the missing effect-size estimates are zero or trivial in size.

There are two major problems with setting missing effect-size estimates equal to zero. First, it is unlikely that all the missing effect-size estimates are zero (or trivial in size). If the null hypothesis

[2] Little and Rubin (1987) use the method of maximum likelihood to impute missing values. Their procedure does not, however, use information about the direction and statistical significance of results. If no studies in the research synthesis provide information about the direction or statistical significance of results, then their procedure should be used.

in a research study is not rejected, two conclusions are possible: (a) the treatment effect is zero, or (b) the treatment effect is not zero but the power of the statistical test was too low to reject the false null hypothesis. In the social sciences, research designs typically have low statistical power due to small sample sizes (e.g., Cohen, 1988). Setting missing effect-size estimates equal to zero is therefore an overly conservative practice that underestimates the magnitude of the population effect size. A second problem with imputing the value zero for all missing effect-size estimates is that it artificially deflates the variance of the effect-size estimates (Little & Rubin, 1987; Pigott, 1994).

Set Missing Effect-Size Estimates to the Mean Obtained From the Studies With Effect-Size Estimates

Zero is not the only value that can be imputed for missing effect-size estimates. One can also impute the mean effect-size estimate obtained from the studies with effect-size estimates. In our sample of meta-analyses from *Psychological Bulletin*, missing effect size estimates were set equal to the mean effect-size estimate in only 1 of the 59 (2%) meta-analyses. This method assumes that the studies with missing effect-size estimates are missing completely at random.

There are two major problems with imputing the mean effect-size estimate for missing effect-size estimates. First, it is unlikely that the studies with missing effect-size estimates are missing completely at random. If treatment effects are larger in studies with effect-size estimates than in studies without effect-size estimates, then this method will overestimate the magnitude of the population effect size. Second, imputing the mean effect-size estimate for all missing estimates artificially deflates the variance of the effect-size estimates (Little & Rubin, 1987; Pigott, 1994).

Buck's Method

Buck (1960) proposed that regression techniques should be used to estimate missing values and that the missing observations should be replaced with the conditional mean. For example, effect size can be regressed on type of treatment or on degrees of freedom. In our sample of meta-analyses from *Psychological Bulletin*, none used Buck's method.[3] Buck's method assumes that the studies with missing effect-size estimates are missing completely at random and that effect-size estimates are linearly related to the other variables in the data set.

The major problem with Buck's (1960) method is that the studies with missing effect sizes are not likely to be missing completely at random. Another problem with this method is that it underestimates the variance of the effect-size estimates, but Little and Rubin (1987) offered an adjustment for this problem.

In summary, all of the current methods for dealing with missing effect-size estimates have serious problems that limit their usefulness. The combined procedure we are proposing overcomes these problems. Before describing the combined procedure, however, it is first necessary to describe effect-size and vote-counting procedures.

Effect-Size Procedures for Standardized Mean Differences

Model and Notation for Standardized Mean Differences

Meta-analysis is concerned with combining the statistical results from primary studies. When the primary studies in question compare two groups, either through treatment versus control comparisons or through orthogonal contrasts, the effect size is expressed as some form of standardized difference between the group means, often called the d index (Cohen, 1988). Suppose that the data arise from a series of k independent studies, each of which compares a treatment or experimental group (E) with a control group (C). Let Y_{ij}^E and Y_{ij}^C denote the respective jth observations in the experimental and control groups of the ith study, and let n_i^E and n_i^C denote the respective sample sizes for the experimental and control groups. Suppose also that the assumptions for the validity of the two-sample t test are met in each study. That is, the observations in the experimental and control groups of the ith study are independently normally distributed with means μ_i^E and μ_i^C, respectively, and common variance σ_i^2. Define the population effect size in the ith study as the standardized mean difference

[3] Procedures for handling missing estimates were not discussed in the remaining 23 (39%) meta-analyses.

$$\delta_i = \frac{\mu_i^E - \mu_i^C}{\sigma_i}, \tag{1}$$

and define the sample estimator of δ_i as

$$d_i = \frac{\overline{Y}_i^E - \overline{Y}_i^C}{S_i}, \tag{2}$$

where $\overline{Y}_i^E$ and $\overline{Y}_i^C$ are the respective sample means in the experimental and control groups and S_i is the pooled within-group sample standard deviation.

The sample standardized mean difference is a biased estimator of the population standardized mean difference. Hedges (1981) has provided the following small sample correction factor for this bias:

$$C(\nu_i) \approx 1 - \frac{3}{4\nu_i - 1}, \tag{3}$$

where $\nu_i = n_i^E + n_i^C - 2$. Thus, an unbiased estimator of δ_i is given by

$$d_i^U = C(\nu_i)d_i. \tag{4}$$

When $n_i^E = n_i^C$, d_i^U is not only an unbiased estimator of δ_i, it is also the unique minimum variance unbiased estimator of δ_i (see Hedges, 1981).

The two-sample t statistic for testing the significance of the mean difference in the ith study can be written as

$$t_i = \sqrt{\frac{n_i^E n_i^C}{n_i^E + n_i^C}}(d_i - \delta_i), \tag{5}$$

where t_i has a t distribution with degrees of freedom $\nu_i = n_i^E + n_i^C - 2$. The null hypothesis H_0: $\mu_i^E = \mu_i^C$ is rejected at significance level α if t_i is greater than t_{α,ν_i}, the one-sided critical value from the t distribution with degrees of freedom ν_i at significance level α.

Effect-Size Procedures for Estimating Standardized Mean Differences

The weighted average standardized mean difference (d_+) is calculated by averaging the individual standardized mean differences after each has been weighted by the inverse of its variance. That is,

$$d_+ = \frac{\sum_{i=1}^{k} w_i d_i}{\sum_{i=1}^{k} w_i}, \tag{6}$$

where the weight w_i, given by

$$w_i = \frac{2(n_i^E + n_i^C)n_i^E n_i^C}{2(n_i^E + n_i^C)^2 + n_i^E n_i^C d_i^2},$$

is the reciprocal of the estimated sampling variance of d_i (Cooper, 1989, p. 107). The variance of d_+ is given by

$$\mathrm{Var}(d_+) = 1 \Big/ \sum_{i=1}^{k} w_i. \tag{7}$$

The weighted average standardized mean difference (d_+) is a biased estimator of δ. An unbiased weighted average standardized mean difference is given by

$$d_+^U = \frac{\sum_{i=1}^{k} w_i^U d_i^U}{\sum_{i=1}^{k} w_i^U}, \tag{8}$$

where

$$w_i^U = \frac{2(n_i^E + n_i^C)n_i^E n_i^C}{2(n_i^E + n_i^C)^2 + n_i^E n_i^C (d_i^U)^2}.$$

Another positive feature of d_+^U is that it is a consistent estimator of δ. The variance of d_+^U is given by

$$\mathrm{Var}(d_+^U) = 1 \Big/ \sum_{i=1}^{k} w_i^U. \tag{9}$$

Under the fixed-effects model, the standardized mean differences are assumed to be homogeneous (i.e., $\delta_1 = \ldots = \delta_k = \delta$). Formal statistics for testing the homogeneity assumption have been described in detail elsewhere (e.g., Hedges & Olkin, 1985). If the homogeneity assumption is not met, the meta-analyst can test whether study characteristics moderate the effects of the treatment on the dependent variable. Meta-analytic procedures can then be applied to homogeneous subsets of effects.

The upper and lower bounds of a $100(1 - \alpha)\%$ confidence interval for the population standardized mean difference δ are given by

$$d_+ - z_{\alpha/2}\sqrt{\mathrm{Var}(d_+)} \leq \delta \leq d_+ + z_{\alpha/2}\sqrt{\mathrm{Var}(d_+)}, \tag{10}$$

where $z_{\alpha/2}$ is the two-sided critical value of the standard normal distribution at significance level α. The null hypothesis H_0: $\delta_1 = \ldots = \delta_k = \delta = 0$ is rejected at significance level α if the confidence interval in Equation 10 does not contain the value 0. If the correction factor (see Equation 3) was applied to the sample standardized mean differences, then the upper and lower bounds of a $100(1 - \alpha)\%$ confidence interval for δ are given by

$$d_+^U - z_{\alpha/2}\sqrt{\mathrm{Var}(d_+^U)} \leq \delta \leq d_+^U + z_{\alpha/2}\sqrt{\mathrm{Var}(d_+^U)}. \quad (11)$$

It is worth noting that the confidence interval for δ is shorter when it is based on d_+^U than when it is based on d_+, because w_i^U is greater than w_i.

Example 1: Effect-Size Procedure

Example 1 illustrates how to obtain an estimate and a 95% confidence interval for δ using the effect-size procedure. The data set was taken from previous meta-analytic reviews on intoxicated aggression by Bushman and his colleagues (Bushman, 1993, in press; Bushman & Cooper, 1990). One explanation of intoxicated aggression is that alcohol increases aggression because people expect it to. Those who behave aggressively while intoxicated can therefore blame the bottle for their actions. According to MacAndrew and Edgerton (1969), violent and other antisocial behaviors occur when alcohol is consumed because, in many societies, drinking occasions are culturally agreed-on time-out periods when people are not held accountable for their actions.

In laboratory experiments, this expectancy explanation of intoxicated aggression can be tested by comparing the level of aggression for participants who expect an alcoholic beverage but receive a nonalcoholic beverage (i.e., placebo) with the level of aggression for participants who expect and receive a nonalcoholic beverage (i.e., control). The placebo versus control comparison gives the pure effects of alcohol-related expectancies on aggression uncontaminated by the pharmacological effects of alcohol. To enhance the credibility of the placebo drink, researchers have poured the beverage from "legitimate" bottles and have placed a small amount of alcohol on the surface of the drink and on the rim of the glass. In the laboratory, aggression was measured by having participants give noxious physical (e.g., electric shocks, noise blasts) or verbal (e.g., negative written evaluations) stimuli to a confederate or by having participants take away positive stimuli from a confederate (e.g., money loss).

Table 1 lists the results from 13 laboratory experiments that included placebo and control groups; all participants were male social drinkers.[4] Standardized mean differences could be estimated for 9 of the 13 studies. Although the standardized mean differences in Table 1 are not homogeneous, they suffice for purposes of illustration. Suppose that one wants to obtain an estimate and a confidence interval for δ using the 9 sample standardized mean differences in Table 1. Because the sample sizes are small, Hedges's (1981) correction factor will be used. Using the sample sizes, sample standardized mean differences, weights, and correction factors in Table 2, one obtains the weighted average (see Equation 8)

$$d_+^U = \sum_{i=1}^{9} w_i^U d_i^U \Big/ \sum_{i=1}^{9} w_i^U = 17.675/85.549 = 0.207,$$

and the variance (see Equation 9)

$$\mathrm{Var}(d_+^U) = 1 \Big/ \sum_{i=1}^{9} w_i^U = 1/85.549 = 0.012.$$

The 95% confidence interval for δ (see Equation 11) is given by

$$d_+^U - z_{.05/2}\sqrt{\mathrm{Var}(d_+^U)} \leq \delta \leq d_+^U + z_{.05/2}\sqrt{\mathrm{Var}(d_+^U)}$$

$$= 0.207 - 1.96\sqrt{0.012} \leq \delta \leq 0.207 + 1.96\sqrt{0.012}.$$

Simplifying, one has $[-0.005, 0.419]$. Because the confidence interval includes the value zero, the null hypothesis that $\delta_1 = \ldots = \delta_9 = \delta = 0$ is not rejected.

Vote-Counting Procedures for Standardized Mean Differences

The Conventional Vote-Counting Procedure

Light and Smith (1971) were the first to propose a formal procedure for "taking a vote" of study results.

> All studies which have data on a dependent variable and a specific independent variable of interest are examined. Three possible outcomes are defined. The relationship between the independent variable and the dependent variable is either significantly positive, significantly negative, or there is no specific relationship in either direction. The number

---

[4] A list of the studies given in Table 1 can be obtained from Brad J. Bushman.

Table 1
Studies Testing Effects of Alcohol-Related Expectancies on Aggression

Study	Dependent measure	n^E	n^C	d	p	Direction
1	Electric shocks	19	19	1.028	<.05	+
2	Electric shocks	24	24	0.903	<.05	+
3	Electric shocks	24	24	0.552	<.05	+
4	Electric shocks	24	24	0.328	*ns*	+
5	Electric shocks	15	15	−0.540	*ns*	−
6	Electric shocks	16	16		*ns*	+
7	Electric shocks	12	12		*ns*	+
8	Electric shocks	10	10		*ns*	+
9	Electric shocks	10	10		*ns*	−
10	Noise blasts	12	12	0.193	*ns*	+
11	Negative evaluation	12	12	−0.660	*ns*	−
12	Money loss	24	24	−0.332	*ns*	−
13	Money loss	24	24	−0.036	*ns*	−

Note. Nonsignificant results were assumed for studies that did not report p values. n^E = placebo group sample size; n^C = control group sample size; d = standardized mean differences; *ns* = statistically nonsignificant at the .05 level; plus sign = positive result (i.e., in predicted direction); minus sign = negative result (i.e., in opposite direction).

of studies falling into each of these three categories is then simply tallied. If a plurality of studies falls into any one of these three categories, with fewer falling into the other two, the modal category is declared the winner. This modal category is then assumed to give the best estimate of the direction of the true relationship between the independent and dependent variable. (p. 433)

The conventional vote-counting procedure proposed by Light and Smith (1971) has been criticized on several grounds. First, as Light and Smith noted, it does not incorporate sample size into the vote. It is well known that as sample size increases, the probability of finding a statistically significant relation between the independent and dependent variables also increases. Second, although it allows one to determine which modal category is the "winner," it does not allow one to determine what the margin of victory is. In other words, the conventional vote-counting procedure does not provide an effect-size estimate. Third, Hedges and Olkin (1980) have shown that it has very low power for the range of sample sizes and effect sizes most common in the social sciences. Hedges and Olkin proposed vote-counting procedures that overcome the problems associated with the conventional vote-counting procedure. In this article, we only describe vote-counting procedures based on unequal sample sizes because our combined procedure does not use vote-counting procedures based on equal sample sizes (see Bushman, 1994, for a description of vote-counting procedures based on equal sample sizes).

Unequal Sample Size Vote-Counting Procedures for Estimating Standardized Mean Differences

If one makes the underlying assumption that the standardized mean differences are positive and homogeneous, the null and alternative hypotheses for the ith study are

$$H_0\colon \delta_i = \delta = 0$$
$$H_A\colon \delta_i > \delta = 0. \tag{12}$$

The test statistic for Equation 12 is

$$t_i = \sqrt{\frac{n_i^E n_i^C}{n_i^E + n_i^C}}(d_i - \delta_i), \tag{13}$$

which has a t distribution with degrees of freedom $\nu_i = n_i^E + n_i^C - 2$. If the correction factor (Equation 3) was used, then replace d_i with d_i^U in Equation 13.

The essential feature of vote-counting procedures is that some d_is are not observed. Even if one does not observe d_i, one can still estimate δ_i if the research report contains information about the direction or statistical significance of results. For each study, there are two possible outcomes: a "success" if $t_i > t_{\alpha,\nu_i}$ or a "failure" if $t_i \leq t_{\alpha,\nu_i}$. If α = .5, a success is defined as a positive mean difference and a failure is defined as a negative or null mean

Table 2
Computations for Effect-Size Analysis When Studies With Missing Estimates Are Omitted

Study	n^E	n^C	d	$C(\nu)$	d^U	w^U	$d^U w^U$
1	19	19	1.028	0.979	1.006	8.432	8.483
2	24	24	0.903	0.984	0.889	10.923	9.711
3	24	24	0.552	0.984	0.543	11.574	6.285
4	24	24	0.328	0.984	0.323	11.846	3.826
5	15	15	−0.540	0.973	−0.525	7.250	−3.806
10	12	12	0.193	0.966	0.186	5.974	1.111
11	12	12	−0.660	0.966	−0.638	5.710	−3.643
12	24	24	−0.332	0.984	−0.327	11.842	−3.872
13	24	24	−0.036	0.984	−0.035	11.998	−0.420
Total						85.549	17.675

Note. n^E = placebo group sample size; n^C = control group sample size; d = standardized mean differences. See Equations 3, 4, and 8 for definitions of $C(\nu)$, d^U, and w^U, respectively.

difference. If $\alpha = .05$, a success is defined as a significant positive mean difference and a failure is defined as a negative, null, or nonsignificant positive mean difference. Of course, other values of α can be used to define a significant positive mean difference (e.g., $\alpha = .01$).

Define V_i as an indicator variable that takes on the value 1 if the outcome is a success or the value 0 if the outcome is a failure. That is,

$$V_i = \begin{cases} 1 & \text{if } t_i > t_{\alpha,\nu_i} \\ 0 & \text{if } t_i \leq t_{\alpha,\nu_i} \end{cases}. \tag{14}$$

The probability of a success is given by

$$\Pr(V_i = 1) = \Pr(t_i > t_{\alpha,\nu_i}) = p_i, \tag{15}$$

and the probability of a failure is given by

$$\Pr(V_i = 0) = \Pr(t_i \leq t_{\alpha,\nu_i}) = 1 - p_i. \tag{16}$$

If we assume that the sample sizes are equal across the k studies, then $p_1 = \ldots = p_k = p$, and each V_i has a Bernoulli distribution with parameter p. The maximum likelihood estimator of p is

$$\hat{p} = \textstyle\sum_{i=1}^{k} v_i/k, \tag{17}$$

where v_i is the observed value of V_i (i.e., 1 or 0).

Because $E(\hat{p}) = p$ and $\text{Var}(\hat{p}) = p(1 - p)/k$, $\hat{p}$ is the proportion of positive or significant positive mean differences obtained from the k studies. The maximum likelihood estimator for the population standardized mean difference δ can be obtained by solving

$$\hat{p} = \Pr(t > t_{\alpha,\nu}),$$

where t has a noncentral t distribution with noncentrality parameter $\delta' = \sqrt{n^E n^C / n^E + n^C}\ \delta$ and degrees of freedom ν, where $\nu = \nu_1 = \ldots = \nu_k$.

The equal sample size assumption underlying vote-counting procedures, however, is extremely restrictive because studies frequently have different sample sizes. If the equal sample size assumption is not met, each V_i has a Bernoulli distribution with parameter p_i. Because the studies under review are assumed to be independent, the log-likelihood function is given by

$$L(\delta) = \sum_{i=1}^{k} [v_i \ln(p_i) + (1 - v_i)\ln(1 - p_i)], \tag{18}$$

where $v_1 = \ldots = v_k$ are the respective observed values of $V_1 = \ldots = V_k$ and $p_i = \Pr(t_i > t_{\alpha,\nu_i})$.[5]

The probability of a positive result is

[5] The vote-counting procedure presented can be used even if the homogeneity assumption is violated, but the log-likelihood function is

$$L(\delta_1, \ldots, \delta_k) = \sum_{i=1}^{k} [v_i \ln(p_i) + (1 - v_i)\ln(1 - p_i)],$$

where $p_i = \Pr_{\delta_i',\nu_i}(t_i > t_{\alpha,\nu_i})$ and t_i is a noncentral t variate with degrees of freedom $\nu_i = n_i^E + n_i^C - 2$ and noncentrality parameter $\delta_i' = \sqrt{n_i^E n_i^C/(n_i^E + n_i^C)}\,\delta_i$.

$$p_i = \Phi(\delta_i'), \tag{19}$$

where $\delta_i' = \sqrt{n_i^E n_i^C / n_i^E + n_i^C}\, \delta_i$ and $\Phi(.)$ is the standard normal cumulative distribution function. The probability of a significant positive result is

$$p_i = \Phi(\delta_i') - \sum_{j=0}^{\infty} \left[a_j I\left(j + \frac{1}{2}, \frac{\nu_i}{2}, x_i\right) + b_j I\left(j + 1, \frac{\nu_i}{2}, x_i\right)\right], \tag{20}$$

where

$$a_j = \frac{1}{2} e^{-(\delta_i'^2/2)} \frac{(\delta_i'^2/2)^j}{j!},$$

$$b_j = \frac{1}{2} \delta_i' e^{-(\delta_i'^2/2)} \frac{(\delta_i'^2/2)^j}{\sqrt{2}\,\Gamma(j + 3/2)},$$

$\Gamma(.)$ is the complete gamma function, $I(., ., .)$ is the incomplete beta function, and

$$x_i = \frac{t_{\alpha,\nu_i}^2}{t_{\alpha,\nu_i}^2 + \nu_i}$$

(see Appendix A).

Because $\nu_1, \ldots, \nu_k$ are known and the data $v_1, \ldots, v_k$ are observed, the log-likelihood function (Equation 18) is a function of δ alone. Thus, $L(\delta)$ can be maximized over δ to obtain the maximum likelihood estimator $\hat{\delta}$. Although there is no closed form solution for the maximum likelihood estimator $\hat{\delta}$, it can be obtained numerically using a computer subroutine.[6] When the number k of studies is large, $\hat{\delta}$ will be approximately normally distributed.

The theory of maximum likelihood also can be used to obtain an expression for the large sample variance of $\hat{\delta}$:

$$\mathrm{Var}(\hat{\delta}) = \left[\sum_{i=1}^{k} \frac{[D_i^{(1)}]^2}{p_i(1 - p_i)}\right]^{-1}, \tag{21}$$

where $D_i^{(1)}$ is the first-order derivative of p_i evaluated at $\delta = \hat{\delta}$ (see Hedges & Olkin, 1985, p. 70). The upper and lower bounds of a $100(1 - \alpha)\%$ confidence interval for the population standardized mean difference δ are given by

$$\hat{\delta} - z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\delta})} \le \delta \le \hat{\delta} + z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\delta})}. \tag{22}$$

One limitation of vote-counting procedures is that they cannot be used if all the results are in the same direction. The method of maximum likelihood cannot be used if $\hat{p}$ is unity or zero because there is not a unique corresponding value of $\hat{\delta}$. The estimator $\hat{p}$ could be unity when sample sizes and standardized mean differences are large. If all of the results are in the same direction, one can use a Bayes estimate of p (see Hedges & Olkin, 1985).

Example 2: Vote-Counting Procedure

Example 2 illustrates how to obtain an estimate and a confidence interval for δ using the unequal sample size vote-counting procedure. Using the data in Table 3 and a computer subroutine (see Footnote 6), one finds that the maximum value of $L(\delta)$ occurs at -14.692, which corresponds with the maximum likelihood estimator $\hat{\delta} = 0.298$. The variance of $\hat{\delta}$ (Equation 21) is 0.022, and the 95% confidence interval is [0.007, 0.589] (i.e., Equation 22: $0.298 - 1.96\sqrt{0.022} \le \delta \le 0.298 + 1.96\sqrt{0.022}$). Because the confidence interval does not include the value zero, the null hypothesis that $\delta_1 = \ldots = \delta_{13} = \delta = 0$ is rejected.

A Procedure for Combining Effect-Size Estimates and Vote Counts to Obtain an Estimate and a Confidence Interval for the Population Standardized Mean Difference

The combined procedure uses the effect-size and vote-counting procedures described previously. Out of k independent studies, suppose that the first m studies ($m < k$) report enough information to compute effect-size estimates and that the remaining $k - m$ studies only report the direction or statistical significance of results. One can use the first m studies to obtain an unbiased estimator d_+^U of the population standardized mean difference δ, and one can use the remaining $k - m$ studies to obtain the maximum likelihood estimator $\hat{\delta}$ of δ. Let $\mathrm{Var}(d_+^U)$ and $\mathrm{Var}(\hat{\delta})$ be the respective variances for d_+^U and $\hat{\delta}$. Because d_+^U and

[6] A FORTRAN computer subroutine for unequal sample size vote-counting procedures can be obtained by writing to Morgan C. Wang, Department of Statistics, University of Central Florida, Orlando, Florida 32828, or by electronic mail via Internet at cwang@pegasus.cc.ucf.edu.

Table 3
Computations for Vote-Counting Analysis

Study	n^E	n^C	p	Direction	v	$t_{\alpha,\nu}$
1	19	19	<.05	+	1	1.688
2	24	24	<.05	+	1	1.679
3	24	24	<.05	+	1	1.679
4	24	24	*ns*	+	1	1.679
5	15	15	*ns*	−	0	0
6	16	16	*ns*	+	1	0
7	12	12	*ns*	+	1	0
8	10	10	*ns*	+	1	0
9	10	10	*ns*	−	0	0
10	12	12	*ns*	+	1	0
11	12	12	*ns*	−	0	0
12	24	24	*ns*	−	0	0
13	24	24	*ns*	−	0	0

Note. n^E = placebo group sample size; n^C = control group sample size; *ns* = statistically nonsignificant at the .05 level; $t_{\alpha,\nu}$ = the one-side critical value from the t distribution with degrees of $\nu_i = n_i^E + n_i^C - 2$. See Equation 14 for explanation of v.

$\hat{\delta}$ are both consistent estimators of δ, the combined estimator of δ,

$$\hat{\delta}_C = \frac{d_+^U/\mathrm{Var}(d_+^U) + \hat{\delta}/\mathrm{Var}(\hat{\delta})}{1/\mathrm{Var}(d_+^U) + 1/\mathrm{Var}(\hat{\delta})}, \tag{23}$$

is consistent as well. The variance of the combined estimator is given by

$$\mathrm{Var}(\hat{\delta}_C) = \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})}. \tag{24}$$

Because $\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(d_+^U)$ and $\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(\hat{\delta})$, the combined estimator is more efficient than either d_+^U or $\hat{\delta}$ (see Appendix B for a proof).

The upper and lower bounds of a $100(1 - \alpha)\%$ confidence interval for the population standardized mean difference δ are given by

$$\hat{\delta}_C - z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\delta}_C)} \le \delta \le \hat{\delta}_C + z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{\delta}_C)}. \tag{25}$$

Because $\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(d_+^U)$ and $\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(\hat{\delta})$, the confidence interval based on the combined procedure is shorter than the confidence intervals based on the effect-size and vote-counting procedures.

Example 3: Combined Procedure

Example 3 illustrates how to obtain an estimate and a 95% confidence interval for δ using the combined procedure. Recall from Example 1 that the unbiased estimator of δ based on effect-size procedures was $d_+^U = 0.207$ with variance $\mathrm{Var}(d_+^U) = 0.012$. Four of the studies in Table 1 do not contain enough information to compute standardized mean differences, but these four studies do report the direction and statistical significance of results. Using the data in Table 4 and a computer subroutine (see Footnote 6), one finds that the maximum value of $L(\delta)$ occurs at -0.215, which corresponds to the maximum likelihood estimator $\hat{\delta} = 0.306$. The variance of $\hat{\delta}$ is $\mathrm{Var}(\hat{\delta}) = 0.112$. Thus, the combined estimator of δ (Equation 23) is

$$\hat{\delta}_C = \frac{d_+^U/\mathrm{Var}(d_+^U) + \hat{\delta}/\mathrm{Var}(\hat{\delta})}{1/\mathrm{Var}(d_+^U) + 1/\mathrm{Var}(\hat{\delta})} = \frac{0.207/0.012 + 0.306/0.112}{1/0.012 + 1/0.112} = 0.217,$$

and the variance of the combined estimator (Equation 24) is

$$\mathrm{Var}(\hat{\delta}_C) = \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})} = \frac{(0.012)(0.112)}{0.012 + 0.112} = 0.011.$$

A 95% confidence interval for δ (Equation 25) is given by

$$\hat{\delta}_C - z_{.05/2}\sqrt{\mathrm{Var}(\hat{\delta}_C)} \le \delta \le \hat{\delta}_C + z_{.05/2}\sqrt{\mathrm{Var}(\hat{\delta}_C)}$$
$$= 0.217 - 1.96\sqrt{0.011} \le \delta \le 0.217 + 1.96\sqrt{0.011}$$

Simplifying, one has [0.011, 0.423]. Because the

Table 4
Computations for Vote-Counting Analysis Used in Combined Procedure

Study	n^E	n^C	p	Direction	v	$t_{\alpha,\nu}$
6	16	16	*ns*	+	1	0
7	12	12	*ns*	+	1	0
8	10	10	*ns*	+	1	0
9	10	10	*ns*	−	0	0

Note. n^E = placebo group sample size; n^C = control group sample size; *ns* = statistically nonsignificant at the .05 level; $t_{\alpha,\nu}$ = the one-sided critical value from the t distribution with degrees of freedom $\nu_i = n_i^E + n_i^C - 2$. See Equation 14 for explanation of v.

confidence interval does not include the value zero, the null hypothesis that $\delta_1 = \ldots = \delta_k = \delta = 0$ is rejected.

Example 4: Other Methods for Handling Missing Effect-Size Estimates

Example 4 illustrates how to obtain an estimate and a 95% confidence interval for δ using the other methods for handling missing effect-size estimates.

Omit studies with missing effect-size estimates (or analyzing only complete cases). This method was illustrated in Example 1.

Set missing effect-size estimates equal to zero. In Table 5, the value zero was imputed for studies with missing estimates. Using the data in Table 5, one obtains the weighted average 0.161 (i.e., Equation 8: 17.675/109.549), the variance 0.009 (i.e., Equation 9: 1/109.549), and the 95% confidence interval [−0.026, 0.348] (i.e., Equation 11: $0.161 - 1.96\sqrt{0.009} \le \delta \le 0.161 + 1.96\sqrt{0.009}$). Because the confidence interval includes the value zero, the null hypothesis that $\delta_1 = \ldots = \delta_{13} = \delta = 0$ is not rejected.

Set missing effect-size estimates to the mean obtained from the studies with effect-size estimates. In Table 6, the mean estimate (obtained in Example 1) was imputed for studies with missing estimates. Using the data in Table 6, one obtains the weighted average 0.205 (i.e., Equation 8: 22.447/109.431), the variance 0.009 (i.e., Equation 9: 1/109.431), and the 95% confidence interval [0.018, 0.392] (i.e., Equation 11: 0.205 −

Table 5
Computations for Effect-Size Analysis When the Value Zero Is Imputed for Missing Effect-Size Estimates

Study	n^E	n^C	d	$C(\nu)$	d^U	w^U	$d^U w^U$
1	19	19	1.028	0.979	1.006	8.432	8.483
2	24	24	0.903	0.984	0.889	10.923	9.711
3	24	24	0.552	0.984	0.543	11.574	6.285
4	24	24	0.328	0.984	0.323	11.846	3.826
5	15	15	−0.540	0.973	−0.525	7.250	−3.806
6	16	16	0	0.975	0	8.000	0
7	12	12	0	0.966	0	6.000	0
8	10	10	0	0.958	0	5.000	0
9	10	10	0	0.958	0	5.000	0
10	12	12	0.193	0.966	0.186	5.974	1.111
11	12	12	−0.660	0.966	−0.638	5.710	−3.643
12	24	24	−0.332	0.984	−0.327	11.842	−3.872
13	24	24	−0.036	0.984	−0.035	11.998	−0.420
Total						109.549	17.675

Note. n^E = placebo group sample size; n^C = control group sample size; d = standardized mean differences. See Equations 3, 4, and 8 for definitions of $C(\nu)$, d^U, and w^U, respectively.

Table 6
Computations for Effect-Size Analysis When the Mean Effect-Size Estimate Is Imputed for Missing Effect-Size Estimates

Study	n^E	n^C	d	$C(\nu)$	d^U	w^U	d^Uw^U
1	19	19	1.028	0.979	1.006	8.432	8.483
2	24	24	0.903	0.984	0.889	10.923	9.711
3	24	24	0.552	0.984	0.543	11.574	6.285
4	24	24	0.328	0.984	0.323	11.846	3.826
5	15	15	−0.540	0.973	−0.525	7.250	−3.806
6	16	16	0.207	0.975	0.202	7.960	1.608
7	12	12	0.207	0.966	0.200	5.970	1.194
8	10	10	0.207	0.958	0.198	4.976	0.985
9	10	10	0.207	0.958	0.198	4.976	0.985
10	12	12	0.193	0.966	0.186	5.974	1.111
11	12	12	−0.660	0.966	−0.638	5.710	−3.643
12	24	24	−0.332	0.984	−0.327	11.842	−3.872
13	24	24	−0.036	0.984	−0.035	11.998	−0.420
Total						109.431	22.447

Note. n^E = placebo group sample size; n^C = control group sample size; d = standardized mean differences. See Equations 3, 4, and 8 for definitions of $C(\nu)$, d^U, and w^U, respectively.

$1.96\sqrt{0.009} \le \delta \le 0.205 + 1.96\sqrt{0.009}$). Because the confidence interval does not include the value zero, the null hypothesis that $\delta_1 = ... = \delta_{13} = \delta = 0$ is rejected.

Buck's method. For Buck's (1960) method, the unbiased effect-size estimate d_i^U was regressed on the degrees of freedom ν_i. In Table 7, the regression equation $\hat{d}_i^U = -0.756 + 0.024\ \nu_i$ was used to obtain the conditional mean for studies with missing estimates. Using the data in Table 7, one obtains the weighted average 0.119 (i.e., Equation 8: 13.019/109.394), the variance 0.009 (i.e., Equation 9: 1/109.394), and the 95% confidence interval [−0.068, 0.306] (i.e., Equation 11: 0.119 −

Table 7
Computations for Effect-Size Analysis When the Conditional Mean Effect-Size Estimate Is Imputed for Missing Effect-Size Estimates

Study	ν	d	$C(\nu)$	d^U	w^U	d^Uw^U
1	36	1.028	0.979	1.006	8.432	8.483
2	46	0.903	0.984	0.889	10.923	9.711
3	46	0.552	0.984	0.543	11.574	6.285
4	46	0.328	0.984	0.323	11.846	3.826
5	28	−0.540	0.973	−0.525	7.250	−3.806
6	30	−0.036	0.975	−0.035	7.999	−0.280
7	22	−0.228	0.966	−0.220	5.964	−1.312
8	18	−0.324	0.958	−0.310	4.941	−1.532
9	18	−0.324	0.958	−0.310	4.941	−1.532
10	22	0.193	0.966	0.186	5.974	1.111
11	22	−0.660	0.966	−0.638	5.710	−3.643
12	46	−0.332	0.984	−0.327	11.842	−3.872
13	46	−0.036	0.984	−0.035	11.998	−0.420
Total					109.394	13.019

Note. The regression equation $\hat{d}_i^U = -0.756 + 0.024\ \nu_i$ was used to obtain the conditional means for Studies 6–9. See Equations 3, 4, and 8 for definitions of $C(\nu)$, d^U, and w^U, respectively.

Table 8
Comparison of Various Procedures for Obtaining an Estimate and Confidence Interval for the Population Standardized Mean Difference (δ)

Procedure	k	Estimator of δ	Variance of estimator	95% CI	CI width
Omit missing estimates	9	0.207	0.012	[−0.005, 0.419]	0.424
Vote counting	13	0.298	0.022	[0.007, 0.589]	0.582
Impute zero for missing estimates	13	0.161	0.009	[−0.026, 0.348]	0.374
Impute mean estimate for missing estimates	13	0.205	0.009	[0.018, 0.392]	0.374
Buck's method	13	0.119	0.009	[−0.068, 0.306]	0.374
Combined procedure	13	0.217	0.011	[0.011, 0.423]	0.412

Note. k = number of independent studies; CI = confidence interval.

$1.96\sqrt{0.009} \leq \delta \leq 0.119 + 1.96\sqrt{0.009}$). Because the confidence interval includes the value zero, the null hypothesis that $\delta_1 = \ldots = \delta_{13} = \delta = 0$ is not rejected.

Table 8 lists the estimates and 95% confidence intervals for the various procedures for handling missing effect-size estimates. Note that the variance is unnecessarily large when studies with missing estimates are deleted and that the variance is unnecessarily small when single or conditional values are imputed for missing estimates. Note also that the estimate of δ is overly conservative when the value zero is imputed for missing estimates. Finally, note that the combined procedure confidence interval is shorter than are the effect-size and vote-counting procedure confidence intervals.

Advantages of the Combined Procedure Over Conventional Procedures for Handling Missing Effect-Size Estimates

Missing effect-size estimates pose a particularly difficult problem in meta-analysis. Often studies do not include enough information to permit the computation of an effect-size estimate. The combined procedure described in this article has at least five advantages over conventional procedures for dealing with the problem of missing effect-size estimates. First, the combined procedure uses all the information available from studies in a research synthesis; other procedures ignore information about the direction and statistical significance of results. Second, the combined estimator is consistent (i.e., if the number of studies k under review is large, the combined estimator will not overestimate or underestimate the population effect size). None of the conventional procedures for handling missing effect-size estimates are consistent; they all either underestimate or overestimate the population effect size. Third, the combined estimator is more efficient than either the effect-size or vote-counting estimators (i.e., the variance of the combined estimator is smaller than variance of the effect-size and vote-counting estimators). Fourth, the variance of the combined estimator is known. All of the conventional procedures for handling missing effect-size estimates either artificially inflate or artifically deflate the variance of the effect-size estimates. Fifth, the combined procedure gives weight to all studies proportional to the Fisher information they provide (i.e., no studies are overweighted). This article focused on the standardized mean difference in fixed effects models, although the combined procedure can be used with any effect-size index in either fixed or random effects models. We currently are working on extending the combined procedure to other effect-size indexes and to random effects models. In summary, the combined procedure seems to be the method of choice in handling missing effect-size estimates when some studies do not provide enough information to compute effect-size estimates but do provide information about the direction or statistical significance of results.

References

Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society, Series B, 22,* 302–303.

Bushman, B. J. (1993). Human aggression while under the influence of alcohol and other drugs: An integrative research review. *Current Directions in Psychological Science, 2,* 148–152.

Bushman, B. J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 193–213). New York: Russell Sage Foundation.

Bushman, B. J. (in press). Effects of alcohol on human aggression: Validity of proposed explanations. In D. Fuller, R. Dietrich, & E. Gottheil (Eds.), *Recent developments in alcoholism: Alcohol and violence* (Vol. XIII). New York: Plenum.

Bushman, B. J., & Cooper, H. M. (1990). Effects of alcohol on human aggression: An integrative research review. *Psychological Bulletin, 107,* 341–354.

Bushman, B. J., & Wang, M. C. (1995). A procedure for combining sample correlation coefficients and vote counts to obtain an estimate and a confidence interval for the population correlation coefficient. *Psychological Bulletin, 117,* 530–546.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82,* 1–20.

Guenther, W. C. (1978). Evaluation of probabilities for the non-central distributions and the differences of two t-variables with a desk calculator. *Journal of Statistical Computation and Simulation, 6,* 199–206.

Hawkins, D. M. (1975). From the central t to the normal integral. *American Statistician, 29,* 42–43.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V. (1986). *Estimating effect-sizes from vote counts or box score data.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88,* 359–369.

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

Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions.* Boston: Houghton Mifflin.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review, 41,* 429–471.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

MacAndrew, C., & Edgerton, R. B. (1969). *Drunken comportment: A social explanation.* Chicago: Aldine.

Pigott, T. D. (1994). Methods for handling missing data in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 163–175). New York: Russell Sage Foundation.

Appendix A

Computation Details for Obtaining an Estimator of the Population Standardized Mean Difference Using Unequal Sample Size Vote-Counting Procedures

Under the alternative hypothesis in Equation 12, the probability density function of the sample effect size t_i (Equation 13) is

$$f_{\delta_i',\nu_i}(t_i) = \frac{e^{-(\delta_i'^2/2)}\left(1+\frac{t_i^2}{\nu_i}\right)^{-(\nu_i+1)/2}}{\sqrt{\nu_i}B\left(\frac{1}{2},\frac{\nu_i}{2}\right)} \sum_{j=0}^{\infty} d_j\left[\frac{\sqrt{2}\delta_i' t_i}{\sqrt{\nu_i+t_i^2}}\right]^2, \quad \text{(A1)}$$

where

$$d_j = \frac{\Gamma\left[\frac{1}{2}(\nu_i+j+1)\right]}{j!\Gamma\left[\frac{1}{2}(\nu_i+1)\right]},$$

$$\delta_i' = \sqrt{\frac{n_i^E n_i^C}{n_i^E+n_i^C}}\,\delta,$$

$\nu_i = n_i^E + n_i^C - 2$, $B(.,.)$ is the complete beta function, and $\Gamma(.)$ is the complete gamma function (Johnson & Kotz, 1970). Under the null hypothesis given in Equation 12, the probability density function given in Equation A1 reduces to

$$f_{\delta_i'=0,\nu_i}(t_i) = \frac{\left(1+\frac{t_i^2}{\nu_i}\right)^{-(\nu_i+1)/2}}{\sqrt{\nu_i}\mathrm{B}\left(\frac{1}{2},\frac{\nu_i}{2}\right)}. \quad \text{(A2)}$$

The probability of a positive result is

$$p_i = \Pr_{\delta_i',\nu_i}(t_i>0) = \int_0^{\infty} f_{\delta_i',\nu_i}(t_i)\,dt = \Phi(\delta_i'), \quad \text{(A3)}$$

where $\Phi(.)$ is the standard normal distribution function (Hawkins, 1975). The probability of a significant positive result at significance level α is

$$p_i = \Pr_{\delta_i',\nu_i}(t_i>t_{\alpha,\nu_i}) = \int_{t_{\alpha,\nu_i}}^{\infty} f_{\delta_i',\nu_i}(t_i)\,dt,$$

$$= \Phi(\delta_i') - \Pr_{\delta_i',\nu_i}(0<t_i<t_{\alpha,\nu_i}), \quad \text{(A4)}$$

where t_{α,ν_i} is the one-sided critical value from the distribution of t such that $\alpha = \int_{t_{\alpha,\nu_i}}^{\infty} f_{\delta_i'=0,\nu_i}(t_i)\,dt$, and $0<\alpha<.5$. To evaluate Equation A4, t_{α,ν_i} must first be obtained by solving the equation:

$$\alpha = \int_{t_{\alpha,\nu_i}}^{\infty} f_{\delta_i'=0,\nu_i}(t_i)\,dt$$

$$= 1 - \int_{-\infty}^{t_{\alpha,\nu_i}} f_{\delta_i'=0,\nu_i}(t_i)\,dt$$

$$= 1 - \frac{1}{2} - \frac{1}{2}I\left(\frac{1}{2},\frac{\nu_i}{2},x_i\right)$$

$$= \frac{1}{2}I\left(\frac{\nu_i}{2},\frac{1}{2},1-x_i\right), \quad \text{(A5)}$$

where $I(.,.,.)$ is the incomplete beta function and $x_i = t_{\alpha,\nu_i}^2/t_{\alpha,\nu_i}^2 + \nu_i$. After t_{α,ν_i} has been obtained, the probability given in Equation A4 can be evaluated as follows:

$$\Pr_{\delta_i',\nu_i}(t_i>t_{\alpha,\nu_i}) = \Phi(\delta_i') - \Pr_{\delta_i',\nu_i}(0<t_i<t_{\alpha,\nu_i})$$

$$= \Phi(\delta_i') - \sum_{j=0}^{\infty}\left[a_j I\left(j+\frac{1}{2},\frac{\nu_i}{2},x_i\right) + b_j I\left(j+1,\frac{\nu_i}{2},x_i\right)\right], \quad \text{(A6)}$$

where

$$a_j = \frac{1}{2}e^{-(\delta_i'^2/2)}\frac{(\delta_i'^2/2)^j}{j!}$$

and

$$b_j = \frac{1}{2}\delta_i' e^{-(\delta_i'^2/2)}\frac{(\delta_i'^2/2)^j}{\sqrt{2}\Gamma(j+3/2)}$$

(Guenther, 1978).

Appendix B follows on next page.

Appendix B

Proof That the Combined Estimator Has Smaller Variance Than Does the Effect-Size Estimator and Vote-Counting Estimator

$$\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(d_+^U)$$

$$\mathrm{Var}(\hat{\delta}_C) - \mathrm{Var}(d_+^U)$$

$$= \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})} - \mathrm{Var}(d_+^U)$$

$$= \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta}) - \mathrm{Var}(d_+^U)^2 - \mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})}$$

$$= \frac{-\mathrm{Var}(d_+^U)^2}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})} < 0$$

$$\Rightarrow \mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(d_+^U)$$

$$\mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(\hat{\delta}) \qquad \text{(B1)}$$

$$\mathrm{Var}(\hat{\delta}_C) - \mathrm{Var}(\hat{\delta})$$

$$= \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})} - \mathrm{Var}(\hat{\delta})$$

$$= \frac{\mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta}) - \mathrm{Var}(\hat{\delta})^2 - \mathrm{Var}(d_+^U)\mathrm{Var}(\hat{\delta})}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})}$$

$$= \frac{-\mathrm{Var}(\hat{\delta})^2}{\mathrm{Var}(d_+^U) + \mathrm{Var}(\hat{\delta})} < 0$$

$$\Rightarrow \mathrm{Var}(\hat{\delta}_C) < \mathrm{Var}(\hat{\delta}) \qquad \text{(B2)}$$

Received April 4, 1995
Revision received August 8, 1995
Accepted August 11, 1995 ■