June 1995

On a Proper Meta-Analytic Model for Correlations

Amir Erez
Cornell University

Matthew C. Bloom
Cornell University

Martin T. Wells
Cornell University

On a Proper Meta-Analytic Model for Correlations

Abstract
Combining statistical information across studies is a standard research tool in applied psychology. The most common approach in applied psychology is the fixed effects model. The fixed-effects approach assumes that individual study characteristics such as treatment conditions, study context, or individual differences do not influence study effect sizes. That is, that the majority of the differences between the effect sizes of different studies can be explained by sampling error alone. We critique the fixed-effects methodology for correlations and propose an advancement, the random-effects model, that ameliorates problems imposed by fixed-effects models. The random-effects approach explicitly incorporates between-study differences in data analysis and provides estimates of how those study characteristics influence the relationships among constructs of interest. Because they can model the influence of study characteristics, we assert that random-effects models have advantages for psychological research. Parameter estimates of both models are compared and evidence in favor of the random-effects approach is presented.

Keywords
research, model, organization, review, information, psychology, applied, study, effect

CAHRS / Cornell University
187 Ives Hall
Ithaca, NY 14853-3901 USA
Tel. 607 255-9358
www.ilr.cornell.edu/CAHRS/

CAHRS
Center for Advanced Human Resource Studies

WORKING PAPER SERIES

On a Proper Meta-Analytic Model for Correlations

Amir Erez
Matthew C. Bloom
Martin T. Wells

Working Paper 95–11

ilr School of Industrial and Labor Relations
CORNELL UNIVERSITY
Advancing the World of Work

On a Proper Meta-Analytic Model for Correlations

Amir Erez
Department of Human Resource Studies
School of Industrial & Labor Relations
393 Ives hall
Cornell University
Ithaca, New York 14853-3901
Phone: (607)255-5427
FAX: (607)255-9862

Matthew C. Bloom
Department of Human Resource Studies
Center for Advanced Human Resource Studies
School of Industrial & Labor Relations
Cornell University

Martin T. Wells
Department of Economic & Social Statistics
New York State School of Industrial & Labor Relations
Cornell University

Working Paper 95-11

www.ilr.cornell.edu/cahrs

Abstract

Combining statistical information across studies is a standard research tool in applied psychology. The most common approach in applied psychology is the fixed effects model. The fixed-effects approach assumes that individual study characteristics such as treatment conditions, study context, or individual differences do not influence study effect sizes. That is, that the majority of the differences between the effect sizes of different studies can be explained by sampling error alone. We critique the fixed-effects methodology for correlations and propose an advancement, the random-effects model, that ameliorates problems imposed by fixed-effects models. The random-effects approach explicitly incorporates between-study differences in data analysis and provides estimates of how those study characteristics influence the relationships among constructs of interest. Because they can model the influence of study characteristics, we assert that random-effects models have advantages for psychological research. Parameter estimates of both models are compared and evidence in favor of the random-effects approach is presented.

Meta-analysis offers the potential to combine information from a number of studies to provide a concise information summary about relationships of interest. As such, meta-analysis is one of the most powerful and useful statistical techniques available to researchers. The ideas of combining information go back as far as Legendre (1805) in an attempt to use astronomical observations collected at several different observatories to estimate the orbits of comets and Gauss (1809) to determine the meridian arcs in geodesy. Earlier in this century, Birge (1932) was among the first to combine estimates across experiments at different laboratories to establish reference values for the fundamental constants of physics. Around the same time techniques for combining information across agricultural experiments were being developed by Cochran (1937), Cochran and Yates (1938), Trippett (1931), and Fisher (1932). These techniques were extended by Glass (1976). In applied psychology, the message about meta-analysis brought by Hunter, Schmidt, & Jackson (1982) caused an important transformation in research. These researchers demonstrated that information from multiple studies can be combined in meaningful ways. In their validity generalization research, they showed that selection procedures could be generalized across occupations and organizations. This means that separate and costly validity studies on a particular selection procedure may not be needed for each individual occupation.

In applied psychology, the most common meta-analytic technique for correlations is the fixed-effects model (Hunter & Schmidt, 1990). In fact, until now researchers in the field have relied almost exclusively on the fixed-effects model and have not used other methods, such as the random-effects models. Given the availability of these alternative models, we argue that psychologists should take advantage of random-effects models because of their superior statistical and analytic qualities. In 1992 a panel of experts on meta-analysis, commissioned by the National Research Council to evaluate the state of knowledge about meta-analysis, stated that the fixed effects model carries with it restrictive and often incorrect methodological assumptions. The panel concluded by stating "[meta-analytic] modeling would be improved by the increased use of random effects models in preference to the current default of fixed effects models (National Research Council, 1992:185)." Drawing upon advances in theoretical and applied statistics, this paper presents an overview of the random-effects model and compares it to the fixed-effects approach.

Random- versus Fixed-Effects Model--Conceptual Differences

The underlying assumption of meta-analysis is that combining information from independent, but similar studies, improves the estimation of population parameters over that obtained from any single study. It is useful to compare a meta-analysis to a regression situation

where the researcher wishes to understand the variability in a random variable, say organizational commitment (Y_i). In the simplest case, the researcher has only the mean ($\bar{Y}$) to use in estimating Y_i. A more sophisticated approach is to use information from an explanatory variable, for example the amount of pay an incumbent receives (X_i), in a regression of Y_i on X_i such that $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. The use of β_1 to relate X_i and Y_i is superior to $\bar{Y}$ because much more can be learned about how Y_i varies systematically with X_i. Furthermore, including ε_i is important because it accounts for factors excluded from the model that also cause variance in Y_i. The model might be improved by using multiple predictor variables, X_{ij}, such as various job characteristics in addition to pay information, that are theoretically important for understanding Y so that $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_j X_{ij} + \varepsilon_i$. Finally, the theory might suggest that group differences, such as company size or the industry group (W_{qj}), influence the effect of each X_{ij} on Y_i. An improved model would account for these differences in the effect of X_{ij} by modeling how the β_j's systematically vary due to these group differences. Often this is accomplished through interactions, but an improved approach is to account for the systematic variance in the β_j's by a second stage model where $\beta_j = \gamma_0 + \gamma_q W_{qj} + \ldots + u_j$ (Bryk & Raudenbush, 1992). In this model group differences in the β_j's can be explained by group-level covariates, the W_{qj}'s--and u_j operates like ε_i by accounting for factors that cause systematic variation in β_j that are excluded from the model. Meta-analyses can be grouped in a similar fashion, drawing distinctions between "fixed-effects" and "random-effects" models. Fixed-effects are analogous to using the mean while random-effects are analogous to using a regression.

Suppose *k* studies provide information about an unknown correlation coefficient ρ. The meta-analyst wishes to use information contained in all *k* studies to derive an estimate of the true effect size ρ. However, since the *k* studies are not identical, statistical theory directs the meta-analyst to take into account the different sources of variation that might influence the estimation of ρ. In the random-effects model, two sources of variation are taken into account: within studies and between-studies variance. Random-effects models assume that the studies are heterogeneous, that is, the studies differ on important factors that influence study results. This is indeed likely to be the case in meta-analyses as the researcher is combining information across different studies where study context, treatments, research procedures, and group and individual characteristics are likely to be different. These two sources of variation are estimated in the random-effects model in hierarchical manner. The first level accounts for within-studies variance that influences individual study parameter estimates. Factors such as sample size, restriction of range, and reliability of the measurement tools are the main causes of this source

of variance (Hunter & Schmidt, 1990, Part 11). The second level error component, the uj's, accounts for the influence of the between-studies characteristics such as study context (e.g., organization size, human resources strategies, cultural differences) on observed study effect sizes. The statistical form of this random-effects model is represented by the following set of equations:

$$r_i = \rho_i + e_i \qquad e_i \sim N(0,\sigma^2) \qquad [1]$$

$$\rho_i = \rho + u_i \qquad u_i \sim N(0,\tau^2) \qquad [2]$$

where r_i and ρ_i are the observed and population correlation coefficients of study *i*, respectively, ρ is the population correlation coefficient for the entire population of studies, e_i is the within-studies variance, and u_i is the effect of between-studies differences on the correlation coefficient of study *i*, $1 \leq i \leq k$. From this set of equations it is clear that the random effects model accounts for the influence of differences across studies on individual study effect size estimates. Here, the obtained estimate of ρ_i is similar to the β_j in a regression. The total variance of r_i is divided into two components:

$$\mathrm{Var}(r_i) = \mathrm{Var}(u_I + e_i) = \tau^2 + \sigma^2. \qquad [3]$$

By including covariates, the model can be extended to include factors which account for systematic group differences that influence the ρ_i 's and hence explain variance in the individual r_I's. This is analogous to the use of W_{qj}'s in a regression situation mentioned above.

The premise of the fixed-effects model, on the other hand, is that differences between study effect sizes are only due to within-studies error variance; variation in parameter estimates (e.g., study correlation coefficients) is not due to systematic differences in study characteristics. For example, in organizational research this means that compensation policies, training programs, selection systems and other organizational policies do not influence psychological phenomena under study. The only variance that exists is within-study or random sampling error. Consequently, rather than a distribution of correlations across studies, the fixed-effects model assumes there is only one fixed and true population correlation, ρ. Accordingly,

$$r_i = \rho + e_I \qquad e_I \sim N(0,\sigma^2) \qquad [4]$$

for $1 \leq i \leq k$. In psychology, the most often used fixed-effects approach (Hunter & Schmidt, 1990) asserts that the average study correlation ($\bar{r}$) is the appropriate estimate of the population correlation coefficient. This is analogous to using the mean $\bar{Y}$ as an estimator. Although the fixed-effects model is clearly a submodel of the more general random-effects approach, fundamental differences separate these two models. While the random-effects model takes into account different sources of variation, the fixed-effects model places a restrictive

assumption on parameter estimation: that between-studies variance is zero (τ^2=0). It is therefore essential to ascertain whether this assumption is justified before applying the fixed-effects model.

Considering the nature of constructs studied in psychology and other social sciences, it seems that the 'homogeneity of studies' assumption is rarely met. Asserting that between-studies variance is zero is tantamount to stating that study context and procedures, individual differences, organizational practices, situational conditions, and cultural differences have no affect on the relationships among constructs of interest. The homogeneity assumption implies that pure relationships between variables can be effectively isolated in studies. Until recently, the predominant belief in the social sciences opposed this view. The following represents just a short list of the views of leading methodologists in the field about the homogeneity assumption:

> There is little question but that sizable differences, correlations, etc., in samples,... speak more strongly of sizable differences, correlation, etc., in the population (Bakan, 1966:429).
>
> ...the notion that correlations between arbitrarily paired [psychological] variables will be, while not literally zero, of such minuscule size to be of no importance, is surely wrong (Meehl, 1990:208).
>
> It is difficult to focus on the critical features of a particular problem in behavioral science when they occur in the midst of so much other stuff that may or may not affect the result (Campbell, 1990:46).

Even the strongest advocates of the fixed-effects model have stated that:

> ... studies are never perfect. Thus, the relationship between study correlations and actual correlations is more complicated … The complexity of formulas depend on two things: (a) the extent of variation in artifacts [within studies variance] and (b) the extent of variation in actual correlations [between-studies variance] (Hunter & Schmidt, 1990:43).

Accordingly, Hunter and Schmidt (1990) concluded that, when between-studies variance exists, using the fixed-effects approach to overall meta-analysis is not meaningful. These authors recommend that, in the case where substantial differences between studies exist, sub-groups of studies should be formed on the basis of moderator variables. Meta-analysis should be conducted on these sub-groups of similar studies and not on the overall pool of studies. This suggests that, before applying the fixed-effects approach, one should first estimate and verify that between studies variance is indeed inconsequential. Only when it is inconsequential can one use the fixed effects approach. The Hunter and Schmidt (1990; Part II) fixed-effects model purports to do exactly that. The preliminary assumption of this method is similar to that of the random-effects model, that two sources of variance might influence

observed correlations. These sources are the within-studies variance (σ_e^2) and the between studies variance (σ_r^2). The fixed-effects model (Hunter & Schmidt, 1990) beings with the following assumption:

$$\sigma_r^2 = \sigma_r^2 + \sigma_e^2 \qquad [5]$$

where σ_r^2 is the total variance of observed study correlations[1].

To determine if between-study variance is consequential, the fixed-effects model (Hunter & Schmidt, 1990) proceeds as follows. Having selected a group of studies which investigate specific theoretical relationships, the procedure begins with the computation of a weighted average correlation across studies ($\bar{r}$) and the variance of these observed study correlations, σ_r^2. Next, the analyst adjusts the variance of observed correlations (σ_r^2) for artifacts (σ_e^2). If artifact adjustments account for at least 75% of the variance in observed correlations Hunter and Schmidt (1990) direct the analyst to assume that the between-studies variance is zero (i.e., $\sigma_r^2 = 0$). Here, one uses the fixed-effects approach and assumes that $\bar{r}$ is a reasonable estimate of the true population correlation, ρ. When between-studies variance is substantial, the fixed-effects model advises the analyst to divide the studies into homogeneous groups and to proceed with a separate meta-analysis on each these homogeneous groups. Within each group, sources of true between studies variance are not allowed to influence the parameter estimates, and $\bar{r}$ is always the sole estimate of the population correlation. Note that no procedure for modeling how between-studies differences influence observed and population correlations is available. As a result, this method reduces to several fixed-effects models and cannot inform us about potential influences of between-study variance or within-study results.

The random-effects approach, on the other hand, expressly models both between-studies and within-studies (i.e., error) variance, and correctly allows them to influence parameter estimates. If between-studies variance is truly zero, parameter estimates of the fixed- and random-effects models will be the same. In the case of correlations, both estimate a single ρ. However, if between-studies variance is not equal to zero, this variance will be taken into account by the random-effects method when estimating the true mean correlation, ρ. The fixed-effects model simply assumes it is zero. The estimate of true population correlation remains meaningful in the random-effects case even in the presence of substantial between-studies variance and, as a result, the division of studies into homogeneous groups becomes unnecessary. If one wished for a more detailed assessment of ρ, it would be possible

[1] We note that it seems technically illogical to talk about variances of non-random quantities, i.e., σ_r^2, unless one is dealing with a random-effects model.

to model ρ_i in Equation 3 as a function of some observed covariates $\mathbf{W}_q = (W_{q1},...,W_{qj})$, so that now $\rho_i = \mathbf{W_q}'\beta+\mathbf{u_i}$. In this case attention shifts from an overall or 'grand' mean correlation to $\beta = (\beta_1,...,\beta_\varphi)$, the components of the overall mean correlation. Thus, meta-analytic research may be improved by using random-effects which account for and explain between-studies variance.

Methodological Limitations of the Fixed-Effects Model

Reliance on fixed-effects model in meta-analytic research in applied psychology has started a shift away from the situational specificity hypothesis which asserts that differences between studies influence effect sizes. Although one would expect to find between-studies variance in almost every meta-analysis conducted, the majority of meta-analytic studies in the field of applied psychology have found the opposite. However, these studies have relied on the fixed-effects model so findings of no between-studies variance might be overstated. If situational contingencies such as contextual factors, individual differences, and treatment implementations do cause real differences in correlations, use of the fixed-effects model is open to serious challenge. We present statistical evidence of the potential consequences of a failure to include estimates of between-studies variance below.

Statistical Properties of the Fixed- and Random-effects Models.

Johnson, Mullen, and Salas (1995) analyze the efficacy of three methods of meta-analysis; the Hedges & Olkin (1985); Rosenthal & Rubin (1978, 1988; Rosenthal, 1991) and Hunter, et al. (1982) fixed-effects models. Across a series of simulations, Johnson, et al. (1995) found that the results of the Hedges & Olkin (1985) and Rosenthal & Rubin (1978, 1988; Rosenthal, 1991) models converged together and conformed to conventional statistical expectations. On the other hand, they found that the Hunter, et al. (1982) approach consistently produced results which violated conventional statistical expectations and diverged significantly from those of the other two models. We specify the reasons behind some of Johnson, et al.'s (1995) findings and then extend the Rosenthal and Rubin (1978, 1988; Rosenthal, 1991) model to include random-effects.

The population equation for variance due to sampling error is:

$$s_e^2 = \frac{\sum_{i=1}^{k}\left[Ni\left[\frac{(1-r_i^2)^2}{N_i - 1} \right]\right]}{\sum N_i} \qquad [6]$$

where *k* = the number of studies (Hunter & Schmidt, 1990:107). Hunter & Schmidt (1990:107) state that since "... the average (ρ^2) - (average $\rho)^2$," Equation 7 can be estimated by:

$$s_e^2 = \frac{(1-\bar{r}^2)^2 k}{\sum N_i} \qquad [7]$$

However, it is well known (e.g., Hardy, Littlewood & Polya, 1934) that, unless Var(ρ)=0, the average (ρ^2) > (average $\rho)^2$, that is $E(\rho^2) \geq (E(\rho))^2$, where E(.) denotes expectation (Hunter & Schmidt, 1990:169). The average (ρ^2) is equal to (average $\rho)^2$ only when Var(ρ) is zero. This can be seen by considering the formula for the variance of a random variable X:

$$\text{Var}(X) = E(X^2) - [E(X)]^2. \qquad [8]$$

Since variances are always positive, the only time that $E(X^2) - [E(X)]^2$ is when Var(X)=0. If we substitute ρ for X in Equation 8, we see that Equation 7 is an appropriate estimate of sampling error only when the true between-studies variance is zero. The result is that the fixed-effects model attributes all variance to sampling error because it lumps variance from both between- and within-studies sources together.

We illustrate the effects of lumping variances together by the example presented in Table 1 where the true population values, ρ, are known. The cases differ in terms of the amount of total variance in observed study correlations that is accounted for by true between-studies sources (τ^2) and within-studies or sampling error (σ_e^2) sources. For each case we have computed the average (ρ^2), the (average $\rho)^2$, τ^2, and the fixed-effects model estimates of σ_ρ^2, σ_r^2, and σ_e^2.

TABLE 1
Components of variance in study correlations
and fixed-effects method estimates

Case	% of variance from between-studies sources		Study #1 (n=100)	Study #2 (n=100)	Study #3 (n=50)	Study #4 (n=50)	True σ_e^2	True τ^2	$E(\rho^2)$	$(E(\rho))^2$	σ_ρ^2	σ_r^2	σ_e^2
Case 1	100%	ρ =	.34	.16	.12	.38	0	.01103	.0735	.0625	-.000214	.011033	.011877
		r_i =	.34	.16	.12	.38							
Case 2	50	ρ =	.32	.18	.17	.33	.005633	.0054	.0679	.0625	-.000214	.011033	.011877
		r_i =	.34	.16	.12	.38							
Case 3	25	ρ =	.30	.20	.19	.31	.00816	.00287	.0654	.0625	-.000214	.011033	.011877
		r_i =	.34	.12	.12	.38							
Case 4	0	ρ =	.25	.25	.25	.25	.011033	0	.0625	.0624	-.000214	.011033	.011877
		r_i =	.34	.16	.16	.34							

In Case 1, all of the variance in observed correlations is due to between-studies factors (τ^2= .01) sampling error is zero. We see that the $E(\rho^2)$ is larger than the $(E(\rho))^2$. It is clearly inappropriate to use the fixed-effects model here because it produces the opposite result; that between-studies variance (σ_ρ^2) is zero[2] and sampling error accounts for the total variance in observed correlations. In fact the fixed-effects estimate of sampling error (σ_e^2 = .012) is greater than the total variance (σ_e^2 = .011). In Case 4, all of the variance in observed correlations is due to sampling error. The fixed-effects model is appropriate here because the $E(\rho^2)$ equals the $(E(\rho))^2$. However, even when sampling error accounts for 75% of total variance (Case 3), the fixed-effects model produces erroneous results because the $E(\rho^2)$ is not equal to $(E(\rho))^2$. Although we can clearly see that the true between- and within-studies variance changes from case to case, all estimates produced by the fixed-effects method stay the same. Application of the fixed-effects model often results in an estimate of σ_e^2 (i.e., τ^2) that is close to zero regardless of its true size. As was the result here, *negative* estimates are even possible (Hunter & Schmidt, 1990). Thus, instead of estimating $\sigma_\rho^2 = \sigma_r^2 - \sigma_e^2$, the fixed-effects model is actually estimating $\sigma_\rho^2 = \sigma_r^2 - (\tau^2 + \sigma_e^2)$. Because it assumes no true between-studies variance exists and because it does not expressly estimate between-studies variance, the result is not surprising: the variance remaining after removing "sampling error" is close to (or even less than) zero[3]. In an actual meta-analysis the homogeneity assumption of the fixed-effects model is highly questionable because values of ρ and τ^2 are not a *priori* known.

Upon closer inspection, we see that Equation 7 is not an estimate of σ_e^2, sampling error, it is *k* times the maximum likelihood estimate[4] of σ_r^2, the variance of $\bar{r}$. Since it is a maximum likelihood estimate, it is a consistent estimate of *k* times σ_r^2. The values obtained from Equation 8 and the artifact correction method's estimate of σ_r^2 converge to the same value as the number of studies gets larger[5]. Application of the fixed-effects model results in the subtraction of one estimate of the variance of $\bar{r}$ from another estimate of the variance of $\bar{r}$. This explains both the

[2] In this case, the fixed-effects estimate of between-studies variance, of σ_ρ^2, is negative. We discuss the occurrence of negative variance estimates in the fixed-effects procedure in more detail below.

[3] Koslowsky and Sagie (1994) show via simulation that the fixed-effects estimate of sampling error usually accounts for between 80% and 100% of correctable artifact.

[4] $\hat{s}_r^2 = \dfrac{\sum N_i (r_i - \bar{r})^2}{\sum N_i}$

[5] Upon application of the weak law of large numbers it is *k* times the method of moments estimator of the variance of $\bar{r}$; Bickel & Doksum, 1977.

propensity for the fixed-effects model to produce near zero estimates of σ_ρ^2 and the occurrence of negative estimates. Because of these limitations in the fixed-effects model, we present a random-effects model which produces estimates of both within- and between-studies variance in the sections below. Furthermore, it allows those variance components to correctly influence estimates of the population correlation.

A Random-Effects Estimator of ρ

The following discussion is a procedural overview of the random-effects model. We provide the details and statistical derivation of a general random-effects model and a random-effects-with-covariates model in the appendix. Application of the proposed general random-effects model is very straightforward and results can even be calculated by hand. The iterative procedure for the model with covariates is easily adapted to many mathematical computer programs. The proposed random-effects procedures are further facilitated because corrections for sampling error and range restriction artifacts may be unnecessary. Such corrections are inherent in the procedure (Bryk & Raudenbush, 1992; Raudenbush, 1988).

A General Random-Effects Model

Combing Equations 1 and 2, the general random-effects model is:

$$r_i = \rho + u_i + e_i \,. \qquad [9]$$

The general model entails estimating three parameters: ρ, u_i, and e_i. The variance of the u_i's is τ^2 (from Equation 3). By expressly estimating between-studies variance, the random-effects model provides a rigorous method to statistically test the homogeneity assumption (i.e., the standard Cochran -χ^2 test of Ho: $\tau^2 = 0$; Cochran, 1937). Study characteristics which do produce real variation in population values are incorporated in the between-studies variance term, τ^2. In the random-effects model $\hat{r}$ is analogous to the grand mean in an ANOVA analysis. It represents the gross or overall population value after partialling out differences due to study characteristics. If a subset of study characteristics are important, differences in population correlations due to theoretically meaningful factors can also be estimated. This can be accomplished through a random-effects-with-covariates model which we present below.

The general random-effects method proceeds as follows. First, because they come from a skewed distribution, raw correlations are transformed to follow a normal distribution. Transforming the correlations normalizes their distribution and stabilizes their variance allowing us to use maximum likelihood estimation. Maximum likelihood estimators reach the Cramer-Rao lower bound and are therefore efficient and consistent (Bickel & Doksum, 1977). This

transformation is very straightforward. First, transform the individual correlations, r_i's, into r_i*'s using Hotelling's (1953) transformation:

$$r_i^* = [r_i - \frac{1}{2(N_i - 3)} r_i(1 - r_1^2)]. \quad [10]$$

Second, transform the r_i*'s using Fisher's z-transformation which constructs the z_i*'s. By transforming the raw correlations into z_i*'s they now follow a normal distribution with a known stable variance (Bobko, 1995; Fisher, 1946). The within-studies variance of the z_i*'s is $1/n_i$. The procedure continues by computing estimates of the population correlations and between-studies variance. The starting assumption is that between-studies variance or τ^2 is zero. Using this value, Equation 11 is solved.

$$w_i = \left[\frac{1}{n_i} + \hat{t}^2 \right]^{-1} \quad [11]$$

The obtained value for w; is used in Equation 12.

$$\hat{x} = \frac{\sum_{i=1}^{k} w_i z_i^*}{\sum_{i=1}^{k} w_i} \quad [12]$$

Next, $\hat{x}$ the from Equation 12 is used to solve Equation 13 for a new $\hat{t}^2$.

$$t^2 = \frac{\sum_{i=1}^{k} w_i^2 \left[(z_i^* - \hat{x})^2 - \frac{1}{n_i} \right]}{\sum_{i=1}^{k} w_i^2} \quad [13]$$

Continue exchanging $\hat{t}^2$ and $\hat{x}$ in Equations 12 and 13 until the difference in obtained estimates is very small (e.g., 10^{-10}). In various simulations we conducted, convergence was usually achieved in less than seven iterations. Finally, back-transform $\hat{x}$ into $\hat{r}$ using Equation 14.

$$\hat{r} = \tanh(\hat{x}) = \frac{e^{(\hat{x})} - e^{(-\hat{x})}}{e^{(\hat{x})} + e^{(-\hat{x})}} \quad [14]$$

This. model without covariates can be computed by hand or using a short computer program[6]. Confidence intervals of the estimated population correlation can be computed using Equation 15.

$$\hat{x} - 1.96\sum_{i=1}^{k}\left(\frac{1}{n_i} + \hat{t}^2\right)^{-1} \leq x \leq \hat{x} + 1.96\sum_{i=1}^{k}\left(\frac{1}{n_i} + \hat{t}^2\right)^{-1} \qquad [15]$$

This general model provides an estimate of the true population correlation, $\hat{x}$, and an overall estimate of the between-studies variance, $\hat{t}^2$. It also takes between-studies variance into account when estimating the population correlation. Researchers should use this simple model when they assume no important moderators exist. A model with covariates, which allows the researcher to test for causes of between-studies variance (moderators), operates much like a standard regression with indicator or dummy variables. The effects of theoretically important factors on study correlations can be tested using a standard t-test. The model to be estimated is then:

$$\xi_i = \mathbf{x}_i^T\beta + \delta_i \qquad \delta_i \sim N(0, \tau^2)$$

Because of the complex iterative calculations, the model with covariates requires a computer program. As an example, the participation literature suggests that the form of employee participation might influence its effects on job performance and satisfaction (Cotton, Vollrath, Lengnick-Hall, & Froggatt, 1990; Leana, Locke, & Schweiger, 1990; Wagner, 1994). The covariates of interest here are different forms of participation and they can easily be included in a random-effects model. We present such an analysis later in the paper.

A number of statistical and practical advantages are offered by these random-effects estimators. First, because between-studies variation in effect sizes is accounted for by the model and is correctly allowed to influence parameter estimates, random-effects models are not constrained by the restrictive assumptions of the fixed-effects method. The general model also includes an estimate of the variance due to between-studies characteristics (i.e., τ^2). Operating like the residual in a standard ANOVA or ordinary least squares regression (OLS), τ^2 captures the effect of study characteristics not expressly included in the analysis. The model with covariates can be expanded to include a number of factors which represent theoretically meaningful study differences (similar to the moderators proposed by Hunter & Schmidt, 1990), allowing theory testing about the importance of these differences. Another advantage of the random-effects procedure is that confidence intervals, which offer a superior approach to

[6] A program written in S-Plus for Windows is available from the authors upon request. It can be easily altered for other mathematical programs.

hypothesis testing (Bakan, 1966; Cohen, 1994; Lykken, 1968, Rozeboom, 1962), can be computed for the parameter estimates. These procedures also preclude adjusting for sampling error because such adjustments are inherent to the estimation procedure. Finally, the estimator of t can be easily re-expressed in p units through reverse transformation. This allows the researcher to discuss the results of the meta-analysis in correlational terms.

Comparing Parameter Estimates from the Fixed and Random Effects Models

Although it is always advisable to use proper methods of estimation, statistical differences in the accuracy and correctness of methods do not always lead to practical differences in estimations. We also recognize that when τ^2 is likely to be different from zero, careful researchers would not use the fixed-effects approach. We conducted the following simulation study to demonstrate the practical consequences of misapplication of the fixed-effects method when between-studies differences are significant. In this simulation, the true population correlation, between-studies variance, and within-studies variance were set in advance. This allowed us to compare the accuracy of random- and fixed-effects estimation as values of τ^2 change.

Method

Data sets were created by assigning the number of studies (*k*), the true population correlation (ρ), and the between-studies variance (τ^2) for each data set. Then, a random numbers generator was used to create each of the *k* studies comprising a data set. The sample size of each study (N_i) was allowed to vary uniformly from 40 to 80. For each study in the data set, the computer produced an observed correlation and a sample size. Next, based on the data set of k studies created, we estimated the population correlation ($\hat{r}$), the between-studies variance ($\hat{t}^2$), and the standard error of the point estimates (SE($\bar{r}$) and SE($\hat{r}$)) using the fixed-effects (Hunter & Schmidt, 1990) and random-effects procedures. In addition, for each point estimate obtained by the two methods, we constructed a confidence interval.

Results

Table 2 presents the results of the simulation. With respect to the population correlation estimates, although neither method produced exactly precise results, the random-effects model produced better estimates of the true population correlations in all cases. Moreover, as the between-studies variance increased, the random-effects parameter estimates remained stable while those of the fixed-effects model became less and less accurate. This results are depicted in Figures 1 & 2. The fines are kernel smoothed (Hardle, 1990) summaries of all the estimates, while the points denoted by "X" and "+" are the point estimates.

TABLE 2
Comparison of parameter estimates from the random-effects and fixed-effects methods

				Fixed-effects Method		Random-effects Method		
Case	True population correlation (ρ)	True between-studies variance (τ^2)	Number of studies	Average study correlation ($\bar{r}$)	Standard error of the point estimate (SE($\bar{r}$))	the point $\hat{r}$	$\hat{t}^2$	Standard error of estimate (SE($\hat{r}$))
Case 1	.2	.1	75	.16	.002	.16	.08	.04
Case 2	.2	.6	75	.15	.002	.17	.50	.08
Case 3	.2	.1	100	.18	.001	.20	.09	.03
Case 4	.2	.6	100	.11	.001	.17	.61	.08
Case 5	.4	.1	75	.35	.001	.40	.10	.04
Case 6	.4	.6	75	.24	.002	.32	.63	.09
Case 7	.6	.1	75	.56	.001	.60	.09	.03

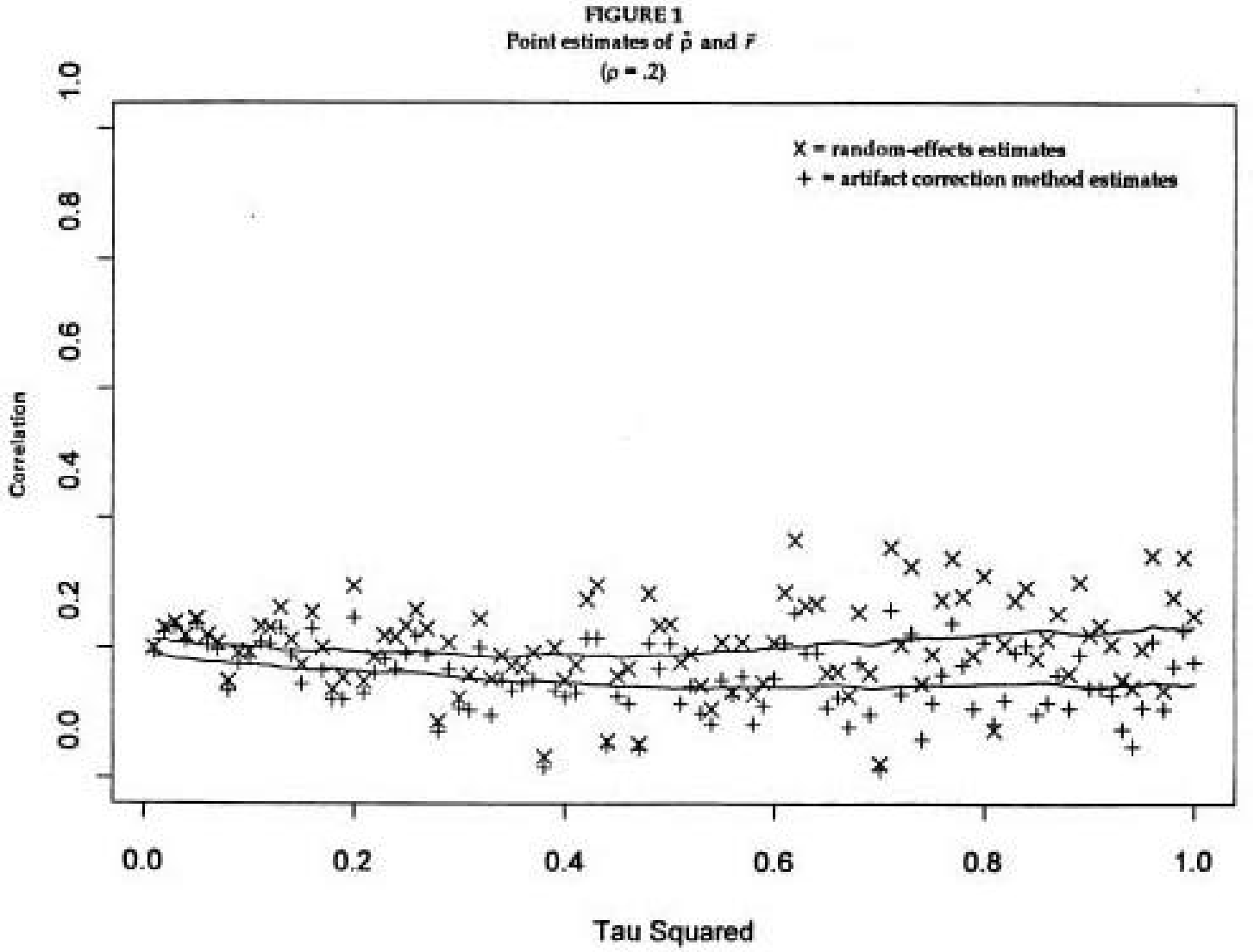

Figure 1. Point estimates of $\hat{r}$ and $\bar{r}$ (ρ = .2)

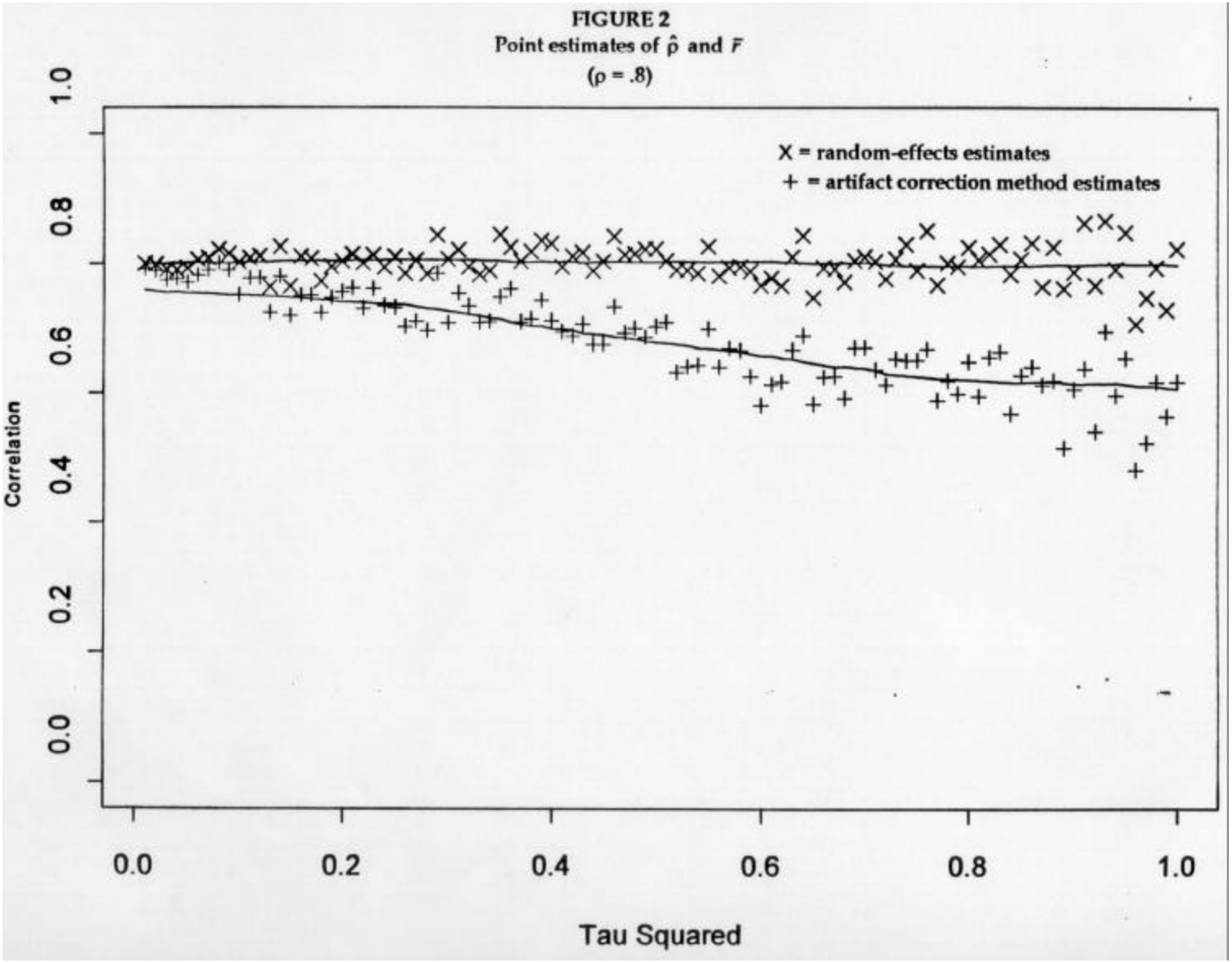

Figure 2. Point estimates of $\hat{r}$ and $\bar{r}$ (ρ = .8)

In all cases presented in Table 2, the estimates of between-studies variance produced by the random-effects model are very close to the true values of τ^2. Moreover, the confidence intervals estimated by the fixed-effects model are much too narrow. This is no doubt due to the fact that the between-studies variance is not accounted for. In fact, much can be learned about differences between the accuracy of the methods by focusing on the confidence intervals. Figures 3, 4, and 5 plot the kernel smoothed point estimates and both upper and lower confidence limits. This smoothing again makes it easier to see the difference in the procedures. First, as Figures 3, 4, and 5 show, the fixed-effects model neglects to estimate between-studies variance and, hence, the confidence intervals are very narrow. Even as τ^2 increases, the confidence intervals of the fixed-effects model remain constant. On the other hand, the random-effects confidence intervals appropriately become wider as τ^2 increases; as true between-studies differences increase, the confidence intervals become wider to account for these differences. Finally, comparing Figures 3, 4, and 5 reveals that, as the true correlation coefficient increases from .2 to .8, the confidence intervals of the fixed-effects model become even narrower and fail to include the true population parameter, ρ. Thus, as correlations become stronger, confidence intervals of the fixed-effects model become narrower and the point estimates become worse (as Figures 1 & 2 also demonstrate). Indeed, Figures 3, 4, and 5 indicate that the confidence intervals of the fixed effects model degrade as p increases. On the other hand, the random-effects confidence intervals only worsen slightly as the value of the true correlations increase. In sum, the results of this simulation indicate that the random-effects model produces more precise estimates of the population correlation and much more accurate confidence intervals. The results of this simulation confirm the findings of Johnson, et al. (1995).

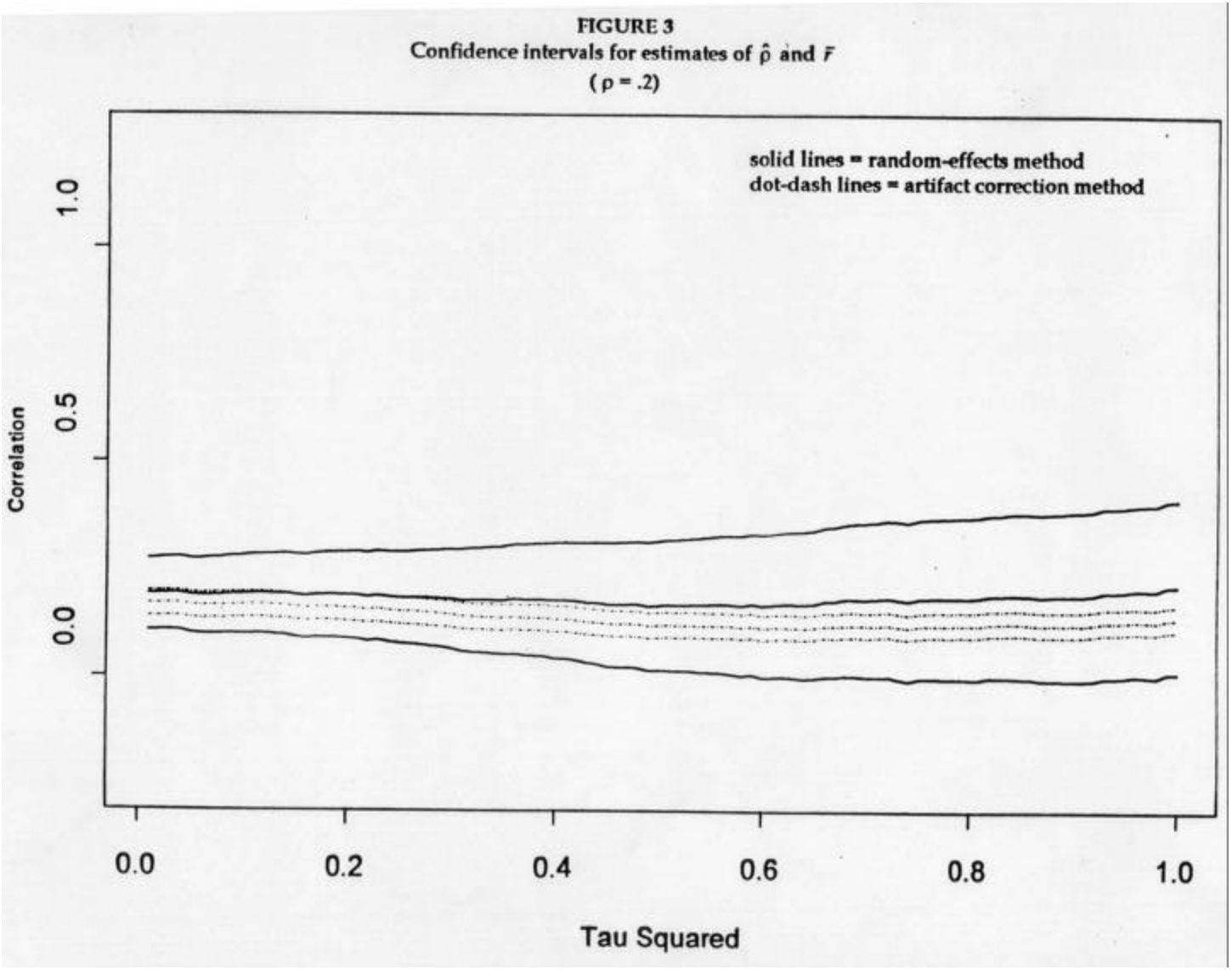

Figure 3. Confidence intervals for estimates of $\hat{r}$ and $\bar{r}$ (ρ = .2)

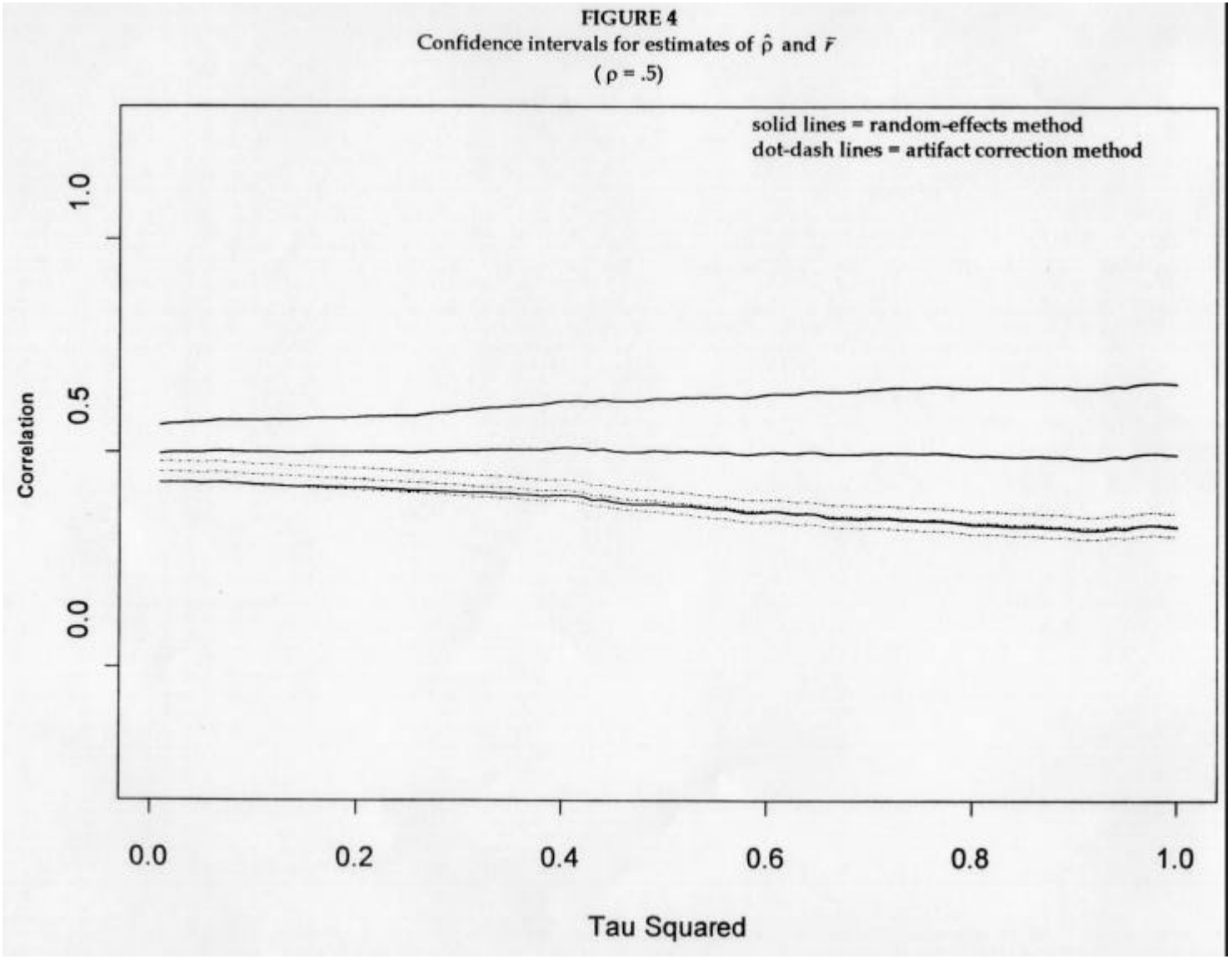

Figure 4. Confidence intervals for estimates of $\hat{r}$ and $\bar{r}$ (ρ = .5)

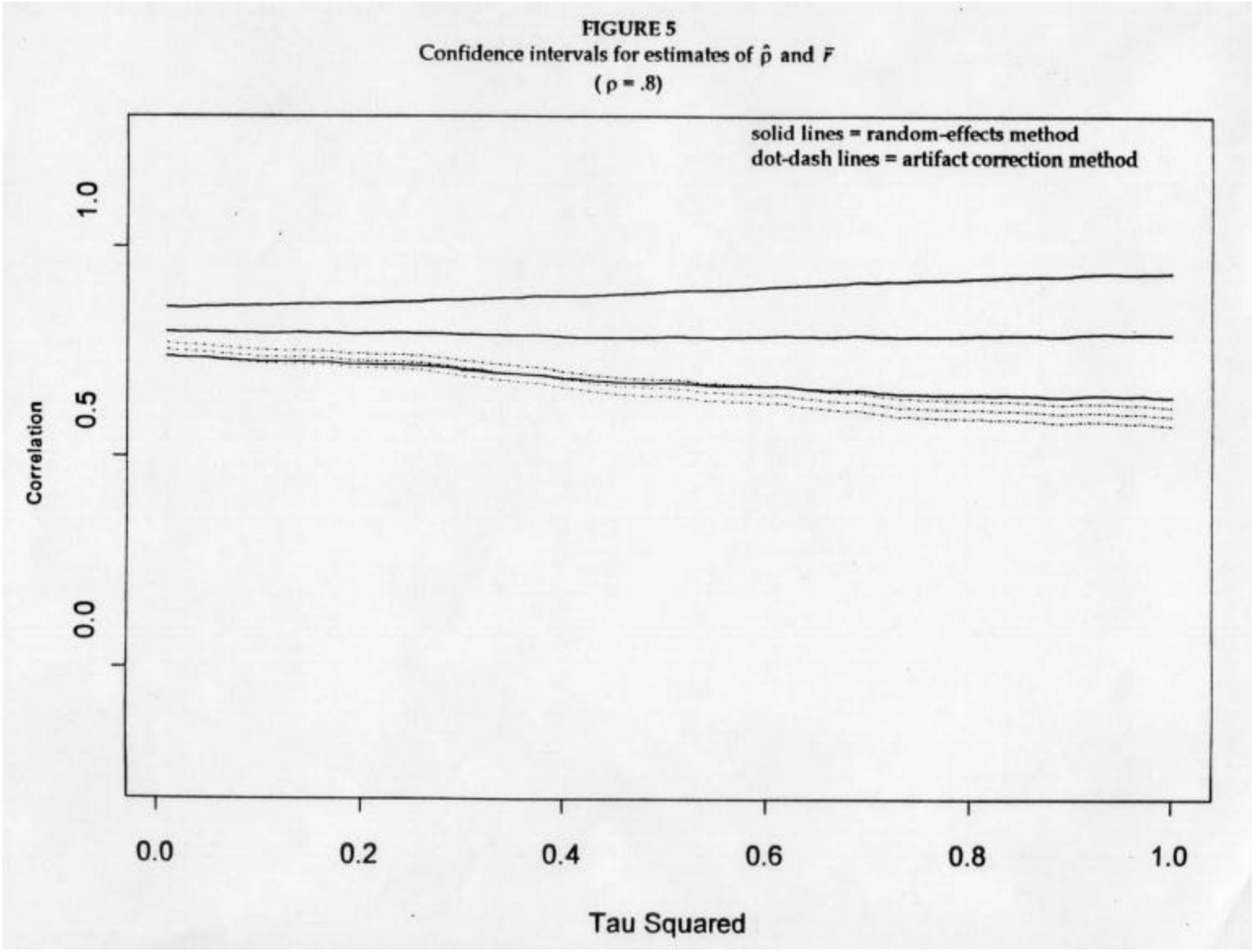

Figure 5. Confidence intervals for estimates of $\hat{r}$ and $\bar{r}$ (ρ = .8)

The random-effects model can be extended to include covariates that account for between-studies variation. We simulated meta-analytic data sets where τ^2 was caused by industry differences in the sample of organizations from which data for the studies were drawn. For convenience, we assumed the organizations came from two different industries with different population correlations. We used a *k* of 60, allowing the sample size to vary uniformly between 40 and 80. We estimated the pooled or general random-effects model in addition to the model with second-level covariates for the sub-groups. Table 3 presents the data and summary results. It is clear the grand mean or overall population parameter estimate, ρ, is in between the values for the two sub-populations. The importance of including meaningful covariates can be seen in terms of the estimates of ρ and τ^2. When a second-level covariate denoting industry was included, estimates of these sub-population parameters were obtained which are very close to the true sub-population values. This second-level model can be extended to include multiple covariates and to model more complex relationships. Furthermore, this model avoids the problems of conducting two, separate meta-analysis as would be prescribed by the fixed-effects model.

TABLE 3
Results of including second-level covariates in a random-effects model

Simulation	True population correlation for industry #1	True population correlation for industry #2	True τ^2	$\hat{r}$ for industry #1	$\hat{r}$ for industry #2	$\hat{t}^2$	Pooled or `grand mean' estimate of $\hat{r}$	pooled estimate of $\hat{t}^2$
Simulation #1	.3	.7	0	.31	.70	.0001	.57	.07
Simulation #2	.3	.7	.l	.31	.71	.07	.58	.14
Simulation #3	.3	.7	.2	.27	.71	.20	.57	.29
Simulation #4	.3	.7	.3	.26	.69	.33	.55	.36
Simulation #5	.3	.7	.4	.32	.70	.40	.57	.48

To further illustrate the random-effects-with-covariates model we analyzed the meta-analytic data on participation presented in Wagner (1994). Wagner (1994) was addressing the debate about whether the form of participation was important for explaining its effects. Cotton, et al, (1990) argued that different forms of participation would have different effects on employee performance and job satisfaction. Leana, et al., (1990) argued the form did not matter; participation's effects on performance and satisfaction were negligible. Wagner (1994) reviewed data from existing studies again. We only used data on the participation-job performance relationship. Wagner (1994) used the fixed-effects model of Hunter and Schmidt (1990) to compute the population correlation between participation and job performance for six different forms of employee participation. He broke the sample into six sub-samples - one for each form of participation -and found low correlations across these sub-groups. Wagner (1994) concluded that the form of participation did not matter. However, he notes several limitations of the fixed-effects procedure including lack of a precise test of differences among subgroups means and the relatively small sample size caused by the need to conduct six subgroup meta-analyses. The random-effects-with-covariates model avoids both of these limitations. Separate meta-analyses are not necessary because the form of participation can be included as a series of indicators variables. The coefficients obtained for these covariates allow the use of the usual t-test that the parameter estimate is zero to determine if statistically significant differences between forms of participation exist. The results of the random-effects with covariates analysis are presented in Table 4.

TABLE 4
Analysis of participation data from Wagner (1995) using a random-effects model

Factor	$\hat{x}$	s.e.	t-test
Intercept or grand mean correlation	$\hat{x}_0$ = .12	.08	1.40
Participation in work decisions	$\hat{x}_1$ = .16	.08	1.89
Consultative participation	$\hat{x}_2$ = .002	.11	.02
Informal Participation	$\hat{x}_3$ = .20	.09	2.26*
Employee Ownership	$\hat{x}_4$ = .08	.13	.61
Representative participation	$\hat{x}_5$ = .11	.10	1.31

Note: *p < .05 (two-tailed test)

Similar to an ordinary regression, we omitted one of the six indicator variables to avoid colinearity. The omitted form of participation is what Wagner (1994) calls short-term participation. Only one indicator coefficient was significant, the coefficient for informal participation ($\hat{x}_3$= .20, p<.05) suggests a difference between short-term and informal participation. After back-translation, the estimated correlation for informal participation and performance is .35, for short-term participation and performance it is .14. Analyses of this sort could be conducted in many other meta-analytic situations. Although we present this analysis for illustrative purposes only, it leads us to conclusions different from those of Wagner (1994); the form of participation may matter if one is comparing short-term and informal participation.

Discussion

Perhaps the most valuable contribution of the random-effects model is the discipline it imposes on the process of meta-analysis. Rather than assuming away potentially important differences between studies, the random-effects model provides a more rigorous means of

modeling and then testing for their existence and impact on a series of studies. The meta-analyst specifies a theoretically-based model of the relationships under study and then uses statistical analyses to test that model. This is the common approach used by psychologists in regression analyses. The fixed-effects approach, on the other hand, assumes a single model fits all meta-analytic situations.

One of the fundamental differences between random- and fixed-effects approaches is their divergent perspective on the situational specificity hypothesis. This hypothesis contends that there may be real differences between studies that cause coefficients to vary. Whereas the random-effects model explicitly tests the validity of this hypothesis, the fixed-effects model just assumes it away. This is similar to the choice of covariate modeling versus $\bar{Y}$ to explain a regression situation. The fixed-effects model asserts that $\bar{r}$ is the best estimator and attempts to explain away the sampling error. The random-effects approach assumes that, at a minimum, $\bar{r}$ is not the best estimator because unexplained causes of variance in the r_i's exist. The situational specificity hypothesis is sometimes misinterpreted as implying that, since parameter estimates vary from setting to setting, results cannot be generalized from study to study. The fixed-effects meta-analytic framework of Hunter and Schmidt (1990) is directed against this view. These authors claim that, in most cases, study results can be combined in order to achieve a single meaningful parameter. We fully agree with this conclusion. However, the fixed-effects model goes a step beyond by presuming that differences between studies usually do not exist or are not meaningful. As we have demonstrated, this conclusion can lead to erroneous results when taken as an *a priori* assumption. Surprisingly, when between-studies differences do exist, the fixed-effects model appears to accept the hypothesis that an overall parameter is unobtainable due to the heterogeneity of study groups. It suggests that studies be divided into more homogeneous subgroups and a separate meta-analysis conducted on each sub-group. Consequently, parameter estimates cannot be generalized across these sub-groups. The random-effects model asserts the opposite. Rather than breaking studies into separate meta-analyses, real differences between studies are accounted for in the model and the estimate of an overall parameter is meaningful. This approach is analogous to including an indicator variable in an OLS regression to account for group membership (e.g., male-female) rather than computing a separate regression for each group. Furthermore, these sub-group differences can be explicitly modeled through second-level covariates.

Many of these advantages derive from the generalizability of random-effects results to a super-population of all hypothetical studies on a relationship of interest (fixed-effects results can only be generalized to *the studies included in the meta-analysis;* Bryk & Raudenbush, 1992;

Hedges & Olkin, 1987; National Research Council, 1992). By providing a theoretically and statistically strong base, the general random-effects model introduced here can be extended into a family of hierarchical linear models capable of incorporating a variety of research problems. The random-effects-with-covariates model is one such extension. Others include modeling effects at three levels, say individual employee, firm, and industry. For example, a meta-analysis in organizational research might nest studies within industries and then nest industries within geographic regions. Such an analysis could be conducted using a hierarchical random-effects model (Bryk & Raudenbush, 1992). These models also extend our ability to draw out information which is important for understanding the nature and function of psychological constructs. These models can effectively combine information from seemingly dissimilar studies to provide answers to sophisticated research questions. For example, DuMouchel and Harris (1983) used cancer data from different studies on a variety of mammalian species and carcinogenic agents to gain a better understanding of the carcinogenic effects of a particular diesel fuel on humans even though *no human data were available.* This explanatory power 'borrows strength' by drawing out important information from a group of studies which cannot be obtained from the individual studies nor from a fixed-effects approach. Applied psychologists can gain considerable advantages in investigating complex relationships using the approaches outlined here. Random-effects models can also be easily extended to include a wide variety of study parameters in addition to correlations. Bryk and Raudenbush (1992) give an overview of the general models and the required transformations for continuous, dichotomous, and logistic effect sizes, among others.

Random-effects procedures shed light on meaningful sub-populations or sub-groups in the data that have different population correlations. By separating variability into within- and between-groups components, random-effects models draw attention to factors accounting for differences between sub-populations. The factors contributing to between-studies variance can be identified and modeled by random-effects procedures so that analysts can then test hypothesis about cross-level effects. In addition to improving parameter estimates, including second-level factors to link parameters from different levels efficiently utilizes information from groups to detect hierarchical structures (i.e., sub-populations) in the data. These procedures for diagnosing and testing sub-group effects provide potentially powerful exploratory techniques useful in many theory-testing situations.

Conclusion

The random-effects approach we have outlined here is a general form of a family of models based upon hierarchical linear modeling theory. It is extremely flexible, allowing a number of theoretical specifications to be incorporated and tested. These extensions include diagnosing sub-population effects and hypothesis testing of theoretically meaningful covariates. A variety of effect sizes, in addition to correlations, can be used including continuous and dichotomous variables (Bryk & Raudenbush, 1992). Owing to the theoretical and methodological problems associated with the fixed-effects model which were discussed throughout this paper, we believe that future meta-analyses should be conducted using the random-effects model. In particular, the a priori assumption that between-studies differences are zero or unimportant is tenuous. We recommend that potential differences should be explicitly modeled and tested. If the potential for theoretically relevant between-study differences exists in past meta-analysis, researchers may wish to apply the random-effects model to check their results for the effects of the homogeneity assumption.

Despite the methodological problems outlined herein, the message brought to the field of applied psychology by Hunter and Schmidt (1990) is invaluable - results from different studies can be combined and a generalizable coefficient can be obtained. Advances in meta-analytic modeling now allow meta-analysts to go beyond the estimation of the basic ‘grand mean' population effect by developing and testing richer, more complex models. Such models offer the promise of contributing extensively to our understanding of psychological constructs and relationships.

References

Bakan, D. (1966). The test of significance in psychological research, *Psychological Bulletin,* 66, 423-437.

Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics,* San Francisco: HoldenDay, Inc.

Birge, R. T. (1932). Probable values *of e, h, elm,* and *a. Physical Review,* 40, 228-261.

Bobko, P. (1995). *Correlation and regression: Principles and applications for industrial psychology and management,* New York: McGraw Hill.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods,* Newbury Park, CA: Sage Publications

Campbell, J. T. (1990). The role *of* theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.). *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 39-73), Palo Alto, CA: Consulting Psychologists Press.

Cochran, W. G. (1937). Problems arising in the analysis *of* a series *of* similar experiments. *Journal of the Royal Statistical Society, Supplement,* 4,102-118.

Cochran, W. G., & Yates, F. (1938). The analysis *of* groups *of* experiments. *Journal of Agricultural Science,* 28, 556-580.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist,* 49, 997-1003.

Cotton, J. L., Vollrath, D. A., Lengnick-Hall, M. L., & Froggatt, K. L. (1990). Fact: The form *of* participation does matter - A rebuttal to Leana, Locke, and Schweiger. *Academy of Management Review,* 15, 147-153.

David, F. N. (1938). *Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples.* Cambridge: Cambridge University Press.

Devlin, R. L., Gnanadesikan, R., & Kettering, J. R. (1976). Some multivariate applications *of* elliptical distributions. In S. Ikeda (Ed.), *Essays in probability and statistics,* Tokyo: Shinko Tsusho.

DuMouchel, W. H., & Harris, J. E. (1983). Bayes methods for combining the results *of* cancer studies in human and other species. *Journal of the American Statistical Association,* 78, 293-315.

Fisher, R. A. (1932). *Statistical methods for research workers,* London: Oliver & Boyd.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3-8.

Gauss, C. F. (1809). *Theoria motus corporum celestium.* Hamburg: Perthes et Besser.

Handle, W. (1990). *Applied nonparametric regression,* Cambridge: Cambridge University Press.

Hardy, G. H., Littlewood, J. E. & P61ya, G. (1934). *Inequalities,* London: Oxford University Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods of meta-analysis.* Orlando, FL: Academic Press.

Hotelling, H. (1953). New light on the correlation coefficient and its transformation. *Journal of the Royal Statistical Society, Series B, 15,* 193-225.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis,* Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Johnson, B.T., Mullen, B., & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology, 80,* 94-106.

Koslowsky, M., & Sagie, A. (1994). Components of artifactual variance in meta-analytic research. *Personnel Psychology, 47,* 561-574.

Leana, C. R., Locke, E. A., & Schweiger, D. M. (1990). Fact and fiction in analyzing research on participative decision making: A critique of Cotton, Vollrath, Froggatt, Lengnick-Hall, and Jennings. *Academy of Management Review, 15,* 137-146.

Legendre, D. T. (1805). *Novelles method a l determination des orbites des cometes.* Paris: Courcier.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70,* 151-159.

Mathieu, J. E., & Zajac, . M. (1990). A review and meta-analysis of the antecedent, correlates, and consequences of organizational commitment. *Psychological Bulletin,* 108, 171-194.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports,* 66, 195-244.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, D.C.: National academy Press.

Nunnaly, J. C. (1978). *Psychometric theory,* New York: McGraw-Hill.

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics,* 13, 85-116.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. Ed.), Beverly Hills, CA: Sage.

Rosenthal, R., & Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences,* 3, 377-415.

Rosenthal, R., & Rubin, D. (1988). Comment: Assumptions and procedures in the file drawer problem. *Statistical Science,* 3, 120-125.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test, *Psychological Bulletin,* 57, 416-428.

Silver, N., & Dunlop, W. (1987). Averaging coefficients: Should Fisher's z-transformation be used? *Journal of Applied Psychology,* 72, 146-148.

Strube, M. (1988). Averaging correlation coefficients: Influence of heterogeneity and set size. *Journal of Applied Psychology,* 73, 559-569.

Trippett, L. C. (1931). *The method of statistics.* London: Williams & Norgate.

Wagner, J. A., III. (1994). Participation's effects on performance and satisfaction: A reconsideration of research evidence. *Academy of Management Review,* 19, 312 330.

APPENDIX

Because we refer to equations contained in the body of the paper, we continue numbering appendix equations consecutively with the body. Rather than working in the natural parameterization of Equations 1 and 2, we prefer to apply Fisher's z-transformation to both the sample and population correlations (Fisher, 1932). Such a transformation is known as a variance stabilizing transformation and the use of such transformations is common place in statistical practice (Bickel & Doksum, 1977:221). It is well known that the sample correlation is 1) a biased estimate of the true correlation, 2) has a variance which depends on the true correlation, and 3) has a skewed distribution for large |ρ| (see Figure 3.3 in Bobko, 1995:51). Therefore, a model such as Equation 1 may not be realistic. Bobko (1995:51) notes that, because of the skewness in the sampling distribution of the sample correlations, averaging untransformed correlations can lead to an under-estimation of the true average. (See Silver and Dunlop, 1987 and Strube, 1988 for computer simulations of this observation.) Specifically, if r is a correlation based on a sample of N = n + 1 and p is the true correlation:

$$E(r) = r - \frac{r(1-r^2)}{2n} + O(n^{-2})$$

$$Var(r) = \frac{(1-r^2)^2}{n} + O(n^{-2}).$$

The symbol $O(n^{-\alpha})$ denotes an error of approximation e(ρ) term such that $e(\rho)n^{\alpha}$ is bounded by some constant as n tends toward infinity, for instance $1/n^2 = O(n^{-2})$ That is, $O(n^{-2})$ gives a rate at which e(ρ) tends to zero; as *n* increases the value of $O(n^{-2})$ decreases to zero rapidly. Upon applying the z-transformation

$$z = a\tanh(r) = \frac{1}{2}\log_e\left(\frac{1=r}{1-r}\right)$$

$$x = a\tanh(r) = \frac{1}{2}\left(\frac{1+r}{1-r}\right),$$

where *atanh(.) is* the arc-hyperbolic tangent function. It follows that

$$E(r) = x + \frac{r}{2n} + O(n^{-2})$$

$$Var(r) = \frac{1}{n} + O(n^{-2}).$$

Therefore, the variance has been stabilized. A further transformation will reduce the bias remaining in the z-transformed sample correlations. Define $z^* = r[-\frac{1}{2n} r(1-r^2)]$. In this case, it can be shown that

$$E(z^*) = x + O(n^{-2})$$

$$Var(z^*) = \frac{1}{n} + O(n^{-2}) .$$

For more details on these transformation see Hotelling, 1953). Furthermore, the probability density function of z* looks nearly normal, particularly for N $\geq$ 25. Devlin, Gnanadesikan, and Kettenring (1976) and Bobko (1995) make the recommendation that one can treat the transformed correlations as if they were normally distributed with a mean equal to zero and variance equal to 1/(N - 3). See Fisher (1932:201) and David (1938) for more details on the approximate normality.

Upon taking these transformed results into account, it is quite clear why Equation 4 may be an inadequate model. We prefer to model the relationships in the transformed scale. If r_i, ρ^i (i=1,...,k), and ρ denote the sample, individual population, and overall population correlations, respectively, define

$$z_i^* = a \tanh[r_i - \frac{1}{2n_1} r_i (1-r_i^2)]$$

$$x_i = a \tanh[r_i]$$

$$x = a \tanh[r] .$$

for $1 \leq i \leq k$. Then we propose the model:

$$z_i^* = x_i + e_i \qquad \varepsilon_i \sim N(0, \frac{1}{2n_i}) \qquad i = 1,\ldots,k \qquad [16]$$

$$\xi_I = \xi + \delta_i \qquad \delta_i \sim N(0,\tau^2) \quad i = 1,\ldots,k \qquad [17]$$

where $n_i = N_i - 3$. This is a general model because potential differences between studies are accounted for via a general random effect, δ_1. Under the model defined in Equations 16 and 17, it *k* follows that the marginal distributions of the are independent normals with a mean equal to ξ and a variance equal to . Therefore, the likelihood function for (ξ, τ^2) is then equal to

$$f(\xi,\tau^2) = (2p)^{-\frac{k}{2}} \left[\prod_{i=1}^{k} \left(\frac{1}{n_i} + t^2\right)\right]^{-\frac{1}{2}} \exp\left[\frac{-\frac{1}{2}\sum_{i=1}^{k}(z_i^* - x)^2}{(n_i + t^2)}\right]. \qquad [18]$$

It follows from this likelihood function that the maximum likelihood estimators ($\hat{x}$, $\hat{t}^2$) of (ξ, τ^2) are solutions to the equations

$$\hat{x} = \frac{\sum_{i=1}^{k} w_i z_i *}{\sum_{i=1}^{k} w_i} \qquad [19]$$

$$\hat{t} = \frac{\sum_{i=1}^{k} w_i^2 \left[(z_i * - \hat{x})^2 - \frac{1}{n_i} \right]}{\sum_{i=1}^{k} w_i^2}, \qquad [20]$$

where $w_i = \left[\frac{1}{n_i} + \hat{t}^2 \right]^{-1}$. It is easy to solve these equations via iterations. Start at a value, say $\tau^2 = 0$, solve Equation 19, then Equation 20. Given a new value of $\hat{t}^2$, solve Equation 18 again, then compute a new $\hat{t}^2$. Continue this process until there is only a slight difference between successive estimates (say, 10^{-10}). In various simulations we conducted, convergence was usually achieved in less than seven iterations.

Using the usual large sample properties of maximum likelihood estimators and Fisher information arguments it can be shown that the estimate of $\hat{x}$ is approximately normal, at ξ with a variance equal to $\left[\sum_{i=1}^{k} \left(\frac{1}{n_i} + \hat{t}^2 \right)^{-1} \right]^{-1}$. These results make it possible to construct an approximate pivot as to yield the 95% confidence interval for ξ:

$$\hat{x} - 1.96 \sum_{i=1}^{k} \left(\frac{1}{n_i} + \hat{t}^2 \right)^{-1} \leq \xi \leq \hat{x} + 1.96 \sum_{i=1}^{k} \left(\frac{1}{n_i} + \hat{t}^2 \right)^{-1}. \qquad [21]$$

One usually wishes to recover the true mean population correlation rather than the transformed version. In this case define:

$$\hat{r} = \tanh(\hat{x}) = \frac{\exp(\hat{x}) - \exp(-\hat{x})}{\exp(\hat{x}) + \exp(-\hat{x})} \qquad [22]$$

as the back-transformed point estimate of ρ. This is also the maximum likelihood estimator of ρ due to the invariance of the maximum likelihood estimation technique (see Bickel & Doksum,

1977:141). As *tanh(.) is* a monotone function one may apply it to all three parts of Equation 21 and still maintain the inequality in addition to the implied confidence statement. Consequently, the 95% confidence interval for ρ is

$$\tanh\left[\hat{x} - 1.96\sum_{i=1}^{k}\left(\frac{1}{n_i} + \hat{t}^2\right)^{-1}\right] \leq r \leq \tanh\left[\hat{x} + 1.96\sum_{i=1}^{k}\left(\frac{1}{n_i} + \hat{t}^2\right)^{-1}\right].$$

A Random-Effects with Covariates Model

Now, consider the model

$$\xi_i = \mathbf{x_i}^T\boldsymbol{\beta} + \delta_i \quad \delta_i \sim N(o,\tau^2)$$

where $\mathbf{x_i}^T$ is a vector of covariates for the i^{th} individual under study. Again, the estimation is based on the marginal density of the z_i*'s, which, since the z_i* have independent normal distributions with mean and variance equal to $\mathbf{x_i}^T\boldsymbol{\beta}$ and $\left(\frac{1}{n_i} + t^2\right)$, respectively, is given by

$$\left(\prod_{i=1}^{k}\left[2p\left(\frac{1}{n_i} + t^2\right)\right]^{-\frac{1}{2}}\right)\exp\left\{\left[-\frac{1}{2}\sum_{i=1}^{k}\left(z_i * - x_i^T b\right)^2\right] / \left(\frac{1}{n_i} + t^2\right)\right\}. \qquad [23]$$

Following the previous approach, we can estimate β and $\hat{t}^2$ by differentiating Equation 16 with respect to β and $\hat{t}^2$, and setting the equations equal to zero. Letting $\hat{b}$ and $\hat{t}^2$ denote the likelihood estimates, the estimates obtained can be expressed as

$$\hat{b} = \left(X^T W^{-1} X\right)^{-1}\left(X^T W^{-1} z *\right) \qquad [24]$$

where z* = (z_l*, ..., z_k*)^T, X is the *k* x *l* matrix with rows x_i^T, and W is the *k* x *k* diagonal matrix with diagonal element $W_{ii} = \frac{1}{n + \hat{t}^2}$ and

$$\hat{t} = \frac{\sum\left[\left\{z_i * - x^T\hat{b}\right\}^2 - \frac{1}{n_i}\right] / \left[\frac{1}{n_i} + \hat{t}\right]^2}{\sum_{i=1}^{k}\left(\frac{1}{n_i} + \hat{t}^2\right)^{-2}} \qquad [25]$$

As before, there are no single closed form expressions for β and τ^2. However, one can use the iterative scheme previously discussed.