Journal of Applied Psychology
1994, Vol. 79, No. 6, 978–986

0021-9010/94/$3.00

A Test of Two Refinements in Procedures for Meta-Analysis

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

This study used Monte Carlo simulation to examine the increase in accuracy resulting from 2 statistical refinements of the interactive Schmidt–Hunter procedures for meta-analysis: the use of the mean correlation instead of individual correlations in the estimation of sampling error variance, and a procedure that takes into account the nonlinear nature of the range-restriction correction. In all of the cases examined, these refinements increased the accuracy of the interactive procedure in estimating the variance of population correlations and resulted in more accuracy than other procedures examined. The use of the mean correlation in the sampling error variance formula also increased the accuracy of variance estimates for the multiplicative and Taylor Series procedures.

Meta-analysis has become a widely used research technique in behavioral science. Among the different available meta-analytic approaches, the artifact-distribution-based meta-analytic methods for correlation coefficients are used most often, with numerous applications of this technique in different areas of behavioral science (Hunter & Schmidt, 1990; Schmidt, 1992). Computer simulation studies have examined the accuracy of procedures of this type (Callender & Osburn, 1980; Kemery, Mossholder, & Roth, 1987; Mendosa & Reinhardt, 1991; Raju & Burke, 1983). Most applications of meta-analysis to large real-world databases have yielded positive nonzero estimates of the variance of population correlations (e.g., Pearlman, Schmidt, & Hunter, 1980), although the variance estimates have generally been very small. Simulations (e.g., see Raju & Burke, 1983) have shown that although all of the mean population correlation estimates were very accurate, all of the procedures except the Schmidt–Hunter noninteractive procedure overestimated the true variance of population correlations. As a result, there have been continuous efforts to improve the accuracy of the existing meta-analytic procedures and to explain why they overestimate the true variance of population correlations.

This study focused on two major modifications of the Schmidt–Hunter noninteractive (Pearlman et al., 1980; Schmidt, Hunter, Pearlman, & Shane, 1979) and interactive (Schmidt, Gast-Rosenberg, & Hunter, 1980) procedures for meta-analysis and used computer simulations to test the estimates produced by these improved procedures. We hypothesized that with these improvements overestimation of true variance of population correlations across studies would be reduced and that more accurate estimates of the population parameters could be obtained. The two improvements tested in this study were (a) the use of the mean observed correlation instead of individual correlations in sampling error variance estimation (Hunter & Schmidt, 1990, pp. 208–209) and (b) a procedure that takes into account the nonlinear effect of the range-restriction correction (Hunter & Schmidt, 1990, pp. 209–211).

This article is one in a series of related articles. The first article (Law, Schmidt, & Hunter, 1994) used simulation studies of a variety of validity distributions to evaluate the nonlinear range-correction procedure for estimating SD_ρ. To more clearly address the effects of this refinement, we conducted that study under conditions of no sampling error (all Ns were infinite). The major finding was that the nonlinear refinement increased the accuracy of estimates of SD_ρ for the Schmidt–Hunter interactive procedure, making that procedure generally more accurate than other procedures. We also found that even gross violations of the normality assumption entailed by the nonlinear refinement did not prevent the refinement from increasing accuracy. The next article in this series (Hunter & Schmidt, 1994) examined the question of whether use of $\bar{r}$ instead of r in the formula for the sampling error variance of the (Pearson) correlation coefficient would increase the accuracy of estimates of sampling error variance. This was accomplished analytically, not through computer simulation. The major finding was that use of $\bar{r}$ increased accuracy in all cases except those in which the mean attenuated population correlation (ρ_{xy}) was larger than .60. Because attenuated population correlations are rarely as large as .60 in real data (they are usually in the .00 to .40 range), Hunter and Schmidt (1994) concluded that use of $\bar{r}$ would usually increase the accuracy of meta-analytic results. However, Hunter and Schmidt addressed only the homogeneous case, in which $SD_\rho = 0$. The heterogeneous case, in which SD_ρ is greater than zero, is more complicated to address analytically and is still being studied. The difficulties in analytically evaluating the heterogeneous case can be circumvented to some extent by using computer-simulation studies to evaluate use of $\bar{r}$ in the heterogeneous case, which was one purpose of the present study. The other purpose of this study was to examine the improvement in accuracy of estimates of SD_ρ resulting from the nonlinear range refinement under conditions of finite (and realistic) sample sizes

Kenneth S. Law, Department of Management, Hong Kong University of Science and Technology, Hong Kong; Frank L. Schmidt, Department of Management and Organization, College of Business, University of Iowa; John E. Hunter, Department of Psychology, Michigan State University.

Correspondence concerning this article should be addressed to Frank L. Schmidt, Department of Management and Organization, College of Business, University of Iowa, Iowa City, Iowa 52242. Electronic mail may be sent via Internet to frank.schmidt@uiowa.edu.

(both in individual studies and in meta-analysis as a whole). The two major hypotheses were (a) that even in the heterogeneous case use of $\bar{r}$ would improve accuracy (for all procedures) and (b) that the nonlinear refinement would continue to improve the accuracy of the interactive procedure's estimates of SD_ρ. A final hypothesis was that the interactive procedure used with both the nonlinear refinement and with $\bar{r}$ in the sampling error variance formula would yield estimates of SD_ρ that were more accurate than those from any other procedure.

In the final study in this series, Schmidt et al. (1993) used the interactive procedure to apply the two refinements mentioned above to large sets of real data to determine the practical and theoretical implications in the area of validity generalizability for cognitive-ability selection tests. The general conclusion was that results using the more accurate procedures further reduced support for the hypothesis of situational specificity of validity in personnel selection.

Use of $\bar{r}$ Instead of r in the Sampling Error Variance Formula

The well-known formula for the sampling error variance of sample correlation coefficients is

$$S_e^2 = \frac{(1 - \rho_{xy}^2)^2}{N - 1}, \tag{1}$$

where N is the sample size and ρ_{xy} is the (uncorrected) population correlation. (A variation of this formula has N in the denominator rather than $N - 1$. The formula using $N - 1$ has smaller bias for all except very large sample sizes [e.g., see Hunter & Schmidt, 1994] and is used in all of our meta-analytic programs as well as by most researchers for most purposes.) The value of ρ_{xy} is unknown, and to use this formula, some method must be found to estimate it. In single studies, the estimate of ρ_{xy} that is typically used—because it is the only one available—is the observed correlation in the study at hand. In our early meta-analyses of employment test validities (e.g., Schmidt et al., 1979, 1980), we followed this tradition: The value used to estimate the sampling error variance in every study was the observed correlation in that study. We have since found that this procedure is not optimal. The mean observed correlation ($\bar{r}_{\text{obs}}$)—a good estimate of ρ_{xy}—is typically about .20 in this literature. Sample sizes are usually small, so there are substantial departures in both directions from ρ_{xy}. When the sampling error is large and positive (e.g., .20, so that $r = .40$), the estimated S_e^2 is substantially reduced (by 23% in this example). However, this effect is not symmetrical. When sampling error is large and negative (e.g., $-.20$, so that $r = .00$), the estimated S_e^2 is increased by only a small amount (by 9% in this example). Thus, on balance, the sampling error in a set of correlations is underestimated. The smaller the sample size is in the studies being analyzed, the greater this underestimation will be. Also, the smaller the (attenuated) population correlation is, the greater the underestimation will be because smaller ρ_{xy} values yield larger sampling error variances, sample sizes being equal. The result is underestimation of the amount of variance accounted for by sampling error and overestimation of SD_ρ. This distortion can be eliminated by using the mean correlation for the set of studies rather than individual correlations in the formula for sampling error. The mean correlation contains little sampling error, and extreme values are unlikely. The result should be more accurate estimates of SD_ρ.

By using analytic methods and by examining only the homogeneous case (no variation in population correlations, $SD_\rho = 0$), Hunter and Schmidt (1994) found that, under most conditions, use of $\bar{r}$ yielded more accurate estimates. However, when the attenuated population correlation (the true-score population correlation reduced by unreliability, range restriction, or both) was larger than about .60, the estimate based on the mean correlation had slightly more negative bias (underestimation of actual sampling variance) than the estimate based on the individual correlations. The importance of this exception is limited by the fact that in real data, attenuated population correlations as large as .60 are rare. In the abilities domain of validity generalization, for example, the average value of the attenuated population correlation is about .20 and is rarely above .25. No values of this correlation at or above .60 appeared in the realistic simulation studies included in the present research. Thus, the greater bias of sampling error variance estimates based on $\bar{r}$ in this range was not of any consequence in this study (as would be the case in virtually all real data sets).

In a Monte Carlo study, Millsap (1988) used r rather than $\bar{r}$ in the formula for sampling error variance. In his study, all population correlations were equal, so S_ρ^2 was zero, and the variance of the observed correlations was solely due to sampling error variance (i.e., $S_r^2 = S_e^2$). However, he found that his formula-derived estimates of S_e^2 were slightly smaller than the observed S_r^2 figures, and this discrepancy was larger for smaller sample sizes. He attributed this finding to inaccuracy in the formula, which is an approximation, but the phenomenon described here is, in part, the explanation for his findings. Millsap also found that the negative bias in his formula-derived estimates of sampling error variance was larger when measures had lower reliability. This finding is explained by the fact that lower reliability leads to lower values of ρ_{xy}, the attenuated population correlation. Lower ρ_i values have larger sampling error variances for any fixed sample size, thus intensifying the process described above. Therefore, it was not unreliability (measurement error) per se that caused the increase in Millsap's underestimation but rather the reduced value of the attenuated population correlation.

Nonlinearity in the Range Correction and a New Correction Procedure

In artifact-distribution-based methods of meta-analysis, the mean ($\bar{\rho}$) and standard deviation (SD_ρ) of true correlations are estimated from the mean ($\bar{r}_{\text{res}}$) and standard deviation (SD_{res}) of the residual distribution. The residual distribution is the distribution of observed correlations expected across studies if N were always infinite (i.e., there is no sampling error) and reliability and range restriction were always constant at their respective mean values (Law et al., 1994). To correct the residual distribution for unreliability, we could divide every value in that distribution by the mean of the square roots of the reliabilities. However, because that value is a constant, we can instead simply divide both $\bar{r}_{\text{res}}$ and SD_{res} by that constant and obtain the same result. This is what our artifact-distribution-based procedures for meta-analysis have traditionally done. However, these procedures have done exactly the same thing in correcting the residual distribution for the effects of mean range restriction, and here this approach does not work as accurately. Our traditional procedures use the mean level of range restriction (in the form of the ratio of the restricted predictor standard deviation to the unrestricted predictor standard deviation) to correct $\bar{r}_{\text{res}}$. This procedure increases $\bar{r}_{\text{res}}$ by some factor (e.g., 1.50). Then, SD_{res} is multiplied by this same factor to estimate the standard deviation of a distribution in which each correlation has been corrected for the mean level of range restriction. However, unlike the reliability correction, the range-restriction correction is not linear in r. The correction is not the same for every value of r in the residual distribution; instead, it is larger for smaller correlations and smaller for larger correlations. Thus, the approximation based on the assumption of linearity in artifact-distribution-based procedures for meta-analysis leads to overestimates of SD_ρ. Simulation studies (Callender & Osburn, 1980; Raju & Burke, 1983) have demonstrated that the interactive procedure—theoretically, our most sophisticated method (see Schmidt et al., 1980)—yields estimates of SD_ρ that are too large by .02 or more. This overestimation occurs in simulated data in which sample sizes are infinite (eliminating sampling error) and uncorrected sources of artifactual variance such as computational errors, outliers, and non-Pearson correlations do not exist. This overestimation stems from failure to take into account the nonlinearity of range-restriction corrections. This nonlinearity can be taken into account by separately correcting each value in the residual distribution for the mean level of range restriction. To take this nonlinearity into account, the following method can be used (Law et al., 1994): After determining the mean and the standard deviation of the residual distribution, 60 additional values in that distribution are specified by moving out from the mean in 0.1 SD units to 3 SD units above and below the mean. Then, each of these values is corrected individually for range restriction by using the mean of the S/s ratio. The formula used to correct each value is

$$R_i = \frac{r_i(S/s)}{\{[(S/s)^2 - 1]r_i^2 + 1\}^{1/2}}, \quad (2)$$

where r_i is the value of the correlation in the residual distribution; R_i is the correlation corrected for range restriction; S is the unrestricted standard deviation; and s is the mean restricted standard deviation.

Each range-corrected correlation is then corrected for the mean effect of unreliability; the resulting value is symbolized as $\hat{\rho}_i$. The relative frequency (f_i) of each value of r is indexed by the normal curve ordinate associated with its z score in the residual distribution. (Law et al., 1994, provided both conceptual and empirical justification for the normality assumption.) These frequencies are applied to the corresponding corrected correlations ($\hat{\rho}_i$). The frequency-weighted mean of the distribution of the corrected correlations ($\bar{\rho}$) is then determined, and the following frequency-weighted variance formula is used to find S_ρ^2:

$$S_\rho^2 = \frac{\Sigma f_i(\hat{\rho}_i - \bar{\rho})^2}{\Sigma f_i}, \quad (3)$$

where f_i is the frequency in the normal distribution associated with $\hat{\rho}_i$. The square root of this value (i.e., SD_ρ) is a more accurate estimate of the standard deviation of true validities.

This procedure was discussed in more detail by Law et al. (1994). They used computer simulation to compare the accuracy of this procedure to that of the older linear procedure and found the new procedure to be considerably more accurate. The increases in accuracy became more pronounced as the variability of the population correlations became larger. The new procedure increased accuracy even when the normality assumption described above was seriously violated. In Law et al.'s study, sample size was infinite, and thus there was no sampling error. If there is sampling error, as in real data, and if sampling error is incompletely corrected for as a result of use of r_i in the sampling error variance formula, the inaccuracy introduced by the failure to allow for the nonlinearity of the range correction may be intensified. In the present study, we examined the nonlinear range-correction procedure under realistic conditions of sampling error and contrasted the results obtained when r_i versus $\bar{r}$ was used in the sampling error variance formula. We believe the results provide useful information on the operational accuracy of these meta-analytic methods.

Method

The simulation methods used by Callender and Osburn (1980) and by Raju and Burke (1983) were not used in this study because samples of only infinite size were considered in these earlier studies. Thus, with these methods, the effect of improved sampling variance estimation could not be tested. Instead, methods similar to those used by Osburn, Callender, Greener, and Ashworth (1983) were used. However, Osburn et al. focused on validity generalization, not meta-analysis as a general technique. Thus, they considered the effects of criterion reliability but not predictor reliability. Therefore, modification of their design was required.

The simulation was divided into three different cases. Case 1 examined accuracy in predicting the mean and the variance of true population correlations across different procedures when true population correlations were constant (i.e., the true variance of population correlations was zero, $V_\rho = 0$). Three values of the true score correlation were simulated (ρ = .30, .50, and .70). Case 2 focused on accuracy in estimating the true mean population correlation and V_ρ for different procedures with variable true population correlations (i.e., true $V_\rho > 0$). The hypothetical distribution of population correlations used by Callender and Osburn (1980) was considered unrealistic (Law et al., 1994) because in real data, population correlations in a meta-analysis would be unlikely to range from .06 to .94 (true V_ρ = .0833). Osburn et al. (1983) tested the accuracy of different procedures with four true population correlation distributions of different variability. However, except for the distribution labeled as having *small* variance, the other distributions of true population correlations had unrealistically large variances. (Interested readers are referred to the original article by Osburn et al., 1983, for a discussion of their distributions.) Therefore, three more realistic distributions of true population correlations were used in the present study. These distributions are listed in Table 1.

These three distributions have mean population correlations of .30, .50, and .70, respectively, which correspond to the values in Case 1 of this study. Each of these distributions had a true variance of .0030, which was similar to the mean V_ρ estimate that Schmidt et al. (1993) found by using a refined estimation procedure with large empirical databases. The distributions used in this study should therefore be more realistic and empirically based than most of those used in past simula-

Table 1
Case 2. Hypothetical Distributions of Population Correlations Developed for This Study

Frequency	Population 1 (ρ)	Population 2 (ρ)	Population 3 (ρ)
.10	.20	.40	.60
.20	.25	.45	.65
.40	.30	.50	.70
.20	.35	.55	.75
.10	.40	.60	.80

Note. The mean population correlations for Populations 1, 2, and 3 were .30, .50, and .70, respectively. For all populations, the standard deviation was .0548, and the variance was .0030.

tion studies. In addition, by holding true variance of population correlations constant at .0030, we could test the accuracy of different procedures with respect to different values of the mean true population correlation; such an analysis had not been conducted previously. However, to ensure generalizability of the results, we also tested three distributions of true validities with larger variance; these were the same as those used by Osburn et al. (1983) and are presented in Table 2. These three distributions are collectively referred to as Case 3 in this study.

Estimates from 12 different combinations of procedures were compared in this study. These procedures originated from five meta-analytic models—the Schmidt–Hunter interactive and noninteractive models, the Callender–Osburn (Callender & Osburn, 1980) multiplicative model, and the two Taylor Series models of Raju and Burke (1983). The sampling variance estimation refinement was applicable to all five procedures. Therefore, sampling variance estimates calculated with r and $\bar{r}$ for each of these five procedures were compared, which produced 10 combinations. The nonlinear range-restriction correction was applicable only to the Schmidt–Hunter procedures (Law et al., 1994). Therefore, the remaining two combinations were the interactive and noninteractive procedures with nonlinear range-restriction correction. On the basis of the findings of Law et al. (1994), the nonlinear range-restriction correction was expected to increase accuracy in Case 2 analyses (in which $V_\rho > 0$); however, it was expected to have a smaller beneficial effect in Case 1 analyses (in which $V_\rho = 0$). The use of $\bar{r}$ in the sampling variance formula was expected to improve accuracy regardless of the value of V_ρ. As stated earlier, one objective of this study was to determine the combined effects of these two refinements. Raju, Burke, Normand, and Langlois's (1991) procedure was found by Law et al. (1994) to be considerably less accurate in estimating V_ρ than the more traditional procedures; therefore, it was not included in the present study. Also, because of the reasons explained in detail by Law et al., Thomas's (1990) method also could not be included.

Table 2
Case 3. Osburn et al.'s (1983) Distributions of Population Correlations Used for This Study

Frequency	Population 1 (ρ)	Population 2 (ρ)	Population 3 (ρ)
1	.74	.82	.90
1	.74	.78	.85
2	.68	.74	.80
2	.65	.70	.75
3	.62	.66	.70
4	.59	.62	.65
4	.56	.58	.60
5	.53	.54	.55
6	.50	.50	.50
5	.47	.46	.45
4	.44	.42	.40
4	.41	.38	.35
3	.38	.34	.30
2	.35	.30	.25
2	.32	.26	.20
1	.29	.22	.15
1	.26	.18	.10

Note. The mean population correlation for all populations was .50. The variances for Populations 1, 2, and 3 were .0122, .0218, and .0340, respectively.

The artifact distributions used by Schmidt et al. (1979, 1980) were used as the distributions of reliabilities and range-restriction levels. As is the usual case in meta-analysis, the artifacts were assumed to be uncorrelated; that is, it was assumed there was no correlation between degree of range restriction and reliability values. The rationale for this assumption was discussed by Schmidt et al. (1980). These artifact distributions were used both in the generation of the data that were analyzed and in the meta-analysis of those data. Thus, the artifact distributions used in the meta-analyses were those known to be exactly appropriate for the data. A reviewer suggested that we also examine the effect on accuracy of using incorrect or inappropriate artifact distributions. Use of such distributions would cause error in the estimated means and variances. However, even under these circumstances, addition of the refinements examined in this study would make these estimates more accurate, relatively speaking; that is, the results would be less inaccurate with the refinements than without them. Thus, use of inappropriate artifact distributions would have no effect on the evaluation the accuracy improvements produced by the refinements, and therefore, the effects of using inappropriate artifact distributions are not relevant to the purposes of this research. Concern about the possible effects of inappropriate artifact distributions seems to stem from the belief that discrepancies between our artifact distributions and those appropriate for empirical data might be large. This concern does not appear to be justified. Our distributions were carefully developed on the basis of examination of numerous validity studies and data sets in which predictor and criterion reliabilities and range-restriction values were either given or could be computed. These distributions were later checked against figures from large empirical databases, yielding considerable empirical evidence that these distributions closely match those found in real data (Alexander, Carson, Alliger, & Cronshaw, 1989; Rothstein, 1990; Schmidt, Hunter, Pearlman, & Hirsh, 1985, question and answer 26, pp. 750–756). Thus, reported validity generalization findings are unlikely to be affected in any significant way by mismatching of artifact distributions.

In this study, the number of studies in a meta-analysis and the sample size of each study within one meta-analysis (simulation run) were randomly assigned. The number of studies in a meta-analysis was restricted to be greater than 30 but less than 100. These limits were chosen to represent the usual number of studies in a meta-analysis in the literature. Within a study, the sample size was also randomly assigned with limits of 30 and 150. These limits were chosen in consideration of the fact that Lent, Auerbach, and Levin (1971) found a median sample size of 68 in the personnel research literature. Five hundred simulation runs were conducted for each population correlation or distribution of population correlations, each with a different value for number of studies and sample size for each study.

At the beginning of each simulation, a random number was first generated by using the IMSL (International Mathematical and Statistical Library) subroutine DRNUNE (Double-Precision Random Number Uniform Distribution Function). This number was linearly transformed to impose lower and upper limits of 30 and 100, respectively, and was then used as the number of studies in this particular meta-analysis (this particular run of the 500 replications of the simulation). Thus, a different number of studies was generated for each of the 500 simulations. Another random number was subsequently generated and linearly transformed with limits of 30 and 150. This number was then used as the sample size for one study within this meta-analysis (simulation run). Next, one value for criterion reliability (ρ_{yy}) and one value for predictor reliability (ρ_{xx}) were randomly chosen from the two reliability distributions. The true predictor (x_t) and criterion (y_t) values were normally distributed with a mean of zero and a standard deviation of unity. A random true score, x_t, was sampled from that standard, normal true-score distribution by using the IMSL subroutine DRNNOF (Double-Precision Random Number Normally Distributed Function). A value for measurement error was randomly sampled from a normal distribution with a mean of zero and a standard deviation equal to the standard error of measurement (σ_{e_x}). The predicted true score ($\hat{y}_t$) corresponding to the generated true score x_t was found by using the preset population correlation between true scores x_t and y_t. The actual true score y_t (given a particular x_t value) was then sampled from a normal distribution with a mean of $\hat{y}_t$ and a standard deviation of $\sigma_{y_t \cdot x_t}$ (i.e., the conditional standard deviation of true scores). Next, a measurement error for the true score y_t was sampled from a normal distribution with a mean of zero and a standard deviation of σ_{e_y} (the standard error of measurement for y; σ_{e_y} was estimated with a procedure similar to that for σ_{e_x}). Finally, the observed y score is the sum of actual true score y_t and error score y_e.

This data generation procedure can be mathematically represented by the following equations:

$$x = x_t + e_x, \tag{4}$$

$$y = y_t + e_y, \tag{5}$$

$$\sigma^2_{x_t} = \sigma^2_{y_t} = 1.00, \tag{6}$$

$$\sigma_{e_x} = \sigma_x\sqrt{1 - \rho_{xx}} = \frac{\sigma_{x_t}}{\sqrt{\rho_{xx}}}\sqrt{1 - \rho_{xx}} = \sqrt{\frac{1 - \rho_{xx}}{\rho_{xx}}} \text{ and} \tag{7a}$$

$$\sigma_{e_y} = \sqrt{\frac{1 - \rho_{yy}}{\rho_{yy}}}, \tag{7b}$$

$$\hat{y}_t = \rho_{x_t y_t}\left[\frac{\sigma_{y_t}}{\sigma_{x_t}}\right]x_t = \rho_{x_t y_t}x_t, \tag{8}$$

$$\sigma_{y_t \cdot x_t} = \sqrt{1 - \rho^2_{x_t y_t}}, \tag{9}$$

where x is the observed x score; x_t is the true x score, $N(0, 1)$; y_t is the true y score, $N(0, 1)$; e_x is the error score for x, $N(0, \sigma_{e_x})$; e_y is the error score for y, $N(0, \sigma_{e_y})$; σ^2_x is the variance of the observed x score, $\sigma^2_x > 1.00$; σ^2_y is the variance of the observed y score, $\sigma^2_y > 1.00$; σ_{e_x} is the standard error measurement of x; σ_{e_y} is the standard error measurement of y; ρ_{xx} is the reliability of x; ρ_{yy} is the reliability of y; $\rho_{x_t y_t}$ is the correlation between true score x and true score y; $\hat{y}_t$ is the predicted true score y given a true score (x_t); and $\sigma_{y_t \cdot x_t}$ is the conditional standard deviation of y_t (conditional on x_t).

To incorporate range restriction on the observed x–y pairs, a selection ratio was randomly selected from the distribution of range-restriction values. The critical z score corresponding to the generated selection ratio was calculated. Any generated x–y pair with x smaller than the critical value was rejected. This procedure was repeated until the desired number of x–y pairs (sample size for each correlation) was obtained. The correlation between these x–y pairs was calculated as the observed correlation for that study. This procedure for generating observed correlations was repeated until the desired number of correlations for a meta-analysis was obtained.

There were 500 simulated meta-analyses for each value of the population correlation in Case 1. The simulation procedure for Cases 2 and 3 was the same as that for Case 1. The only difference was that instead of using one specific true-score population correlation for all 500 runs of the simulation, a randomly drawn true population correlation from the prescribed distributions in Table 1 or Table 2 (with replacement) was used each time in generating each observed correlation for a meta-analysis. Law et al. (1994) discuss the importance of sampling with replacement. Again, 500 simulated meta-analyses were conducted for each value of the mean true population correlation.

Results and Discussion

Results for Simulation Case 1

The simulation results for 500 runs for different fixed population correlation values (ρ = .30, .50, and .70) are presented in Tables 3 and 4. It can be seen that mean estimates of the true population correlations of the six procedures were very similar. Differences across procedures were within rounding error. (In actual usage, estimates of the true mean population correlation are routinely rounded to two decimal places.) Mean population correlation estimates of the interactive and noninteractive procedures with linear range-restriction correction are presented together because these procedures yield identical estimated mean population correlations when the older linear range-restriction correction procedure is used (see Law et al., 1994). We conclude that all of these procedures are about equally accurate in estimating the true population correlation when it is constant. In terms of estimating the mean population correlation, the refined Schmidt–Hunter procedures do not increase the accuracy of the original procedures because the older procedures are already very accurate.

Table 4 summarizes the accuracy of different procedures for estimating the variance of population correlations for different values of true mean population correlations. Because there was only one population correlation in this case, the true variance of population correlations is zero in all cells, and the mean estimated population variance is also the mean error. Both of the refinements led to a smaller estimation of the true variance of population correlations and, thus, to increased accuracy of the V_ρ estimates. For the interactive procedure, application of the two refinements reduced the mean V_ρ estimate from .0089 to .0064, a 28% increase in accuracy. For the noninteractive procedure, the mean SD_ρ estimate decreased from .0046 to .0033 (a 28% decrease) when the two refinements were used. It appears that the nonlinear range-restriction correction procedures were more accurate than the corresponding linear range-restriction correction procedures even in Case 1, where $V_\rho = 0$. The reason is that even though estimates of the residual standard deviation (SD_{res}) are usually very small in Case 1, these estimates are increased less when they are transformed into estimates of SD_ρ (and V_ρ).

As expected, the procedures that used the mean correlation in estimating sampling variance were more accurate than the corresponding procedures that used individual correlations, as demonstrated analytically by Hunter and Schmidt (1994). On average, the error in estimating V_ρ decreased by about .001 when $\bar{r}$ was used instead of r. This was true even for the multiplicative and Taylor Series procedures. When the nonlinear range-restriction correction was used for the Schmidt–Hunter procedures, error in estimating V_ρ decreased futher.

Table 3
Case 1. Mean Population Correlation (M_ρ) Estimates by Different Procedures When True V_ρ = 0

Procedure	True M_ρ = .30			True M_ρ = .50			True M_ρ = .70		
	M	Error	% error	*M*	Error	% error	*M*	Error	% error
Interactive/noninteractive, linear	.2991	−.0009	−0.30	.4975	−.0025	−0.50	.6974	−.0026	−0.37
Interactive, nonlinear	.2982	−.0018	−0.60	.4960	−.0040	−0.80	.6951	−.0049	−0.70
Noninteractive, nonlinear	.2984	−.0016	−0.53	.4968	−.0032	−0.64	.6966	−.0034	−0.49
Multiplicative	.2989	−.0011	−0.40	.4966	−.0034	−0.68	.6952	−.0048	−0.69
First Taylor Series	.2963	−.0037	−1.20	.4930	−.0070	−1.40	.6910	−.0090	−1.30
Second Taylor Series	.2991	−.0009	−0.30	.4975	−.0025	−0.50	.6974	−.0026	−0.37

Note. Mean values of the M_ρ estimates are based on 500 simulation runs. V_ρ = variance of population correlation.

In comparison with other procedures, the refined Schmidt–Hunter procedures gave more accurate estimates of the true variance of population correlations than the multiplicative procedure. As shown in Table 4, mean V_ρ estimates produced by the refined interactive and noninteractive procedures were .0064 and .0033, respectively, as compared with .0089 for the multiplicative procedure using $\bar{r}$. The two Taylor Series procedures using $\bar{r}$ yielded the same mean V_ρ estimate up to the third decimal place as the refined Schmidt–Hunter interactive procedure (rows 10 and 12 vs. row 3).

All procedures overestimated the true variance of population correlations (V_ρ). However, it is an inherent property of all of these procedures that any negative variance estimate must be set to zero. Therefore, with a true V_ρ of zero, no procedure could produce an underestimate. One reviewer was concerned that, under the circumstances of Case 1 (V_ρ = 0), results for the accuracy of the V_ρ estimates might be biased; this is not the case. The purpose of this study was to evaluate the accuracy of these procedures as they are actually used to analyze data. In real applications of these methods, negative variance estimates cannot be used and are not used. Therefore, it is irrelevant whether such negative estimates would create bias if they were used. In the case of the interactive and noninteractive procedures, an essential property of the methods is that the estimate of residual variance is set to zero when the observed estimate of residual variance is negative, leading to an estimate of zero for V_ρ. The other procedures do not define or compute a residual variance (see Law et al., 1994, for a discussion of this point). However, an essential property of each of these other methods is that any negative estimates of V_ρ are set to zero. This rule is similar to that in Cronbach's generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; see discussion in Hunter & Schmidt, 1990, p. 413). Thus, the figures for V_ρ in Table 4 are not biased estimates. It is simply a fact that in cases in which V_ρ is equal to zero, all procedures overestimate V_ρ.

Results for Simulation Case 2

The simulation results for Case 2 ($V_\rho > 0$) are presented in Tables 5 and 6. Table 5 shows the mean population correlation estimates when true V_ρ is equal to .0030. Once again, all procedures yielded very accurate estimates (to two decimal places) of the mean population correlations. Furthermore, the refined procedures did not contribute more accurate estimates because the original estimates were already very accurate.

Table 6 shows the accuracy of V_ρ estimates for different procedures. As in Case 1, most of the procedures, on average, overestimated V_ρ. However, in this case, an underestimation of V_ρ

Table 4
Case 1. Variance of Population Correlation (V_ρ) Estimates by Different Procedures When True V_ρ = 0

Procedure	True M_ρ = .30	True M_ρ = .50	True M_ρ = .70	Mean
Interactive, linear, individual *r*	.0082	.0089	.0096	.0089
Interactive, linear, mean *r*	.0071	.0078	.0087	.0079
Interactive, nonlinear, mean *r*	.0065	.0065	.0062	.0064
Noninteractive, linear, individual *r*	.0062	.0043	.0032	.0046
Noninteractive, linear, mean *r*	.0053	.0037	.0028	.0039
Noninteractive, nonlinear, mean *r*	.0048	.0031	.0020	.0033
Multiplicative, individual *r*	.0079	.0093	.0127	.0100
Multiplicative, mean *r*	.0068	.0082	.0117	.0089
TSA1, individual *r*	.0077	.0074	.0063	.0071
TSA1, mean *r*	.0066	.0065	.0057	.0063
TSA2, individual *r*	.0078	.0074	.0061	.0071
TSA2, mean *r*	.0067	.0064	.0055	.0062

Note. Mean V_ρ estimates are based on 500 simulation runs. M_ρ = mean population correlation; TSA1 = first Taylor Series; TSA2 = second Taylor Series.

Table 5
Case 2. Mean Population Correlation (M_ρ) Estimates by Different Procedures When True V_ρ = .0030

Procedure	True M_ρ = .30			True M_ρ = .50			True M_ρ = .70		
	M	Error	% error	M	Error	% error	M	Error	% error
Interactive/noninteractive, linear	.2981	−.0019	−0.63	.4992	−.0008	−0.16	.6964	−.0036	−0.51
Interactive, nonlinear	.2974	−.0026	−0.87	.4980	−.0020	−0.40	.6947	−.0053	−0.76
Noninteractive, nonlinear	.2976	−.0024	−0.80	.4987	−.0013	−0.26	.6960	−.0040	−0.57
Multiplicative	.2979	−.0021	−0.70	.4983	−.0017	−0.34	.6942	−.0058	−0.83
First Taylor Series	.2953	−.0047	−1.60	.4947	−.0053	−1.10	.6900	−.0100	−1.40
Second Taylor Series	.2981	−.0019	−0.63	.4992	−.0008	−0.16	.6964	−.0036	−0.51

Note. Mean values of the M_ρ estimates are based on 500 simulation runs.

was possible because the true value of V_ρ was .0030 and the minimum possible value was zero. The only procedure that gave an underestimated mean V_ρ was the noninteractive procedure for large values of mean population correlations (mean population correlation = .50 and .70). This finding implies that accuracy of this procedure may vary with the values of true mean correlations. The multiplicative procedure, on the other hand, showed the opposite pattern: the larger the mean population correlation, the larger were the V_ρ estimates. As expected, both of the refinements led to more accurate mean estimates of V_ρ for the interactive procedure. However, for the noninteractive procedure, this was not the case; as the true mean population correlation increased from .30 to .50 to .70, true variance estimates produced by the noninteractive procedure decreased, as did the accuracy of estimation. When the true mean population correlation equaled .50 or .70, the refined noninteractive procedures underestimated the true variance of population correlations. From this analysis, we conclude that when true population correlations vary, the refined procedures give a more accurate estimate of V_ρ when they are used with the interactive procedure.

When the refined Schmidt–Hunter procedures were compared with the multiplicative procedure and the two Taylor Series procedures, results similar to those for Case 1 were observed. The two Schmidt–Hunter procedures gave more accurate mean V_ρ estimates than the multiplicative procedure. The two Taylor Series procedures with $\bar{r}$ yielded the same V_ρ estimates to the third decimal place as the interactive procedure with refinements (rows 10 and 12 vs. row 3).

Results for Simulation Case 3

Tables 7 and 8 show the estimates of different procedures when the very wide distributions of population correlations used by Osburn et al. (1983) were analyzed. Results from Tables 7 and 8 demonstrate the usefulness of the two refinements when the true variance of population correlations is very large (V_ρ = .0122, .0218, and .0340). It is clear from Tables 7 and 8 that the conclusions derived from Cases 1 and 2 can be applied equally to wide distributions of population correlations with large variances. In Table 7, it can be seen that the mean population correlation estimates were very similar across procedures and again were the same within rounding error.

For mean V_ρ estimates, the effect of the refined procedures are more pronounced than in Case 2. The V_ρ estimates for the interactive refined procedure are extremely accurate with these wide distributions of true population correlations. For example, for the widest of these distributions, the mean V_ρ estimate of the

Table 6
Case 2. Variance of Population Correlation (V_ρ) Estimates by Different Procedures When True V_ρ = .0030

Procedure	M_ρ = .30	M_ρ = .50	M_ρ = .70	M	Error	% error
Interactive, linear, individual r	.0062	.0069	.0073	.0068	.0038	126.67
Interactive, linear, mean r	.0052	.0059	.0065	.0059	.0029	95.56
Interactive, nonlinear, mean r	.0048	.0049	.0046	.0048	.0018	58.89
Noninteractive, linear, individual r	.0045	.0031	.0018	.0031	.0001	4.44
Noninteractive, linear, mean r	.0037	.0026	.0015	.0026	−.0004	−13.33
Noninteractive, nonlinear, mean r	.0034	.0021	.0011	.0022	−.0008	−26.67
Multiplicative, individual r	.0060	.0073	.0100	.0078	.0048	158.89
Multiplicative, mean r	.0050	.0063	.0091	.0068	.0038	126.67
TSA1, individual r	.0058	.0057	.0048	.0054	.0024	81.11
TSA1, mean r	.0049	.0049	.0042	.0047	.0017	55.56
TSA2, individual r	.0059	.0057	.0046	.0054	.0024	80.00
TSA2, mean r	.0049	.0049	.0040	.0046	.0016	53.33

Note. Mean V_ρ estimates are based on 500 simulation runs. M_ρ = mean population correlation; TSA1 = first Taylor Series; TSA2 = second Taylor Series.

Table 7
Case 3. Mean Population Correlation (M_ρ) Estimates by Different Procedures When Osburn et al.'s (1983) Distributions of Population Correlations Are Used

Procedure	Moderate M	Moderate Error	Moderate % error	Large M	Large Error	Large % error	Extreme M	Extreme Error	Extreme % error
Interactive/noninteractive, linear	.4977	−.0023	−0.46	.5001	.0001	0.02	.5037	.0037	0.74
Interactive, nonlinear	.4947	−.0053	−1.06	.4945	−.0055	−1.10	.4951	−.0049	−0.98
Noninteractive, nonlinear	.4960	−.0040	−0.80	.4961	−.0039	−0.78	.4968	−.0032	−0.64
Multiplicative	.4968	−.0032	−0.64	.4991	−.0009	−0.18	.5028	.0028	0.56
First Taylor Series	.4931	−.0069	−1.38	.4955	−.0045	−0.90	.4991	−.0009	−0.18
Second Taylor Series	.4977	−.0023	−0.46	.5001	.0001	0.02	.5037	.0037	0.74

Note. Mean values of the M_ρ estimates are based on 500 simulation runs.

interactive procedure with nonlinear range-restriction correction and use of $\bar{r}$ in sampling error estimation was accurate up to the third decimal place (.0353 vs. the true value of .0340; row 3 in Table 8 for the distribution labeled *extreme*), for an error of only 3.82%. The results in Table 8 show that the refined interactive procedure is the most accurate procedure when the true population variance is large. The refined interactive procedure is more accurate than the Taylor Series procedures for the distributions of population correlations in Table 8. As expected, the mean V_ρ estimates of the multiplicative procedure and the two Taylor Series procedures were more accurate when $\bar{r}$ instead of r was used in the sampling error estimation formula, which was the same pattern of findings observed earlier in Cases 1 and 2.

Conclusions

This study examined the usefulness of the nonlinear range-restriction correction process and the use of $\bar{r}$ in sampling variance estimation in increasing the accuracy of estimates of M_ρ and V_ρ in meta-analysis. A previous study (Law et al., 1994) had shown that the nonlinear range correction improved accuracy in the absence of sampling error (all *N*s infinite). The present study examined accuracy under realistic conditions of sampling error (*N*s between 30 and 150) and the effect on accuracy of using $\bar{r}$ in the formula for sampling error variance. From the above findings, we conclude that, when used with the interactive procedure, both of the refinements are effective in reducing error in estimating V_ρ. Thus, our initial hypothesis is supported. For example, when V_ρ is equal to 0 (Case 1), the mean V_ρ estimate produced by the interactive procedure with the two refinements shows a 28% reduction in error when compared with the value produced without the refinements. For a mean true-score correlation of .50, the effect of this increase in accuracy is to increase the 90% credibility estimate from .38 to .40. When V_ρ is .0030 (Case 2), the refined interactive procedure is 53% more accurate in estimating V_ρ as compared with the original interactive procedure. For a mean true-score correlation of .50, the effect of this increase in accuracy is to increase the 90% credibility estimate from .39 to .41. With the high variance distributions of population correlations (Case 3), the refinements produce reductions in error of 98%, 87%, and 89% for the mod-

Table 8
Case 3. Variance of Population Correlation (V_ρ) Estimates by Different Procedures When Osburn et al.'s (1983) Distributions of Population Correlations Are Used

Procedure	Moderate (V_ρ = .0122) M	Error	% error	Large (V_ρ = .0218) M	Error	% error	Extreme (V_ρ = .0340) M	Error	% error
Interactive, linear, individual r	.0164	.0042	34.43	.0297	.0079	36.24	.0456	.0116	34.12
Interactive, linear, mean r	.0149	.0027	22.13	.0278	.0060	27.52	.0434	.0094	27.65
Interactive, nonlinear, mean r	.0123	.0001	0.82	.0228	.0010	4.59	.0353	.0013	3.82
Noninteractive, linear, individual r	.0095	−.0027	−22.13	.0214	−.0004	−1.83	.0366	.0026	7.65
Noninteractive, linear, mean r	.0083	−.0039	−31.97	.0197	−.0021	−9.63	.0343	.0003	0.88
Noninteractive, nonlinear, mean r	.0069	−.0053	−43.44	.0162	−.0056	−25.69	.0281	−.0059	−17.35
Multiplicative, individual r	.0167	.0045	36.89	.0295	.0077	35.32	.0447	.0107	31.47
Multiplicative, mean r	.0152	.0030	24.59	.0277	.0059	27.56	.0425	.0085	25.00
TSA1, individual r	.0137	.0015	12.30	.0248	.0030	13.76	.0379	.0039	11.47
TSA1, mean r	.0124	.0002	1.64	.0232	.0014	6.42	.0360	.0020	5.88
TSA2, individual r	.0137	.0015	12.30	.0250	.0032	14.68	.0383	.0043	12.65
TSA2, mean r	.0124	.0002	1.64	.0233	.0015	6.88	.0364	.0024	7.06

Note. Mean V_ρ estimates are based on 500 simulation runs. TSA1 = first Taylor Series; TSA2 = second Taylor Series.

erate, large, and extreme distributions, respectively, as shown in Table 8. For the large variance distribution in Table 8, the effect of this increase in accuracy is to increase the 90% credibility estimate from .28 to .30. For the high-variance correlation distributions, the refined interactive procedure is the most accurate of the procedures tested in this study.

For all of the procedures, use of $\bar{r}$ in the sampling error variance formula improved the accuracy of the V_ρ estimates. This was true for the multiplicative and Taylor Series procedures as well as for the interactive procedures, and was true under all conditions examined. Thus, these findings show that use of $\bar{r}$ in the sampling error variance formula increases accuracy in the heterogeneous case ($V_\rho > 0$). Previously, this had been demonstrated only for the homogeneous case ($V_\rho = 0$; Hunter & Schmidt, 1994).

An interesting point to be noted from the results in this study is that, for some procedures under study, the accuracy of the mean V_ρ estimate is dependent on the value of the true mean population correlation (M_ρ). For example, in Table 6, the mean V_ρ estimate of the noninteractive procedure and the Taylor Series procedures decreased as true M_ρ increased from .30 to .50 to .70. In contrast, the mean V_ρ estimate of the multiplicative procedure increased as true M_ρ increased from .30 to .50 to .70. This *mean population correlation dependency* of some procedures in estimating V_ρ is a new finding. However, the differences of V_ρ estimates with different M_ρ values are small for most procedures. Therefore, this dependency effect of V_ρ estimates should not be overemphasized. Further studies are needed to explain this phenomenon.

In estimating the mean population correlation, the refined procedures do not yield improvements in accuracy because the original procedures are already very accurate. As is well-known, the major questions of accuracy in meta-analysis have traditionally centered on estimates of V_ρ (or SD_ρ) and not on M_ρ. In estimating V_ρ, the two refinements used with the interactive procedure appear to lead to substantial improvements in accuracy under realistic sample-size conditions. Taken together with the findings of Law et al. (1994), the findings of the present study support the conclusion that, considered across a variety of types of true population correlation distributions, the procedure that is most frequently the most accurate is the interactive procedure used with the two refinements examined in this study.

References

Alexander, R. A., Carson, K. P., Alliger, G. M., & Cronshaw, S. F. (1989). Empirical distributions of range restricted SD_x in validity studies. *Journal of Applied Psychology, 74,* 253–258.

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

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. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. *Journal of Applied Psychology, 79,* 171–177.

Kemery, E. R., Mossholder, K. W., & Roth, L. (1987). The power of the Schmidt and Hunter additive model of validity generalization. *Journal of Applied Psychology, 72,* 30–37.

Law, K., Schmidt, F. L., & Hunter, J. E. (1994). Nonlinearity of range corrections in meta-analysis: Test of an improved procedure. *Journal of Applied Psychology, 79,* 425–438.

Lent, R. H., Auerbach, H. A., & Levin, L. S. (1971). Predictors, criteria and significant results. *Personnel Psychology, 24,* 519–533.

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

Millsap, R. E. (1988). Sampling variance in attenuated correlation coefficients: A Monte Carlo study. *Journal of Applied Psychology, 73,* 316–319.

Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. (1983). Statistical power of tests of the situational specificity hypothesis in validity generalization studies: A cautionary note. *Journal of Applied Psychology, 68,* 115–122.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65,* 373–406.

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.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75,* 322–327.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist, 47,* 1173–1181.

Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65,* 643–661.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, 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,* 257–281.

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.

Thomas, H. (1990). A likelihood-based model for validity generalization. *Journal of Applied Psychology, 75,* 13–20.

Received September 10, 1992
Revision received May 16, 1994
Accepted May 18, 1994 ■