Journal of Applied Psychology
1994, Vol. 79, No. 3, 425–438

Nonlinearity of Range Corrections in Meta-Analysis: Test of an Improved Procedure

Kenneth S. Law, Frank L. Schmidt, and John E. Hunter

The authors evaluated an improved procedure for range-restriction corrections in meta-analysis. When population correlations were approximately normally distributed, the new nonlinear range-correction procedure improved the accuracy of the Schmidt–Hunter (S-H) interactive method in estimating both the mean (M_ρ) and standard deviation (SD_ρ) of population correlations, making it the most accurate of the procedures examined. In the homogeneous case ($SD_\rho = 0$), the nonlinear range correction again improves accuracy of estimates of SD_ρ. In this important case, Taylor Series Approximation (TSA) 1 is considerably less accurate than the S-H interactive and the TSA2 procedures, and N. S. Raju, M. J. Burke, J. Normand, and G. M. Langlois's (1991) procedure yields the least accurate estimates of SD_ρ. Finally, the authors found that the nonlinear range-correction procedure produces improvements in the accuracy of the interactive method even under extreme violations of the normality assumption.

During the past decade, many meta-analyses have been published in a variety of areas (Hunter & Schmidt, 1990b), with most analyses being conducted on correlation coefficients. Because such meta-analyses form the basis of conclusions about cumulative knowledge, the question of the accuracy of meta-analytic procedures in estimating the true mean and variance of population correlations is important. Studies have been conducted to examine the accuracy of different meta-analysis procedures (see e.g., Raju & Burke, 1983). Because data on reliability and range restriction are only sporadically available in individual studies, these computer simulation studies focus on meta-analysis methods that use distributions of artifacts (Hunter & Schmidt, 1990b, chap. 4). Such meta-analysis methods do not correct each correlation individually for the effects of artifacts but, rather, correct distributions of correlations using distributions of artifacts that are characteristic of the research literature in question. For these methods, the nonlinearity of range-restriction effects has caused all existing estimation procedures to suffer some degree of error. Past computer simulation studies (see e.g., Callender & Osburn, 1980; Mendoza & Reinhardt, 1991; Raju & Burke, 1983) have generally concluded that although the artifact-distribution-based meta-analysis procedures studied were fairly accurate, they slightly overestimate the variance of true population correlations. Raju and Burke (1983) concluded that "there is still room for improvement in methods to assess validity generalization" (p. 392). In this study, a refinement is proposed in the Schmidt–Hunter meta-analysis procedure with artifact distributions, and the resulting method is tested for accuracy in a simulation study. The refinement is a nonlinear range-restriction correction performed on the residual distribution to estimate the true population distribution of correlations (Hunter & Schmidt, 1990b, pp. 209–211). This method derives estimates of both the mean and the standard deviation of the distribution of true population correlations from the residual distribution.

There are two parts to meta-analysis: correcting for the unsystematic effects of sampling error and correcting for the systematic effects of other artifacts. Assume for the moment that correction for sampling error has been made; that is, the distribution of sample correlations has been analyzed to estimate the distribution of attenuated study population correlations. In this distribution, the actual study correlations have been attenuated in various degrees, depending on the extent of error of measurement, the extent of range restriction, and so on. Various methods of meta-analysis use different strategies for correcting for the effects of those artifacts.

The original Schmidt–Hunter strategies—the interactive (Schmidt, Gast-Rosenberg, & Hunter, 1980) and noninteractive (Schmidt, Hunter, Pearlman, & Shane, 1979) methods—start from a special case: homogeneous artifact values. If all studies had been conducted with the level of each artifact at a constant value, then the distribution of actual (true score) correlations could be found by correcting each point in the resulting distribution of study population correlations using the uniform artifact values. The Schmidt–Hunter methods handle the case of heterogeneous artifact values in two steps: (a) they transform the correlation distribution yielded by heterogeneous artifacts into the distribution that results from homogeneous artifacts (this step yields the residual distribution), and (b) they correct the estimated homogeneous artifact distribution (residual distribution) using a linear correction procedure. In the multiplicative method of Callender and Osburn (1980), both of these steps are carried out simultaneously. The multiplicative method does not produce a separate estimate of the residual standard

---

Kenneth S. Law (now at the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong) and Frank L. Schmidt, Department of Management and Organizations, College of Business, University of Iowa; John E. Hunter, Department of Psychology, Michigan State University.

Each of us contributed equally to this article.

Correspondence concerning this article should be addressed to Frank L. Schmidt, Department of Management and Organizations, College of Business, University of Iowa, Iowa City, Iowa 52242-1323.

deviation; indeed, the concept of a residual distribution is not defined in the multiplicative method. Therefore, the nonlinear range-correction procedure examined in this article cannot be applied to the multiplicative method (because the procedure is applied to the residual distribution). As a consequence of their deviation (using the Taylor Series Approximations 1 [TSA1] and 2 [TSA2]), the TSA methods of Raju and Burke (1983) contain a built-in approximation to our (integral transform) correction for the nonlinearity of the range-restriction correction. Thus application of our nonlinear refinement would represent a double correction for this effect. Because of this built-in approximate correction, we expected Raju and Burke's TSA methods to be closest in accuracy to the interactive procedure with the nonlinear range-correction procedure added. In passing, we note that the TSA2 formula used in this study was corrected for the slight error in the earlier formula, as discussed by Raju, Burke, Normand, and Langlois (1991, p. 436). Raju et al. noted that correction of this error increases the accuracy of the variance estimates.

Raju et al. (1991) recently introduced a new approach for meta-analysis designed initially for cases in which artifact information is available for each correlation and each correlation is corrected individually. However, they stated that their new method is also very accurate for use with artifact-distribution-based meta-analysis, and therefore, we include an evaluation of that usage of their method in the present study. We refer to this method as the *RBNL91 method* to distinguish it from their two previous methods (TSA1 and TSA2). Raju et al. compared their new method only to TSA1; in this study we compared the accuracy of the RBNL91 method to all of the previously existing methods. Our comparisons are made under conditions that Raju et al. called *Condition 4,* the situation in which artifact values are not available for each individual correlation and artifact distributions are used to make the meta-analysis corrections. Because in the present study there is no sampling error (i.e., all *N*s are infinite), application of the RBNL91 method is very straightforward for Condition 4: Using the mean artifact values for each artifact, one simply corrects each observed (attenuated) population correlation. The estimated mean true correlation is then the mean of the corrected correlations, and the estimated standard deviation of the true correlations is the standard deviation of the corrected correlations (Raju et al., 1991).

On an a priori basis, it appears that the RBNL91 method might not be as accurate as some of the other methods. Consider the mean correlation. Because the range-restriction correction formula is concave, the mean corrected correlation must underestimate the mean true correlation. The greater the variance of the correlations to be corrected, the greater the error of estimation. Correcting the mean attenuated correlation for range restriction using the mean u value should consistently be more accurate than correcting the individual correlations using the mean u value. ($u = s/S$, where s = the restricted standard deviation and S = the unrestricted standard deviation.)

Second, consider the estimated variance of true correlations. The RBNL91 method makes no correction for variance in correlations due to variation in artifact values. Thus it attributes all of the variance in the corrected correlations (beyond sampling error variance, which was not present here) to variance in the true correlations. This would be expected to lead to a consistent overestimation of the standard deviation of true correlations. This overestimation should be most obvious in the case in which there is no real variance in true correlations (SD_ρ = .00), because in that case the relative error in attributing artifact variance to true variance would not be obscured by actual variance in the true correlations.

In most research domains, there is considerable variation in each artifact (i.e., variation in reliability, variation in extent of range restriction, and so on). This heterogeneity in artifact values transforms the actual population correlations into attenuated study-population correlations. Suppose, instead, that every study had been done so that each artifact value was at the same level: the level corresponding to the mean artifact value for the heterogeneous domain of that artifact (e.g., mean dependent variable reliability). The artifact distributions would then be homogeneous across studies. The resulting hypothetical distribution of attenuated study-population correlations has been called the *residual distribution.* Using different procedures, the Schmidt–Hunter interactive and noninteractive methods estimate the mean and variance of the residual distribution given the mean and variance of the unrefined distribution (the distribution of correlations corrected only for sampling error). In the present article, we do not propose changes in this aspect of these methods; that is, the method that we propose and test does not change the way the residual distribution is computed. Instead, it proposes a change in the way the residual distribution is used to estimate the mean and standard deviation of the distribution of true population correlations.

The second step is to correct the estimated residual distribution using the mean artifact values. This step is identical in the interactive and noninteractive methods. For linear factors such as attenuation due to random error, artificial dichotomization (Hunter & Schmidt, 1990a), or imperfect construct validity, the attenuation is, mathematically, multiplication by a constant; and correction (disattenuation) is multiplication by the reciprocal of that constant. But the attenuation produced by range restriction is not linear, and that fact has caused accuracy problems for all methods of meta-analysis. In the original Schmidt–Hunter methods, correction for range restriction is accomplished by using a linear approximation. Attenuation is estimated by the ratio of the mean attenuated correlation to the mean unattenuated correlation, and that ratio is the mean attenuation factor for range restriction. Correction is then accomplished by multiplying the mean and the standard deviation of the residual distribution by the reciprocal of that ratio. The standard deviation can be more accurately estimated by using a ratio modified by Taylor's series, such as that of Raju and Burke (1983). (See Hunter & Schmidt, 1990b, pp. 130–132, for a discussion of such use in meta-analysis of individually corrected correlations.) In the present article, we propose a more direct procedure that is mathematically known as an *integral transform.*

In the new procedure, correction is not applied only to the mean and the standard deviation; rather, each point in the residual distribution is corrected for range restriction with the exact nonlinear range-restriction correction procedure. Each point is also corrected for mean criterion unreliability. As a result, the standard deviation of this corrected distribution estimates the standard deviation of true validities (SD_ρ) and the mean of this distribution estimates $\bar{\rho}$, called *mean true validity* in validity generalization studies. Outside the validity generalization area, corrections are also made for mean unreliability in the independent variable measure. The resulting standard deviation (SD_ρ) then estimates the standard deviation of true score correlations, and the $\bar{\rho}$ value estimates the mean true score correlation. The formal mathematics of the nonlinear correction procedure are presented in the Appendix.

If the residual distribution were known exactly, then the new procedure would be perfectly accurate. Any error in the new procedure must stem from error in the estimation of the residual distribution. There are two potential sources of such error. First, the interactive and noninteractive methods do not perfectly transform the mean and standard deviation from the correlation distribution corrected only for sampling error, and this fact creates some error in the estimated mean and standard deviation of the residual distribution. Second, the exact shape of the distribution of actual (true score) correlations is not known, and therefore the exact shape of the residual distribution is not known. The nonlinear range-correction procedure described in this article uses the normal distribution as an approximation. This is roughly equivalent to assuming a normal distribution for the true score correlations (although because of the nonlinearity of range restriction, a normal, true score correlation distribution does not attenuate to a perfectly normal residual distribution). If the true score correlations vary because of the effects of a large number of small moderator variables, then the normal distribution will be a good approximation (see below). In this study, we tested for the effects of violating the normality assumption by incorporating some strongly nonnormal distributions into our Monte Carlo design.

Thomas's (1990) mixture-decomposition procedure for meta-analysis cannot logically be used if there is any variation across studies in reliability levels (for independent or dependent variable measures), range-restriction levels, or other artifacts that systematically affect the size of observed correlations (Schmidt, Ones, & Hunter, 1992). Thus it could not be compared in this study with the two Schmidt–Hunter procedures, the Callender–Osburn multiplicative procedure, the two Raju–Burke TSA procedures, and the Raju et al. (1991) procedure. Thomas's mixture-decomposition procedure is based on analyses of only observed correlations (correlations uncorrected for measurement error, range restriction, or other statistical artifacts that systematically affect the size of the observed correlations). The method makes no corrections for the effects of these artifacts, either initially or at a later stage in the analysis. Because of this, Thomas's method treats variation in correlations that is due to these artifacts as if it were real variance in the underlying population correlations. This can result (a) in the identification of multiple population-correlation values that do not, in fact, exist and (b) in overestimation of the variance of population correlations (as a result of a; Schmidt et al., 1992).

Law (1992) used computer simulation to examine the accuracy of Thomas's procedure under the conditions for which it was designed, that is, conditions in which there is no variability across studies in measurement error, range restriction, or other systematic artifacts. Even under these restricted conditions (highly unlikely in real data sets), the procedure frequently encountered problems in correctly identifying the underlying population correlations.

Method

In both the noninteractive procedure and the interactive procedure of the Schmidt–Hunter model, the residual distribution is defined as the distribution of population correlations that would have been observed if reliability and range-restriction levels were uniform across studies at their mean values and sample sizes in every study were infinite. Both procedures produce an estimated mean and an estimated standard deviation for the residual distribution (e.g., see Schmidt et al., 1980). We do not propose any changes in the way these estimates are calculated. To find the distribution of population true score correlations from this attenuated distribution, each value in the residual distribution should be corrected using the mean reliability and mean range-restriction values. However, the original Schmidt–Hunter model assumes that the distribution of true population correlations is equal to the residual distribution multiplied by a constant factor. This factor is estimated as the ratio of the mean corrected population correlation to the mean observed correlation. That is, both the mean and the standard deviation of the residual standard deviation are multiplied by this constant, yielding the estimated mean and standard deviation of the population true score correlations. This procedure is only approximate, because the desired correction is not a constant factor for all values of the correlation in the residual distribution. For the unreliability corrections, the appropriate correction factor is a constant, namely, $[E(r_{yy}^{\frac{1}{2}})E(r_{xx}^{\frac{1}{2}})]^{-1}$ times each correlation value in the residual distribution. However, for the range-restriction correction, the correction is not linear and, hence, the correction factor is not constant throughout the range of correlations. Specifically, the (nonlinear) range-restriction correction is larger for smaller correlations and smaller for larger correlations. Thus, the greater the variability of the correlation is in the residual distribution, the greater the effect of the nonlinearity of the correction. The refinement in this study focuses on the nonlinear range-restriction correction effect. One possible way of taking this nonlinear range-restriction correction effect into consideration is to first assume that the residual distribution is approximately normal (see discussion below). Then, using the estimated mean and standard deviation of the residual distribution, one chooses an appropriate sample of points from this normal distribution and corrects these values individually and separately. The frequency-weighted mean and variance of the corrected points are then used to estimate the mean and variance of the population correlations. Thus, estimates of both the mean and the variance can be different when this procedure is used versus when it is not.

How many points in the distribution should be used? A number that is too small may not be representative of the whole distribution; a number that is too large will require unnecessary computer time. To reduce the error in the normal distribution to a low level, we chose to use 30 equally spaced points above and below the mean. These points, plus the mean, yield 61 values to be corrected. Researchers who desire to do so could increase or decrease the number of points. Within broad limits, variations in the number of points used will have negligible effects on the results.

Our method assumes that the distribution of population correlations in the residual distribution is normal. The assumption of normality is commonly made in social science research under circumstances in which it cannot be known with certainty to be correct or in which it is only an approximation. Cohen (1983) noted that "there is abundant evidence throughout applied psychometrics and statistics that the failure of the normality assumption, unless extreme, bears only marginally on the validity of the conclusions drawn" (p. 252). The rationale for the normality assumption here is that, in real data, residual variation is likely to be caused by either a number of essentially uncorrelated, uncorrected artifacts; by a large number of small and essentially independent moderators; or by some combination of both. Distributions that are the product of a large number of essentially uncorrelated causes tend toward normality (Hunter & Schmidt, 1990b). This logic provides a rationale for the normality assumption. Moreover, in this study the effects of violating this assumption on the accuracy of the resulting estimates is examined through computer simulation studies. Specifically, if the distribution of true score correlations is markedly nonnormal, then the residual distribution will also be markedly nonnormal. The first such case that we examined is the case in which the distribution of population true score correlations is rectangular (ranging with equal frequency from .01 to .99). The second case is that in which the population distribution is dichotomous; that is, when there are only two values for the unattenuated population correlation. The third case examined is one in which the distribution of population true scores is (positively) skewed. Finally, we also examined the homogeneous case, in which the population correlation has a point distribution; that is, when $SD_\rho = .00$. These analyses allowed us to determine whether error was introduced into meta-analysis estimates by violation of the normality assumption.

The Monte Carlo methods used by Callender and Osburn (1980) and by Raju and Burke (1983) in their studies were, with some modification, used here; readers are referred to the original articles for detailed descriptions of these simulation methods. As in those studies, the simulations were conducted using infinite sample sizes, because the concern was with the accuracy of corrections for artifacts other than sampling error. That is, sampling error was excluded from the simulation study to allow sharper focus on the accuracy of corrections for other artifacts. The different meta-analysis methods have used the same formula for sampling error variance. Accuracy of the sampling error corrections is not at issue in meta-analysis; it has been examined elsewhere (e.g., Callender & Osburn, 1988) and has been found to be quite high.

Five of the six analyses examined here have positive nonzero values for SD_ρ, the standard deviation of population true score correlations. In light of the cumulative evidence against situational specificity for aptitude and ability tests in employment selection (Schmidt, Hunter, Pearlman, & Hirsch, 1985; Schmidt et al., 1993), the case in which $SD_\rho = 0$ (the homogeneous case) is realistic in that area, and we examined that case in the present study. However, when $SD_\rho = 0$, the expected estimate of the residual distribution standard deviation is 0. Hence, there should be only one value in the residual distribution and, therefore, there should be no difference between the nonlinear and linear range-restriction corrections. Thus results for the case in which $SD_\rho = 0$ should be the same for the two procedures. However, this is true only if the procedures always estimate the residual standard deviation as 0 when $SD_\rho = 0$. If, despite the refinement, the procedures continue to have a positive bias, then the estimate of residual standard deviation will be a small positive number, and the nonlinear procedure will be slightly more accurate than the linear procedure, even though $SD_\rho = 0$. Finally, the homogeneous case represents examination of yet another violation of the normality assumption; namely, the case in which the distribution of population correlations is a point distribution.

Analysis 1

In the first analysis, we used the wide and approximately normal distribution of 23 population correlations that were presented by Callender and Osburn (1980, Table 3) and were also used by Raju and Burke (1983, Table 4). Both of these previous studies used random sampling without replacement. For example, in Raju and Burke's (1983) article, each observed correlation was produced by a random draw without replacement from the distribution of 23 population correlations and from each of the three artifact distributions (criterion reliability, predictor reliability, and range restriction). One value was drawn from each of the four distributions, and an observed correlation was then calculated from the randomly selected population correlation, by using the randomly drawn values for predictor and criterion reliability and range restriction. The process was repeated until all of the 100 values in each distribution were drawn; thus, sampling was done without replacement. However, we used a different procedure that should be more accurate: We applied the meta-analysis procedures to all possible combinations of population correlations and artifact levels.

In reality, most studies are conducted independently. Thus the artifact values in one study are not influenced by those in another. This is mirrored in all current methods of meta-analysis based on artifact distributions: All assume independence of artifacts. The sampling-without-replacement Monte Carlo strategy used by Callender and Osburn (1980) and by Raju and Burke (1983) violates the independence assumption. Thus, the Monte Carlo results in these two studies are not exactly accurate. We eliminated that source of error and also eliminated any influence of random error by doing all computations across the entire population of actual correlations as attenuated by all possible combinations of artifacts, with each combination weighted by exactly the weight specified in the corresponding distribution. Our Monte Carlo estimates are thus subject only to rounding error. We used double-precision arithmetic to reduce rounding error to about the eighth digit overall.

There were 23 values for the population correlation, 13 levels of criterion reliability, 8 levels of range restriction, and 7 levels of predictor reliability. (The three artifact distributions used are given in Raju & Burke, 1983, and in Schmidt et al., 1980.) Thus, each meta-analysis was performed on a total of 16,744 unique observed correlations. Each of these correlations was weighted by its appropriate frequency, which is determined as the product of the relative frequencies of the particular population correlation, the particular level of criterion reliability, the particular level of range restriction, and the particular level of predictor reliability. This weighting procedure is consistent with the usual assumption of artifact independence (Callender & Osburn, 1980; Hunter & Schmidt, 1990b, chap. 3; Raju & Burke, 1983).

Eight meta-analysis methods were then applied to these observed population correlations: the two Schmidt–Hunter procedures with linear and nonlinear range correction (four methods), the Callender and Osburn multiplicative procedure (Callender & Osburn, 1980), the two Taylor Series procedures (Raju & Burke, 1983), and the RBNL91 method. Except for the difference in sampling procedures described here, Analysis 1 corresponds to Case 3 in both Callender and Osburn and Raju and Burke. Because the distribution of population true score correlations is approximately normal, the residual distribution will also be approximately normal. Hence, this analysis provides a test of the accuracy of the new procedure when the normality assumption is met.

Analysis 2

In Analysis 2, the distribution of unattenuated population correlations was rectangular, providing a test of the accuracy of the new procedure when the normality assumption is seriously violated. A rectangular distribution of population true score correlations results in a flat and nearly rectangular residual correlation distribution. Analysis 2 is identical to Case 2 of Callender and Osburn (1980) and Raju and Burke (1983). In each meta-analysis, the artifactual effects were fixed. Three values of criterion reliability, predictor reliability, and range restriction were selected from the hypothetical distribution of artifacts used in those studies. The values selected were 1.0, .80, and .50 for predictor reliability; 1.0, .30, and .60 for criterion reliability; and 1.0, .411, and .603 for range-restriction ratio (the ratio of the restricted to the unrestricted standard deviation of the independent variable). These nine values were completely crossed, resulting in 27 different combinations. Each combination of the three artifacts was then applied to 100 population correlations, which were generated by starting with a population correlation of .01 and incrementing by .01 up to the final value of .99. The mean population true score correlation was therefore .4950, and the true standard deviation was .2886. On each of the 27 sets of 100 observed correlations, meta-analyses were performed using each of the meta-analysis procedures listed above except for the RBNL91 method, which was not meaningful in this context. As noted earlier, the RBNL91 method corrects each correlation for the effects of a particular artifact by using the average value for the artifact. In this analysis, the value for each artifact was a constant within any given meta-analysis. Thus, it was known in advance that the method would have perfect accuracy.

Analysis 3

Analysis 3 was also designed to examine the effects on accuracy of violation of the normality assumption. In cases in which a single, two-valued moderator is operating, the distribution of unattenuated (true score) population correlations will be a dichotomous, two-point distribution. This markedly nonnormal distribution will lead to a very nonnormal (bimodal) distribution of residual correlations. Three dichotomous distributions of true score population correlations were examined: $\rho_1 = .50$ and $\rho_2 = .00$; $\rho_1 = .40$ and $\rho_2 = .10$; and $\rho_1 = .30$ and $\rho_2 = .20$. In each dichotomy, the two values of ρ had equal (50%) frequencies. The artifacts were the same as those in Analysis 2; for each of these 27 artifact combinations (shown in Table 2) an observed correlation was created twice—once for ρ_1 once for ρ_2—and the resulting correlations were meta-analyzed by using each of the eight meta-analysis procedures described above.

Analysis 4

Analyses 1 through 3 were all based on symmetrical distributions of true score population correlations. The purpose of Analysis 4 was to determine whether a violation of the normality assumption that took the form of a (nonsymmetrical) skewed distribution of true score population correlations would induce error in the nonlinear range-correction procedure. The skewed distribution of the true score population correlation that we used was as follows: For population correlations of .60, .50, .40, and .30, relative frequencies were 10, 20, 30, and 40, respectively; mean $\rho = .40$, $S_\rho^2 = .01000$, $SD_\rho = .10000$. Again, the artifacts were the same as those in Analysis 2; for each of these 27 artifact combinations (shown in Table 2) an observed correlation was created for each of the four values in this population correlation distribution. The resulting correlations, weighted by relative frequency, were then meta-analyzed by using each of the eight meta-analysis procedures.

Analysis 5

The fifth analysis was conducted to address a different concern. The original, approximately normal distribution of 23 population correlations in Callender and Osburn's (1980) study was quite wide, perhaps too wide to approximate any realistic distribution of population correlations: It ranged from .06 to .94. Such a large dispersion of population correlations would probably be unlikely in real data from studies on any research question. Therefore, Analysis 1 was rerun with the following, narrower distribution of population correlations: For population correlations of .40, .45, .50, .55, and .60, the relative frequencies were .10, .20, .40, .20, and .10; $\text{mean}_\rho = .50$, $S_\rho^2 = .00300$, $SD_\rho = .05477$. Except for the use of this new distribution of population correlations, Analysis 5 was conducted in the same manner as Analysis 1. But because there were only five different values for the population correlation in Analysis 5 (vs. 23 in Analysis 1), the total number of population correlation–artifact combinations was much smaller: 5 levels of population correlation × 13 levels of criterion reliability × 8 levels of range restriction × 7 levels of predictor reliability = 3,640 combinations (vs. 16,744 combinations in Analysis 1).

Analysis 6

The sixth and final analysis was conducted to determine how accurately the various procedures were when $SD_\rho = 0$; that is, in the homogeneous case. Another purpose of this analysis was to determine whether the violation of the normality assumption represented by a point distribution of the population true score correlation would cause inaccuracy in the nonlinear range-correction procedure.

In this analysis, the true correlation is always .50 and thus the standard deviation of true correlations is always 0. We considered two levels of variation in each of the three artifacts examined in this article: a high level of variation and a low level of variation. For each artifact, the high level of variation has exactly twice the standard deviation of the low level. The specific values for the artifacts are as follows: for criterion reliability (r_{yy}), low variabilities are .40, .60, and .80, and high variabilities are .20, .60, and 1.00; for predictor reliability (r_{xx}), low variabilities are .70, .80, and .90, and high variabilities are .60, .80, and 1.00; for range restriction ($u = \frac{s}{S}$), low variabilities are .51, .67, and .83, and high variabilities are .35, .67, and .99. There are two levels of variation on each of three artifacts; hence, there are eight artifact combinations in all, each one containing 27 observed correlations. All eight methods of meta-analysis were applied to each of the eight combinations.

Results and Discussion

Analysis 1

Results for Analysis 1 are shown in Table 1. Estimates of the mean correlation were identical for the two Schmidt–Hunter

Table 1

Analysis 1: Estimated Means and Standard Deviations of Population True Correlations When the Population Rhos Are Quite Variable and Approximately Normally Distributed and Reliability and Range-Restriction Values Are Variable

Procedure	Estimate	Error	% error
M_ρ estimates (actual $M_\rho = .5000$)			
S-H int. & nonint. linear	.5079	.0079	1.6
S-H int. nonlinear	.5000	.0000	0.0
S-H nonint. nonlinear	.5017	.0017	0.3
C-O multiplicative	.5069	.0069	1.4
TSA1	.5032	.0032	0.6
TSA2	.5079	.0079	1.6
RBNL91	.49563	−.0044	−0.8
SD_ρ estimates (actual $SD_\rho = .1754$)			
S-H int. linear	.1988	.0234	13.3
S-H nonint. linear	.1733	−.0021	−1.2
S-H int. nonlinear	.1799	.0045	2.6
S-H nonint. nonlinear	.1572	−.0182	−10.4
C-O multiplicative	.1972	.0218	12.4
TSA1	.1811	.0057	3.3
TSA2	.1820	.0066	3.8
RBNL91	.20358	.0282	16.1

Note. S-H int. = Schmidt, Gast-Rosenberg, & Hunter (1980) interactive procedure; S-H nonint. = Schmidt, Hunter, Pearlman, & Shane (1979) noninteractive procedure; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983); RBNL91 = Raju, Burke, Normand, & Langlois (1991) procedure.

methods used with the linear range correction. Introduction of the nonlinear range correction improved the estimates of the mean correlation for both the interactive and noninteractive procedures. For the noninteractive, the error declined from .0079 (a 1.6% error) to .0017 (a .3% error); for the interactive, the figure was .0079 (1.6% error), declining to .0000 error. Although all procedures gave fairly accurate estimates of the mean population correlation, the Schmidt–Hunter procedures (interactive and noninteractive) with nonlinear range restriction gave, on average, the most accurate estimates; errors here were within rounding error for three decimal places. The least accurate procedures were the Schmidt–Hunter procedures with linear range corrections and the TSA2 procedure; however, even for these procedures the percentage error was only 1.6%, resulting in a final rounded estimate of .51 instead of the correct value of .50. As predicted, the RBNL91 method underestimated the mean correlation—unlike the other procedures—but this effect was negligible.

Turning to the estimates of the true standard deviation of population correlations, one can see that, for the interactive procedure, introduction of the nonlinear range correction again improved accuracy. Error declined from .0234 (a 13.3% error) with the linear correction to .0045 (a 2.6% error) with the nonlinear correction. However, for the noninteractive procedure, error increased from −.0021 (a 1.2% error) to −.0182 (a 10.4% error) when the nonlinear correction was introduced.

These results indicate that the nonlinear range-correction procedure substantially improves accuracy for the interactive procedure when the normality assumption is met and the variance of population true score correlations is substantial. Procedures other than the noninteractive procedure overestimate the standard deviation on average, and of these, the interactive nonlinear procedure was most accurate, followed by the two Raju and Burke (1983) TSA procedures, which are only marginally less accurate in this analysis than the interactive nonlinear procedure. The RBNL91 method yields the least accurate estimates of the standard deviation of population true score correlations.

Although all the other procedures overestimate the true variance of population correlations, the noninteractive procedure yields slight underestimates of the standard deviation. This is because the noninteractive procedure, unlike other procedures, assumes that the effects (not the artifact values, but the effects of the artifacts on the correlation) of the predictor and criterion unreliability and range restriction are all independent (Schmidt et al., 1980), resulting in a slight overestimation of artifactual variances; a slight underestimation of residual variance; and, therefore, a slight underestimation of the variance (and standard deviation) of population correlations.

Analysis 2

The results for Analysis 2 are shown in Table 2. We first note that the mean and standard deviation estimates for the Schmidt–Hunter linear range-restriction-correction procedure, the Callender–Osburn multiplicative procedure, and the SA procedures are identical to the values in Raju and Burke's (1983) study, confirming the accuracy of the simulation program. In this data set, there was no variation across studies in any meta-analysis in criterion reliabilities, predictor reliabilities, or range-restriction values. Because of this, the Schmidt–Hunter interactive and noninteractive procedures yielded identical results, and so separate columns are not shown for these two procedures. The same is true for TSA1 and TSA2. Table 2 shows that the nonlinear range-restriction-correction procedure was more accurate than the original linear range-restriction-correction procedure. The average error in estimating the mean population correlation was .0196 (4% error) for the linear procedure, but only .0013 (.3% error) for the nonlinear procedure; .0013 is within rounding error to the third decimal place. In estimating the standard deviation of the mean population correlation, SD_ρ, the average errors were .0333 (11.5% error) and .0029 (1% error) for the linear and nonlinear methods, respectively. Thus, even in the presence of a gross violation of the normality assumption (a rectangular distribution of population unattenuated ρs), the nonlinear range correction was more accurate than the traditional linear procedure.

In comparison with the Callender–Osburn and Raju–Burke TSA procedures, the nonlinear procedure was more accurate in estimating the mean correlation (0.3% error vs. 4.0% error for both the other procedures). For estimating the true standard deviation of population correlations, the nonlinear range-restriction procedure was most accurate, followed by the TSA models, which in turn were more accurate than the remaining two procedures. Error in the mean estimate of true standard deviation for the nonlinear procedure averaged .0029 (a 1.0% error) in comparison with .0082 (2.8% error) for the TSA models and 11.5% error for the remaining two procedures.

The results in Table 2 can be used to illustrate a point that should be borne in mind generally: The different meta-analysis procedures yield somewhat different figures primarily because they differ in how they handle the range-restriction artifact. In meta-analyses in which range restriction is not a factor (i.e., none of the studies in the meta-analysis suffers from range restriction), all procedures yield results identical to within rounding error. In the nine cases in Table 2 in which there was no range restriction (i.e., $u = 1.00$), all procedures yielded the same estimate of the mean population correlation: .4950, the correct value. Furthermore, all procedures except those with the nonlinear range adjustment yielded the same estimate of the standard deviation of the population correlation: .2886, the correct value. The nonlinear range adjustment causes the estimate to be .2860, which is low by .0026, a 0.9% error.

Conversely, in Analysis 2 one can see that when there is range restriction, all procedures produce some degree of error in estimating both the mean and the standard deviation of population correlations. These errors are greatest when range restriction is most severe ($u = .4110$) and the reliability of the independent and dependent variable measures is high. The most extreme example of this can be seen in the third row of Table 2, where $r_{xx} = r_{yy} = 1.00$ and $u = .4110$. Under these circumstances, all procedures except the Schmidt–Hunter nonlinear procedure substantially overestimated the mean correlation, and all procedures overestimated the standard deviation, although the TSA procedures and the Schmidt–Hunter nonlinear procedure were less inaccurate than the other two procedures. Because Analysis 2 is characterized by population correlations from .01 to .99 (all with equal frequency), the problems created by the nonlinearity of the

Table 2
Analysis 2: Estimated Mean and Standard Deviation of True Population Correlations With Wide and Rectangular Distribution of Population Rhos and Fixed Reliability and Range-Restriction Values

			M_ρ estimate[a,b]				SD_ρ estimate[b]			
r_{xx}	r_{yy}	u	S-H[c] linear	S-H[c] nonlinear	C-O mult.	TSA1 & TSA2	S-H[c] linear	S-H[c] nonlinear	C-O mult.	TSA1 & TSA2
1.00	1.00	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
1.00	1.00	.6030	.5525	.4996	.5525	.5525	.3890	.3085	.3890	.3134
1.00	1.00	.4110	.5881	.4991	.5881	.5881	.4710	.3471	.4710	.3356
1.00	0.60	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
1.00	0.60	.6030	.5237	.4971	.5237	.5237	.3348	.2902	.3348	.2997
1.00	0.60	.4110	.5355	.4983	.5355	.5355	.3558	.2961	.3558	.3050
1.00	0.30	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
1.00	0.30	.6030	.5079	.4957	.5079	.5079	.3085	.2857	.3085	.2934
1.00	0.30	.4110	.5124	.4960	.5124	.5124	.3158	.2865	.3158	.2951
0.80	1.00	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.80	1.00	.6030	.5366	.4984	.5366	.5366	.3579	.2970	.3579	.3055
0.80	1.00	.4110	.5567	.4999	.5567	.5567	.3976	.3122	.3976	.3156
0.80	0.60	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.80	0.60	.6030	.5170	.4965	.5170	.5170	.3233	.2878	.3233	.2970
0.80	0.60	.4110	.5253	.4973	.5253	.5253	.3376	.2909	.3376	.3003
0.80	0.30	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.80	0.30	.6030	.5051	.4955	.5051	.5051	.3041	.2853	.3041	.2924
0.80	0.30	.4110	.5085	.4957	.5085	.5085	.3095	.2858	.3095	.2936
0.50	1.00	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.50	1.00	.6030	.5180	.4966	.5180	.5180	.3251	.2879	.3251	.2973
0.50	1.00	.4110	.5269	.4974	.5269	.5269	.3403	.2916	.3403	.3012
0.50	0.60	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.50	0.60	.6030	.5079	.4957	.5079	.5079	.3085	.2857	.3085	.2934
0.50	0.60	.4110	.5124	.4960	.5124	.5124	.3158	.2865	.3158	.2951
0.50	0.30	1.0000	.4950	.4950	.4950	.4950	.2886	.2860	.2886	.2886
0.50	0.30	.6030	.5012	.4952	.5012	.5012	.2980	.2853	.2980	.2909
0.50	0.30	.4110	.5032	.4953	.5032	.5032	.3012	.2853	.3012	.2915
M			.5146	.4963	.5146	.5146	.3219	.2915	.3219	.2968
Mean error			.0196	−.0013	.0196	.0196	.0333	.0029	.0333	.0082
% error			4.0%	−0.26%	4.0%	4.0%	11.5%	1.0%	11.5%	2.8%

Note. Actual M_ρ = .4950; actual SD_ρ = .2886. S-H linear and nonlinear = Schmidt, Gast-Rosenberg, & Hunter (1980) linear and nonlinear procedure, respectively; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983).

[a] M_ρ estimates of the multiplicative model, the two SA procedures, and the two Schmidt–Hunter linear procedures are all identical. [b] In these data, two TSA procedures gave the same mean and standard deviation estimates. [c] In these data, the Schmidt–Hunter interactive and noninteractive procedures yielded identical estimates.

range-restriction correction are at their maximum. It can be seen in Table 2 that, in every case, addition of the nonlinear range-correction procedure to the Schmidt–Hunter procedure reduces the inaccuracies caused by this problem.

Analysis 3

The results for Analysis 3, the dichotomous case, are shown in Table 3. Estimates of the mean correlation were identical for the two Schmidt–Hunter methods used with the linear range correction. For both the noninteractive and interactive procedures, introduction of nonlinear range correction improved estimates of the mean correlation marginally in the first two dichotomies, despite the severe violation of the normality assumption. In the third dichotomy (ρ_1 = .30; ρ_2 = .20), accuracy declined for the interactive procedure, but only by a trivial .0003. Thus, this gross violation of normality causes no problems in estimating the mean correlation. In all three dichotomies, all procedures provided fairly accurate estimates of the mean correlation. In the first dichotomy (.50 vs. .00), the Schmidt–Hunter noninteractive procedure with nonlinear range correction yielded the most accurate estimate (0.5% error), and the TSA1 procedure yielded the least accurate estimate (2.2% error). In the second dichotomy, the same procedure was again most accurate (0.1% error), and TSA1 was again least accurate (3.1% error). In the third dichotomy, the accuracy was equal for the noninteractive nonlinear procedure and for the interactive and noninteractive linear procedures (0.04% error). The interactive nonlinear procedure had a larger, but still tiny, error (0.2%). The other procedures had larger errors. However, with the exception of the RBNL91 method in the first dichotomy and TSA1 in the second and third dichotomies, all procedures yielded estimates that when rounded two places would be the correct value of .25.

We now turn to the estimates of the true standard deviation of population true score correlations (bottom half of Table 3).

Table 3

Analysis 3: Estimated Mean and Standard Deviation of True Population Correlations With Dichotomous Distributions of Population Rhos

	Meta-analysis procedure							
Actual value in dichotomy	S-H int. linear[a]	S-H nonint. linear[a]	S-H int. nonlinear	S-H nonint. nonlinear	C-O mult.	TSA1	TSA2	RBNL91
M_ρ estimates (actual M_ρ = .2500)[b]								
$\rho_1 = .50, \rho_2 = .00$	.2549	.2549	.2482	.2488	.2544	.2455	.2549	.2443
$S_\rho^2 = .06250$	**.0049**	**.0049**	**−.0018**	**−.0012**	**.0044**	**−.0055**	**.0049**	**−.0057**
% error	2.0	2.0	−0.7	−0.5	1.8	−2.2	2.0	−2.3
$\rho_1 = .40, \rho_2 = .10$	.2515	.2515	.2490	.2497	.2510	.2423	.2515	.2467
$S_\rho^2 = .02250$	**.0015**	**.0015**	**−.0010**	**−.0003**	**.0010**	**−.0077**	**.0015**	**−.0033**
% error	0.6	0.6	−0.4	−0.1	0.4	−3.1	0.6	−1.3
$\rho_1 = .30, \rho_2 = .20$	.2499	.2499	.2496	.2499	.2510	.2423	.2515	.2481
$S_\rho^2 = .02500$	**−.0001**	**−.0001**	**−.0004**	**−.0001**	**.0010**	**−.0077**	**.0015**	**−.0019**
% error	−0.04	−0.04	−0.2	−0.04	0.4	−3.1	0.6	−0.7
SD_ρ estimates (actual SD_ρ varies)[c]								
$\rho_1 = .50, \rho_2 = .00$	.27885	.26454	.26556	.25249	.25324	.26603	.27528	.28357
$SD_\rho = .25000$	**.02886**	**.01454**	**.01556**	**.00249**	**.00324**	**.01603**	**.02528**	**.03357**
% error	11.5	5.8	6.2	1.0	1.3	6.4	10.1	13.4
$\rho_1 = .40, \rho_2 = .10$	.16670	.14216	.16115	.13773	.15186	.16084	.16559	.19305
$SD_\rho = .15000$	**.01670**	**−.00784**	**.01115**	**−.01227**	**.01855**	**.01084**	**.01559**	**.04305**
% error	11.1	−5.2	7.4	−8.2	1.2	7.2	10.4	28.7
$\rho_1 = .30, \rho_2 = .20$	.05577	.00000	.05441	.00000	.05276	.06058	.05925	.12456
$SD_\rho = .05000$	**.00577**	**−.05000**	**.00441**	**−.05000**	**.00276**	**.01058**	**.00925**	**.07456**
% error	11.5	100	8.8	100	5.5	21.1	18.5	149.1

Note. S-H int. = Schmidt, Gast-Rosenberg, & Hunter (1980) interactive procedure; S-H nonint. = Schmidt, Hunter, Pearlman, & Shane (1979) noninteractive procedure; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983); RBNL91 = Raju, Burke, Normand, & Langlois (1991) procedure.

[a] The Schmidt–Hunter interactive and noninteractive procedures yielded the same M_ρ estimates when the linear range-restriction correction was used. [b] Values in boldface are the errors in estimating M_ρ. [c] Actual values of SD_ρ are given in the left column. Errors in estimating SD_ρ are given in boldface; negative values indicate underestimates.

For the interactive procedure, introduction of the nonlinear range correction increased accuracy in all three dichotomies. In the first dichotomy, error decreased from .02886 (11.5% error) to .01556 (6.2% error). In the second dichotomy, error decreased from .01670 (11.1% error) to .01115 (7.4% error). In the third dichotomy, error decreased from .00577 (11.5% error) to .00441 (8.8% error). The percentage reductions in error resulting from the nonlinear range correction in the three dichotomies were, respectively, 46.1%, 33.2%, and 23.5%. Thus, on a percentage basis, the nonlinear range-correction procedure substantially increased accuracy for the interactive procedure, despite the presence of a gross violation of the normality assumption. The same was true for the noninteractive procedure only for the first dichotomy, where error fell from .01454 (5.8% error) to .00249 (1% error). In the second dichotomy, the noninteractive procedure underestimated standard deviation more with the nonlinear range correction (error of −.01227; 8.2% error) than with the linear procedure (error of −.00784; 5.2% error). Thus the nonlinear range correction decreased accuracy slightly. In the third dichotomy, the noninteractive procedure estimated standard deviation at .0000 for both range-correction procedures (error of −.0500; 100% error).

In the first dichotomy, the most accurate estimates of the standard deviation of population true score correlations were given by the noninteractive procedure with nonlinear range correction (error of .00249; 1% error) and by the Callender–Osburn multiplicative procedure (error of .00324; 1.3% error). The least accurate was the RBNL91 procedure (error of .03357; 13.4% error). In the second dichotomy, the most accurate procedure was the multiplicative procedure (error of .01855; 1.2% error), and the least accurate was again the RBNL91 method (error of .04305; 28.7% error). In the third dichotomy, the multiplicative procedure was again the most accurate (error of .00276; 5.5% error), and the RBNL91 method was again the least accurate, with the error being quite large: .07456, a 149% error. The noninteractive procedure yielded estimates of standard deviations of population true score correlations of 0 in both linear and nonlinear form (true value = .0500), for a 100% error. The tendency of the noninteractive procedure to provide underestimates of the variance of population correlations was discussed earlier. In four of the six cases in Table 3 (bottom half), the noninteractive procedure yielded underestimates of standard deviation; it was the only procedure that did so.

In general, Table 3 shows that the Callender–Osburn multiplicative procedure provides the most accurate estimates of standard deviation when the population true score correlation is dichotomous; that is, when there is a single, two-valued moderator. At first glance, this would appear to indicate a decision rule that one should use the multiplicative procedure with such data sets. However, there is a problem. If one knows that there is a dichotomous moderator, then the overall meta-analysis is inappropriate; one should instead conduct two separate meta-analyses, one for each value of the moderator. For that purpose, the multiplicative procedure is not necessarily more accurate than other procedures. The other possibility is one in which the researcher is not aware of the dichotomous moderator (which would usually not be visible in the observed correlations because of the obscuring effects of sampling error and other artifacts). Being unaware of the existence of the dichotomous moderator, the researcher could not apply a decision rule to use the multiplicative procedure in such cases. This leaves only the unusual case in which the researcher strongly suspects that there is a dichotomous moderator but the studies do not provide the information needed to classify the studies into the two moderator categories. In such a case, the multiplicative procedure would be the procedure of choice. But a case such as this would be rare. Also, the resulting overall meta-analysis would not be very informative scientifically. What is needed is two separate meta-analyses—one for each level of the dichotomous moderator—so that the effect of the moderator can be revealed.

Analysis 4

Table 4 shows the results obtained when the normality assumption is violated because of a skewed distribution of population true score correlations. Despite the skewedness, the nonlinear range-restriction correction reduces error in the estimation of the mean correlation for the interactive procedure from .00190 (0.48% error) to −.0001 (0.03% error). The interactive method with the nonlinear range correction yielded the most accurate estimate of the mean population correlation. For the noninteractive procedure, the reduction in error was smaller: from .00190 (0.48% error) to .00095 (0.24% error). All procedures again estimated the mean population correlation fairly accurately, with the largest error being only 0.55% (for Raju et al.'s [1991] method).

Turning to the estimates on the standard deviation of population correlations (bottom part of Table 4), one can see that for the interactive procedure, the nonlinear range-correction procedure again increased accuracy for the skewed population correlation distribution. Error decreased from .00930 (9.3% error) to .00257 (2.6% error). This was not true for the noninteractive procedure; error for that procedure actually increased slightly, as was the case in some earlier analyses. By considering all eight procedures, one can see that the interactive procedure with the nonlinear range correction yielded the most accurate estimate of the standard deviation of the true correlation; the error was only .00257, a 2.6% error. The least accurate estimates of standard deviation were produced by the RBNL91 method, the two TSA methods, and the noninteractive nonlinear procedure. Under conditions of a skewed population correlation distribution, these procedures are apparently substantially inaccurate (errors of 30% or more).

The major purpose of Analysis 4 was to determine whether skewedness in the distribution of population true score correlations would prevent the nonlinear range-correction procedure (with its assumption of normality) from improving accuracy over the older, linear range-correction procedure. These results indicate that for the interactive procedure skewedness does not have this effect. As in Analyses 2 and 3, the correction increases accuracy despite the violation of the normality assumption.

Analysis 5

Analysis 1 showed that with a wide and approximately normal distribution of population true correlations, the nonlinear range-correction procedure provides an improvement in accuracy for the interactive procedure. The results for Analysis 5, with the narrower, approximately normal distribution of population correlations are shown in Table 5. Error in estimating the

Table 4
Analysis 4: Estimated Mean and Standard Deviation of True Population Correlation With Skewed Distribution of Population Correlations

	Meta-analysis procedure							
Estimate	S-H int. linear[a]	S-H nonint. linear[a]	S-H int. nonlinear	S-H nonint. nonlinear	C-O mult.	TSA1	TSA2	RBNL91
M_ρ	.40190	.40190	.39990	.40095	.40140	.39821	.40190	.39781
Error	.00190	.00190	−.00010	.00095	.00140	−.00179	.00190	−.00219
% error	0.48	0.48	−0.03	0.24	0.35	−0.45	0.47	−0.55
SD_ρ	.10930	.07440	.10257	.06995	.10892	.10318	.10322	.13388
Error	.00930	−.02560	.00257	−.03005	.00892	.00318	.00322	.03388
% error	9.3	−25.6	2.6	−30.1	8.9	31.8	32.2	33.9

Note. Actual M_ρ = .40; actual SD_ρ = .10000. S-H int. = Schmidt, Gast-Rosenberg, & Hunter (1980) interactive procedure; S-H nonint. = Schmidt, Hunter, Pearlman, & Shane (1979) noninteractive procedure; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983); RBNL91 = Raju, Burke, Normand, & Langlois (1991) procedure.

[a] The Schmidt–Hunter interactive and noninteractive procedures yielded the same M_ρ estimate when the linear range-restriction correction was used.

Table 5

Analysis 5: Estimated Mean and Variance of Population Correlations With Narrower Dispersion of Population Rhos and With Variable Reliability and Range-Restriction Values

	Meta-analysis procedure							
Estimate	S-H int. linear[a]	S-H nonint. linear[a]	S-H int. nonlinear	S-H nonint. nonlinear	C-O mult.	TSA1	TSA2	RBNL91
			M_ρ estimates (actual M_ρ = .5000)					
M_ρ	.5006	.5006	.4998	.5006	.4996	.4960	.5006	.4965
Error	.0006	.0006	−.0002	.0006	−.0004	−.0040	.0006	−.0035
% error	.12%	.12%	−.04%	.12%	−.08%	−.80%	.12%	−.7%
			SD_ρ estimates (actual SD_ρ = .05477)					
SD_ρ	.06008	.00000	.05495	.00000	.06935	.05431	.05206	.12016
Error	.00531	−.05477	.00018	−.05477	.01458	−.00046	−.00271	.06539
% error	9.7%	−100%	.32%	100%	26.7%	−.84%	−5.0%	119.4%

Note. S-H int. = Schmidt, Gast-Rosenberg, & Hunter (1980) interactive procedure; S-H nonint. = Schmidt, Hunter, Pearlman, & Shane (1979) noninteractive procedure; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983); RBNL91 = Raju, Burke, Normand, & Langlois (1991) procedure.

[a] The Schmidt–Hunter interactive and noninteractive procedures yielded the same estimates of mean true population correlations when linear range-restriction correction was used.

mean correlation was decreased for the interactive procedure from 0.12% (.0006) to 0.04% (−.0002) by the nonlinear range-correction procedure. However, for the noninteractive procedure, there was no decrease in error; it remained constant at 0.12% (.0006). The most accurate estimates of the mean population correlation were produced by the interactive procedure with the nonlinear range correction (error = −.0002, a 0.04% error), followed by the multiplicative procedure (error = −.0004, a 0.08% error). All estimates were quite accurate. Even the procedure with the largest error (TSA1) yielded the correct value of .50 when rounded to the usual two decimal places.

Turning to estimates of the standard deviation of true score population correlations, we see that for the interactive procedure, introduction of the nonlinear range-correction procedure improved accuracy; the error of estimation declined from .00531 (a 9.7% error) to .00018 (a 0.32% error), for a 97% reduction in error. The noninteractive procedure, on the other hand, estimated this small standard deviation (SD_ρ = .05477) at zero for both range-correction procedures (error of .05477; a 100% error). The most accurate estimate of the standard deviation was yielded by the interactive procedure with the nonlinear range correction (0.32% error), followed by the TSA1 procedure, which underestimated standard deviation by −.00046, a .84% error. For this distribution of population correlations, the multiplicative procedure yielded a noticeably less accurate estimate of the standard deviation, with an error of .01458 (a 26.7% error). However, the least accurate procedure was the RBNL91 method, with an error of .06539 (a 119% error).

Analysis 5 (Table 5) is probably more realistic than Analysis 1, and, as a result, should be given more weight. Also, considering errors of the same absolute size, errors in estimating the population standard deviation are more serious than are errors in estimating the population mean. For example, in Analysis 5, an error of .01458 in estimating the mean is a 2.9% error; however, an error of .01458 in estimating the standard deviation is a 26.7% error. Therefore, the best overall interpretation of our results is that, for real data in which the standard deviation is greater than zero and is modest in size and in which population correlations are approximately normal, the most accurate procedures are likely to be the interactive with nonlinear range correction and the two TSA procedures. In the cases examined in which the distribution of population correlations was approximately normal (Analyses 1 and 5), incorporating the nonlinear range-correction procedure into the interactive procedure increased its accuracy in estimating standard deviation, as predicted by Hunter and Schmidt (1990b, pp. 209–211).

Analysis 6

The results for the homogeneous case (SD_ρ = 0) are shown in Table 6. In this situation, the assumption that the population correlations are normally distributed is violated; the distribution is a point distribution at ρ = .50. For the noninteractive procedure, the nonlinear range correction produced no changes at all in any of the cells (domains) in Table 6, whether in estimating mean or standard deviation of population true correlations. Thus, the normality violation did not cause the nonlinear correction to lead to decreased accuracy. For the interactive procedure, introduction of the nonlinear correction led to small increases in accuracy of estimation of standard deviation of the true score population correlation, despite violation of the normality assumption. Estimates of the mean correlation were made slightly less accurate by introduction of the nonlinear correction in five of the eight domains. However, this decrease in accuracy was trivial, never exceeding 0.027%. Thus we conclude that, despite the violation of the normality assumption, the nonlinear range correction does not encounter accuracy problems in the homogeneous case.

Table 6
Analysis 6: Results for the Homogeneous Case

Meta-analysis procedure									
	Var r_{xx}:	Low	High	Low	Low	High	High	Low	High
	Var r_{yy}:	Low	Low	High	Low	High	Low	High	High
	Var u:	Low	Low	Low	High	Low	High	High	HIgh
% error in estimates of M_ρ (actual M_ρ = .5000 in all cases)									
S-H int. linear		−0.071	0.011	0.578	−0.967	0.670	−0.901	−0.441	−0.367
S-H nonint. linear		−0.071	0.011	0.578	−0.967	0.670	−0.901	−0.441	−0.367
S-H int. nonlinear		−0.072	0.011	0.578	−0.994	0.670	−0.927	−0.460	−0.383
S-H nonint. nonlinear		−0.071	0.011	0.578	−0.967	0.670	−0.901	−0.441	−0.367
C-O mult.		−0.276	−0.193	0.381	−1.881	0.474	−1.810	−1.317	−2.239
TSA1		−1.161	−1.476	−3.963	−2.047	−4.260	−2.374	−4.936	−5.246
TSA2		−0.071	0.011	0.578	−0.967	0.670	−0.901	−0.441	−0.367
RBNL91		−0.613	−0.619	−0.661	−2.408	−0.666	−2.429	−2.592	−2.610
% error in estimates of SD_ρ (actual SD_ρ = 0.0 in all cases)									
S-H int. linear		.005	.000	.000	.027	.000	.027	.023	.022
S-H nonint. linear		.000	.000	.000	.000	.000	.000	.000	.000
S-H int. nonlinear		.004	.000	.000	.025	.000	.025	.022	.020
S-H nonint. nonlinear		.000	.000	.000	.000	.000	.000	.000	.000
C-O mult.		.041	.044	.055	.072	.058	.073	.072	.073
TSA1		.017	.021	.066	.027	.069	.030	.073	.075
TSA2		.007	.012	.000	.022	.014	.024	.029	.035
RBNL91		.113	.121	.176	.185	.182	.190	.226	.231

Note. S-H int. = Schmidt, Gast-Rosenberg, & Hunter (1980) interactive procedure; S-H nonint. = Schmidt, Hunter, Pearlman, & Shane (1979) noninteractive procedure; C-O mult. = Callender & Osburn (1980) multiplicative procedure; TSA1 and TSA2 = Taylor Series Approximations 1 and 2 (Raju & Burke, 1983); RBNL91 = Raju, Burke, Normand, & Langlois (1991) procedure; Var = variance.

The top part of Table 6 shows the percentage errors in estimating the mean true correlation of .50. For five of the eight methods shown, the error is always below 1%; these are the two interactive methods, the two noninteractive methods, and the TSA2 method. For these methods, the mean true correlation estimates were identical to four digits in all eight domains. The multiplicative method was a bit less accurate in the cells (domains) having high variance of range-restriction values, but its error was over 2% for only one of the eight combinations.

The method that showed the largest errors for the high-artifact-variance domains was TSA1. The (corrected) TSA2 method was consistently more accurate than TSA1 and was much more accurate for the high-artifact-variance domains. This was equally true for all of the traditional methods; all were noticeably superior to TSA1.

In the article introducing their new method (RBNL91), Raju et al. (1991) compared their new method only to TSA1. They concluded that the new method was superior to TSA1 in most of the domains that they considered. The top part of Table 6 does, in fact, show the RBNL method to be superior to TSA1 in estimating the mean correlation in six of eight of the homogeneous domains. In the low-low-high and high-low-high domains, the method was just as inaccurate as TSA1 (trivially worse). In the other domains, the RBNL91 method was substantially more accurate than TSA1.

However, the fact that the RBNL91 method is more accurate than TSA1 means little in view of the fact that TSA1 was less accurate in the homogeneous domains than were the other traditional methods. When the RBNL91 method is compared with the TSA2 method, the RBNL91 method is less accurate in seven of eight domains and is essentially equal in the high-high-low domain (though in opposite directions), in which the error for both methods was below 1%. The same is true for all of the traditional methods: All of the traditional methods are at least as accurate as the RBNL91 method in all domains, and all but the multiplicative method are more than twice as accurate in six out of the eight domains. The Callender–Osburn multiplicative estimator is twice as accurate as the RBNL91 method in five of the eight domains.

The bottom half of Table 6 shows the estimated standard deviations for each method applied to each of the eight homogeneous domains. Because the true standard deviation is zero in all eight homogeneous domains, a percentage of error is not defined. Therefore, the error measure that we used is the numerical deviation from the true value of zero. Because the true value of the standard deviation of the population true correlation is zero, the estimated standard deviation for any estimator is also the amount of error for that estimator.

As noted earlier, the noninteractive method is known to slightly overestimate the variance due to (corrected) artifacts if range restriction is present. In the homogeneous case without sampling error, the result is that the estimate of residual variance is always a very small negative number and, hence, the noninteractive method correctly estimates the standard deviation of true correlations to be zero in all eight domains. This holds both with and without the nonlinear range-correction procedure.

Three of the other traditional estimators are extremely accurate in all domains: the two interactive estimators and TSA2. The new range-restriction procedure improves the interactive method in all domains in which there is any room for improvement. The interactive nonlinear estimator is more accurate than TSA2 in five of the eight domains, although the differences are very small.

The multiplicative model consistently underestimates the variation due to range restriction. The domains in our study were designed to have the high level of range restriction found empirically for general cognitive ability tests used to predict job performance. Thus, the mean in each domain is $u = .67$. This makes the multiplicative model the least accurate, on average, of the traditional estimators for these domains.

The TSA1 estimator is no more accurate than the multiplicative estimator in four of the eight domains. In the other four domains it noticeably improves on the multiplicative estimator. On the other hand, TSA1 is noticeably poorer than all of the other traditional estimators, including TSA2 and the interactive estimator.

The RBNL91 method is considerably less accurate than any of the other estimators. It is never less than 6 times less accurate than the improved interactive estimator. The errors made by the RBNL91 estimator are large enough to lead to considerable inaccuracy in some domains. Consider the high-high-high case in Table 6. Whereas the true correlations are actually uniformly .50, the RBNL91 estimator would generate an 80% credibility interval from .20 to .80. The 95% credibility interval would be from .03 to .90. That is, the RBNL91 method would incorrectly suggest that the correlation varies from essentially 0 to nearly 1.00. (The applications of the RBNL91 method considered in this study all relied solely on distributions of artifacts. In the future, we intend to examine the accuracy of this method when artifact information is known individually for some of the correlations.)

By comparison, the improved interactive estimator produces a 95% credibility interval from .46 to .54, an interval width that is almost 12 times narrower. In this domain, that means it is almost 12 times more accurate than the RBNL91 estimator. Furthermore, there is no research domain in which .54 is considered qualitatively different from .46. But in every research domain, .90 would be regarded as qualitatively different from .03.

In estimating the mean correlation, TSA1, with errors as high as 5%, was noticeably less accurate than any other estimator. The RBNL91 method was about twice as accurate as TSA1 but was still noticeably less accurate than the other, traditional methods. The multiplicative estimator was about twice as accurate as the RBNL91 estimator, though still slightly poorer than the other, traditional estimators. The two interactive estimators, the two noninteractive estimators, and TSA2 were all accurate to within less than 1% for estimating the mean correlation in all eight homogeneous domains studied here.

In estimating the true standard deviation, the RBNL91 method was much less accurate than any other estimator; in some cases, it was less accurate by a factor of 10 or more. The RBNL91 method generated qualitatively unacceptable errors in estimating the credibility interval for domains with high artifact variation. The multiplicative estimator did more poorly than the other traditional estimators in these domains, as would be predicted from the high mean level of range restriction in these domains. TSA1 was almost as inaccurate as the multiplicative estimator and was noticeably poorer than TSA2 in all domains. The interactive and noninteractive estimators were even more accurate than TSA2, although the difference was not large enough to be important in the eight homogeneous domains studied here.

A technical comment appears appropriate at this point. In Tables 3, 5, and 6, some of the estimates of standard deviations for population true correlations for the noninteractive procedure were zero (as are several in Table 6 for the interactive procedure). These typically result from small negative estimates of residual variance, which are set to zero before the calculation of standard deviation. One reviewer was concerned that these estimates of the accuracy of the procedures might be biased; however, this is not the case. It is important to remember that the purpose of this study was to evaluate the accuracy of these procedures. An essential property of these methods is that the residual variance estimate is set to zero when the observed estimate is negative. This is similar to the case in Cronbach's generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; see discussion in Hunter & Schmidt, 1990b, p. 413). Thus the zero values of the standard deviation estimates in Tables 3, 5, and 6 are not biased estimates of the accuracy of the procedures in question. They are the actual values produced by those methods. For cases in which $SD_\rho = 0$, the estimates of the standard deviation that are zero represent perfect accuracy; however, when $SD_\rho > 0$, such estimates are in error (to one degree or another). Thus, the zero values for standard deviations in Table 6 represent perfect accuracy, whereas the zero values in Tables 3 and 5 (all from the noninteractive method) represent errors of estimation.

Conclusions

In summary, the findings of Analyses 2, 3, and 6 indicate that violations of the normality assumption underlying the nonlinear range-correction procedure do not cause that procedure to be inaccurate in estimating mean population true correlations or their standard deviations. Even under conditions of gross violation of normality, the nonlinear range-correction procedure still produced increased accuracy for the interactive procedure (and often for the noninteractive procedure as well). In Analysis 1, in which the normality assumption was approximately met and the variance of population correlations was large, the new nonlinear range-correction procedure yielded definite improvements over the older nonlinear procedure for the interactive method in accuracy of estimates of both mean and standard deviation. In Analysis 5, in which the normality assumption was also approximately met but in which the population correlations were much less variable (and hence more realistic), the new nonlinear range-restriction correction again yielded improvements in accuracy in estimating mean and standard deviation for the interactive procedure. Thus, we conclude that these findings indicate that the nonlinear range-correction procedure generally yields an improvement in accuracy for the interactive method, as has been predicted by Hunter & Schmidt (1990b, pp. 209–211).

Accuracy can also be compared for all existing procedures. The results of Analysis 1 indicate that when the variance of population correlations is large and the distribution is approximately normal, the interactive procedure with the nonlinear range correlation is the most accurate; however, the two TSA procedures have very similar levels of accuracy. The results of Analysis 5 indicate that when population correlations are approximately normal but less variable, the most accurate procedures are again apt to be the interactive with nonlinear range correction and the two TSA procedures of Raju and Burke (1983). However, the results of Analysis 4 indicate that when the distribution is skewed, the two TSA methods become considerably less accurate for estimating standard deviation, with percentage of error figures of 30% or more. The same is true for the RBNL91 method and the noninteractive method. Under these circumstances, the interactive procedure with the nonlinear correction showed only 2.6% error in estimating standard deviations.

The results of Analysis 6 indicate that in the homogeneous case (in which $SD_{\rho} = 0$), the interactive procedure with the nonlinear range correction and TSA2 are the most accurate procedures. The TSA1 procedure is considerably less accurate in the homogeneous case. This finding is important because there has been a lot of evidence supporting the homogeneous model for aptitude and ability tests in employment selection (Schmidt et al., 1985, 1993). In the homogeneous case, the RBNL91 method yielded the least accurate estimates of the standard deviation of the population true correlation.

Considering the results of all six analyses, we conclude that the most accurate procedure overall appears to be the interactive procedure with the nonlinear correction. However, in most analyses, the TSA2 procedure was similar in accuracy; the major exception is the case in which the population correlations have a skewed distribution. It is difficult to estimate how frequently this occurs in real data (i.e., sets of studies used in a meta-analysis), and therefore, it is difficult to know how much of a problem skewedness is for the TSA2 procedure.

In all cases, errors in estimating the mean population true score correlation are small for all procedures, both on an absolute and on a percentage basis. Errors in estimating the standard deviation of population true score correlations are usually small in absolute value for all procedures but can be large when expressed as percentage errors of estimation for some procedures under some circumstances. Most procedures that we examined are quite accurate by the usual standards of social science, and most are accurate enough for research use with real data in most applications. Nevertheless, the quest for improved accuracy is part of the scientific enterprise. Scientists—including social scientists—should be concerned about accuracy and, other things being equal, should choose the more accurate procedures even when the difference is not large.

References

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65,* 543–558.

Callender, J. C., & Osburn, H. G. (1988). Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology, 73,* 312–315.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Hunter, J. E., & Schmidt, F. L. (1990a). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75,* 334–349.

Hunter, J. E., & Schmidt, F. L. (1990b). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Law, K. S. (1992). Estimation accuracy of Thomas's likelihood-based procedure of meta-analysis: A Monte Carlo simulation. *Journal of Applied Psychology, 77,* 986–995.

Mendoza, J. E., & Reinhardt, R. N. (1991). Validity generalization procedures using sample-based estimates: A comparison of six procedures. *Psychological Bulletin, 110,* 596–610.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68,* 382–395.

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76,* 432–446.

Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65,* 529–540.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529–540.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsch, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology, 38,* 697–798.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt–Hunter Bayesian validity generalization procedure. *Personnel Psychology, 32,* 257–381.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78,* 3–12.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

Thomas, H. (1990). A likelihood-based model for validity generalization. *Journal of Applied Psychology, 76,* 13–20.

(*Appendix follows on next page*)

Appendix

Mathematics of the Nonlinear Range-Correction Refinement

Uniform Artifacts

Suppose that each artifact is uniform across studies; that is, the level of range restriction is the same across all studies and the reliability of the independent and dependent variables is the same across all studies. If each artifact were uniform, then all study population correlations would be attenuated by exactly the same formula $A(\rho)$, where A is a function of ρ. That is, each attenuated study population correlation (ρ_o) would be related to the true study population correlation (ρ) by the same attenuation function *a:*

$$\rho_o = A(\rho). \tag{A1}$$

Each attenuated study population correlation could be corrected by using the same correction function. That is, if C is the inverse function for A, then

$$\rho = C(\rho_o) = A^{-1}[A(\rho)]. \tag{A2}$$

The moments of the true and attenuated study correlations would be related by integral equations (assuming a continuous distribution of ρ; otherwise by the corresponding sums). If the frequency density function for true study population correlations is $f(\rho)$, then the

$$n\text{th moment of } \rho_o = \int [A(\rho)]^n f(\rho) d\rho. \tag{A3}$$

If the frequency density function for the observed attenuated study population correlations is $f(\rho_o)$, then the

$$n\text{th moment of } \rho = \int [C(\rho_o)]^n f(\rho_o) d\rho_o. \tag{A4}$$

In particular, it is this last equation that would be used to transform the distribution of imperfect study population correlations to the distribution of study population correlations for perfectly done studies. If a bare-bones meta-analysis were done in a domain with uniform artifacts, then the mean and standard deviation from the meta-analysis would be estimates of the mean and standard deviation of attenuated correlations, that is, the mean and standard deviation of ρ_o. If the form of the distribution of ρ_o were known, then this last equation could be used to compute the mean and standard deviation of ρ.

In our refinement, we use the normal distribution as an approximation for the distribution of ρ_o. One source of error for the models using the refinement is the fact that not one of the distributions for this study is actually normal and several are severely nonnormal. The results in our study indicate that little error is introduced by the assumption of a normal distribution.

Consider the computation of the true study population correlations if all studies were subject to range restriction with constant level u. Then the correction formula for all study population correlations would be

$$C(\rho_o) = \frac{U\rho_o}{\sqrt{(U^2 - 1)\rho_o^2 + 1}}, \tag{A5}$$

where $U = 1/u$. Denote the mean and standard deviation of attenuated correlations by μ and σ. Denote the normal density function for mean μ and standard deviation σ by f. Then the mean true study population correlation would be

$$\text{mean}(\rho) = \int C(\rho_o) f(\rho_o) d\rho_o. \tag{A6}$$

In our computer programs (available on request), we computed this integral equation numerically. To do this, we considered 61 values of the standard normal deviate centered about 0; that is, values of z: -3.00, $-2.90, \ldots, -0.10, 0.00, 0.10, \ldots$, and 3.00. For each value, we computed the corresponding probability: .0004, .0006, . . . , .0397, .0399, .0397, . . . , and .0004. The value for ρ_o is computed for each value of z by the usual reverse-standard-score formula:

$$\rho_o = \sigma z + \mu. \tag{A7}$$

Denote the ith created value for ρ_o by r_i and denote the corresponding probability by p_i. The corresponding ith value for true correlations is thus

$$c_i = C(r_i). \tag{A8}$$

Then the first and second moments of the true correlation are computed in the usual way:

$$\text{mean } \rho = \sum c_i p_i, \tag{A9}$$

$$\text{mean square } \rho = \sum c_i^2 p_i. \tag{A10}$$

Nonuniform Artifacts

In most domains, each artifact varies from study to study. For example, if different studies use different measures of the independent variable, then some studies will have higher independent-variable reliability than other studies. In this case, the true and attenuated correlations are not related by a single attenuation formula, and the integral equations of the previous section do not apply.

The heart of both the interactive and noninteractive methods of meta-analysis is to transform the results of the bare-bones meta-analysis to results for a hypothetical domain with uniform artifact values. That is, the attenuated study population correlations with nonuniform artifacts have a mean and variance that are estimated from the bare-bones meta-analysis. Denote these as mean (ρ_o) and var (ρ_o).

On the other hand, if the artifacts were uniform across the domain, with each artifact value uniformly equal to the mean value for that artifact, then the mean and variance would have had different values; these values were referred to by Schmidt and Hunter (1977) as the *residual mean* and *residual variance.* Denote these values as resmean (ρ_o) and resvar (ρ_o). If these values were known for the domain, then the integral equations of the previous section could be used to compute the mean and variance of the true study correlations.

The interactive and noninteractive models are two different methods of estimating the residual mean and residual variance from the mean and variance produced by the bare-bones meta-analysis. In both methods, the current practice is to use the mean from the meta-analysis as the estimate for the residual mean. Thus, for the mean correlation, the estimated residual mean is actually the mean observed correlation. On the other hand, the residual variance is smaller than the variance produced by the bare-bones meta-analysis by an amount called the *variance due to artifact variation.* It is this residual variance that is used to compute the moments of the true correlations by the formula of the preceding section.

Received March 25, 1991
Revision received September 29, 1993
Accepted September 29, 1993 ■