Journal of Applied Psychology
2007, Vol. 92, No. 2, 297–308

0021-9010/07/$12.00 DOI: 10.1037/0021-9010.92.2.297

Updating Meta-Analytic Research Findings: Bayesian Approaches Versus the Medical Model

Frank L. Schmidt
University of Iowa

Nambury S. Raju
Illinois Institute of Technology

The authors examine 3 methods of combining new studies into existing meta-analyses: (a) adding the new study or studies to the database and recalculating the meta-analysis (the medical model); (b) using the Bayesian procedure advocated by F. L. Schmidt and J. E. Hunter (1977) and F. L. Schmidt, J. E. Hunter, K. Pearlman, and G. S. Shane (1979) to update the meta-analysis; and (c) using the Bayesian methods advocated by these authors and M. T. Brannick (2001) and M. T. Brannick, S. M. Hall, and Y. Liu (2002) to estimate study-specific parameters. Method b was found to severely overweight new studies relative to the previous studies contained in the meta-analysis, and Method c was found to do the same while also requiring an assumption with a low prior probability of being correct, causing the method to violate Bayesian principles. The authors present an alternative Bayesian procedure that does not suffer from these drawbacks and yields meta-analytic results very similar to those obtained with the medical model. They recommend use of the medical model or this alternative Bayesian procedure.

Keywords: meta-analysis, Bayesian methods, cumulative knowledge, research synthesis

In recent years, meta-analysis has assumed increased importance as research conclusions in most areas have come to be based on meta-analytic results rather than on individual empirical studies or narrative reviews of individual studies. In medicine, psychology, the social sciences, and many other areas, the conclusions presented in textbooks, both introductory and advanced, are increasingly based on meta-analyses (Hunt, 1997). The same is true of the literature reviews that are used to set the stage for individual empirical studies (Hunter & Schmidt, 2004). But because new studies are constantly being conducted, it is apparent that few, if any, meta-analyses are ever the final word. This fact raises the question of how meta-analyses should be updated as new studies become available. Two approaches have been suggested to do this. In the first approach (referred to below as the *medical model*), the meta-analysis is rerun including the newly available study (or studies). The second approach calls for new studies to be incorporated by means of a particular procedure for Bayesian information integration. Although there are a number of different possible Bayesian approaches, the focus of this article is on the procedure most prominent to date in the industrial/organizational literature—that is, the procedure advocated originally by Schmidt and Hunter (1977) and Schmidt, Hunter, Pearlman, and Shane (1979) and more recently by Brannick (2001; Brannick & Hall, 2003; Brannick, Hall, & Liu, 2002). This article examines the properties and the advantages and disadvantages of the medical model and this Bayesian model. It also presents an alternative Bayesian procedure that, unlike the earlier Bayesian procedure, yields meta-analytic results very similar to the results from the medical model.

Frank L. Schmidt, Department of Management and Organizations, Henry B. Tippie College of Business, University of Iowa; Nambury S. Raju, Department of Psychology, Illinois Institute of Technology.

Nambury S. Raju passed away on October 27, 2005.

Correspondence concerning this article should be addressed to Frank L. Schmidt, Department of Management and Organizations, Henry B. Tippie College of Business, University of Iowa, Iowa City, IA 52242. E-mail: frank-schmidt@uiowa.edu

The Medical Model

The first approach is called the medical model because its earliest and most extensive use has been in the evaluation of the efficacy of medical treatments, including drugs. The Cochrane Collaboration is an organization of medical researchers and biostatisticians based in Britain whose purpose is to use meta-analysis to calibrate the absolute and relative effectiveness of different medical interventions for treating a wide variety of diseases (Hunt, 1997; Hunter & Schmidt, 2004, pp. 28–29). The results of these meta-analyses are made available worldwide to medical practitioners on the Internet (www.cochrane.org). For any given treatment for any specific disease, new studies (usually randomized controlled trials) become available with some frequency. When a new study appears, that study is added to the database, and the meta-analysis is rerun. As a result, the new study is weighted into the database in the same manner as all previous studies.

The medical model for updating meta-analysis results is the basis for sequential meta-analysis in medicine. Sequential meta-analysis was introduced by Antman, Lau, Kupelnick, Mosteller, and Chalmers (1992) to reveal the point in time by which enough information has become available to show conclusively that a medical treatment is effective (Hunt, 1997). For example, suppose that the first randomized controlled trial for a particular drug had been conducted in 1975 but had a wide confidence interval, one that included zero. Now suppose three more studies had been conducted by 1976, providing a total of four studies to be meta-analyzed, but the confidence interval for the meta-analytic mean of these studies was still wide enough to include zero. Now, suppose that five more randomized controlled trials had been conducted in 1977, providing nine studies for a meta-analysis up to this date,

and that meta-analysis yielded a confidence interval that excluded zero. The conclusion is then that, given the use of meta-analysis, enough information was already available in 1977 to begin the use of this drug with patients. On the basis of the meta-analysis result and available statistics on the disease, Antman et al. computed how many lives would have been saved to date had the use of the drug begun in 1977. They found that, across different treatments, diseases, and areas of medical practice, a large number of lives would have been saved had medical research historically used meta-analysis in this manner. Stated another way, their sequential meta-analyses showed that, across many diseases, a large number of lives had been lost because of the failure to apply meta-analysis to medical research literatures. The resulting article (Antman et al., 1992) is widely considered to be the most important and influential meta-analysis ever published in medical research, and it ensured a major role for meta-analysis in medical research from that point on (Hunt, 1997).

Psychology and the social sciences appear to operate at present on the medical model or at least a variant of this model. A common sequence appears to be that a meta-analysis is published, and, after a time, when additional relevant studies have become available, that meta-analysis is updated by inclusion of the new studies. For example, a meta-analysis of the relationship between job satisfaction and job performance was first conducted by Petty, McGee, and Cavender (1984) on the basis of 16 studies. This analysis was updated by Iffaldano and Muchinsky (1985) on the basis of an expanded database of 74 studies. Finally, Judge, Thorensen, Bono, and Patton (2001) incorporated 312 independent samples into their update. In general, this seems to be the dominant procedure today for updating of meta-analyses as new studies become available. It will be important later to note that nothing about the medical model precludes analysis of potential moderators. In fact, almost all published meta-analyses include examination of potential moderators.

Bayesian Methods: Some General Properties

A general property of Bayesian methods is that they combine information from the current study (current information) with prior knowledge about likely values of the parameter to produce final parameter estimates. Under the general Bayesian model, the prior information can come from any source. In fact, Bayesian statistical methods have often been criticized on grounds that the prior probabilities (or prior distributions) that are used are subjective in nature (e.g., see Lee, 1989), and, indeed, this is often the case. However, this is not true for the procedures discussed in the present article. The applications discussed here are empirical Bayes analyses—that is, applications of Bayesian equations to data in which the prior distributions have been empirically estimated via meta-analysis. This fact alone, however, does not eliminate potential problems in all Bayesian applications. In the basic Bayesian model, the prior distribution is supposed to incorporate all prior information about both the mean of the population parameter and the uncertainty about that mean. However, the concept of uncertainty is not explicitly defined in the general Bayesian model. There is general agreement that this uncertainty is indexed by the variance or standard deviation of the prior distribution. In subjective Bayesian applications, this variance is based on subjective estimates of that variance or standard deviation. In empirical Bayesian applications, it is acceptable for this variance to be the sampling error variance of the mean of the prior distribution, and later we present a Bayesian method in which this variance is used as the variance of the prior. However, within the general empirical Bayesian model, this variance can also be an empirically derived estimate of the variance of the population parameters that does not include the influence of sampling error variance. As noted later, one could argue that this variance is more informative (has higher information value) than a value that includes sampling error variance. Although it seems somewhat paradoxical, we later show that this sort of prior distribution leads to problems when Bayesian procedures are used to update meta-analysis findings.

Bayesian Procedures of Schmidt et al. (1979) and Brannick et al. (2002)

This article examines the specific Bayesian procedures advocated to date in the industrial/organizational literature for updating meta-analyses when new studies become available and for using prior distributions from meta-analyses to improve the accuracy of parameter estimates for particular studies or situations. In their initial published articles presenting psychometric meta-analysis (and its use in validity generalization), Schmidt and Hunter (1977; Schmidt et al., 1979) advocated the use of these Bayesian methods for meta-analysis results with new studies. However, in their subsequent work (e.g., Hunter & Schmidt, 1990, 2004; Hunter, Schmidt, & Jackson, 1982; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Ones, Viswesvaran, & Schmidt, 1993; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980), they dropped this procedure in favor of what in this article is called the medical model. Recently, Brannick (2001), Brannick and Hall (2003), and Brannick et al. (2002) have advocated this same Bayesian procedure to estimate population parameters in specific settings or studies (although not for the purpose of updating meta-analyses more generally).

It is important to note that the procedures discussed in this article are all random-effects models; that is, they do not assume a priori that the parameter being estimated has only one value (as fixed-effects models do) but instead assume it may have a distribution of values. (As noted later, random-effects models may yield an empirically determined estimate of zero for the variance of population values, but this is very different from assuming a zero value a priori.) The properties of fixed-effects Bayesian models are somewhat different, but fixed-effects models are rarely appropriate in meta-analysis (Hunter & Schmidt, 2000, 2004; National Research Council, 1992) and are not discussed in any detail in this article. It should also be noted that the meta-analysis procedures developed by Callender and Osburn (1980) and Raju and Burke (1983), also random-effects models, did not incorporate the Bayesian analysis procedure discussed here. The later Schmidt–Hunter meta-analysis articles did not present a basis for the preference for the medical model over their original Bayesian procedure. In the course of exploring the properties of the medical model and this Bayesian procedure as tools for updating meta-analysis results, this article presents such a basis.

All Bayesian models specify the same basic mathematical model for combining a new study with the information contained in the previous meta-analysis (e.g., see Box & Tiao, 1973; Lee, 1989; Novick & Jackson, 1974). It is important to note that these

methods are not limited to use with a single new study but can also be used to combine multiple new studies sequentially with the previous meta-analysis. Although the basic mathematical model is always the same, both the nature of the prior distributions used and the information entered from the new study can vary between procedures, producing differences in the final parameter estimates. Because of this, statements made in this article about the specific Bayesian procedures advocated by Schmidt et al. (1979) and Brannick et al. (2002) should not be taken as generalizing to all Bayesian procedures.

We now illustrate the random-effects Bayesian model advocated by Brannick (2001; Brannick and Hall, 2003; Brannick et al., 2002), Schmidt and Hunter (1977), and Schmidt et al. (1979) using the correlation statistic. The estimates of the mean and standard deviation of population true validities ($\hat{\bar{\rho}}$ and $\hat{\sigma}_\rho$) from the original meta-analysis are regarded as indexing all prior knowledge on the relationship in question—that is, knowledge prior to the appearance of the new study. This is referred to as the *prior distribution.*[1] The information from the new study is referred to as the *likelihood ratio;* its mean is the correlation from the study (corrected for measurement error, range restriction, and any other appropriate artifacts), and its standard deviation is the standard error of that estimate (computed as described below). As in all Bayesian procedures, the prior distribution is combined with the likelihood ratio to yield the posterior distribution, which is the final estimate of revised knowledge. Its mean is taken as the most likely value of the correlation, and its standard error is taken as the index of the uncertainty about that value. The equations for combining the prior with the likelihood ratio are given in any Bayesian text (e.g., Lee, 1989; Novick & Jackson, 1974). The exact form of these equations differs depending on the assumed shape of the prior and likelihood distributions (e.g., whether they are assumed to be normal or assumed to follow the beta distribution). The procedures discussed in this article assume normal distributions, an assumption that is frequently made in Bayesian analysis.

The standard error of the observed correlation, $V^{1/2}(r)$, is

$$V^{1/2}(r) = \frac{(1 - r^2)}{\sqrt{N - 1}}. \quad (1)$$

The square of Equation 1 expresses the simple sampling error of the observed correlation. The standard error of the correlation corrected for measurement error and range restriction is somewhat larger, as will be seen later. In random-effects models, a different equation is often used for sampling error variance (discussed below).

The prior distribution is multiplied by the likelihood ratio to produce the posterior distribution. In this process of combination, the Bayesian equation for the mean of the posterior (Novick and Jackson, 1974, p. 140) is

$$_{post}\hat{\bar{\rho}} = \frac{(_{prior}\hat{\sigma}_\rho^{-2})(_{prior}\hat{\bar{\rho}}) + (V^{-1}(r))(r)}{(_{prior}\hat{\sigma}_\rho^{-2}) + V^{-1}(r)} = \frac{\frac{_{prior}\hat{\bar{\rho}}}{_{prior}\hat{\sigma}_\rho^2} + \frac{r}{V(r)}}{\frac{1}{_{prior}\hat{\sigma}_\rho^2} + \frac{1}{V(r)}}, \quad (2)$$

where $_{post}\hat{\bar{\rho}}$ is the estimate of the mean of the posterior distribution, $_{prior}\hat{\sigma}_\rho^{-2}$ is the inverse of the variance of the prior (meta-analytic) distribution, $_{prior}\hat{\bar{\rho}}$ is the mean of the prior distribution, $V^{-1}(r)$ is the inverse of the variance of the likelihood ratio, and r is the estimate of the population correlation from the new study. The Bayesian equation for computation of the variance of the posterior distribution is (Novick & Jackson, 1974, p. 140)

$$_{post}\hat{\sigma}^2 = (_{prior}\hat{\sigma}_\rho^{-2} + V^{-1}(r))^{-1} = \left(\frac{1}{_{prior}\hat{\sigma}_\rho^2} + \frac{1}{V(r)}\right)^{-1} = \frac{(_{prior}\hat{\sigma}_\rho^2)(V(r))}{_{prior}\hat{\sigma}_\rho^2 + V(r)}. \quad (3)$$

As we will see later, this approach to Bayesian analysis often privileges the new study in that it assigns the new study a much larger effective weight than it would have received had it been included in the original meta-analysis. We later show that this stems in part (but only in part) from the nature of the sampling error variance that is used to represent uncertainty in the likelihood ratio (the new study). Use of a different estimate of sampling error variance in the new study (the random-effects estimate, discussed later) reduces the impact of the new study on the final result. However, the fixed-effects sampling error variance used in our example below is the one advocated for the new study.[2]

Consider a specific example. In Pearlman et al. (1980), the meta-analysis examining the validity of perceptual speed measures for predicting job performance obtained an estimated mean operational validity of .47, with $\hat{\sigma}_\rho = .22$. Until a new study is done, this distribution is the Bayesian posterior distribution; when a new study appears, this distribution becomes the prior distribution to be combined with the findings of that study to produce the new Bayesian posterior distribution, which then expresses the updated sum total of knowledge about that relationship. Suppose a new study is conducted with $N = 65$, and its observed (uncorrected) validity is .30. After one corrects for unreliability in the job performance ratings (assuming a reliability of .60, as Pearlman et al., 1980, did) and then for range restriction (assuming direct range restriction, as Pearlman et al., 1980, did, and assuming $u_x = sd_i/SD_a = .603$, the mean range restriction value in Pearlman et al., 1980), the estimate of operational validity in the study is .57. This is the mean of the likelihood ratio representing the results of the single new study.

We also need to calculate the standard deviation of the likelihood ratio, which is the standard error of the .57 mean value. If we apply Equation 1, the standard error of the original observed value of .30 is $(V^{1/2}(r)) = \frac{1 - .30^2}{\sqrt{65 - 1}} = .11375$. However, the standard error of the corrected estimate of .57 is larger, because the corrections increase the amount of sampling error in the estimate (Bobko

[1] In the random effects Bayesian method discussed in this article, it is clear that the prior distribution is the ρ distribution from the previous meta-analysis. For example, in Brannick (2001), see pages 472–474. In Brannick et al. (2002), see page 5. In Schmidt et al. (1979), see page 275. (In addition to the random effects model discussed here, Brannick, 2001, also discussed a fixed effects Bayesian model; because fixed effects models are rarely appropriate in meta-analysis, this article does not discuss that model; National Research Council, 1992.)

[2] Brannick (2001, p. 474) and Brannick et al. (2002, p. 5) explicitly stated that the sampling error variance of the new study is taken as the simple (i.e., fixed-effects) sampling error variance. This was implicit in Schmidt and Hunter (1977) and Schmidt et al. (1979).

& Reick, 1980; Cureton, 1936; Forsyth & Feldt, 1969; Hunter & Schmidt, 2004, Chapter 3; Kelly, 1947; Raju & Brand, 2003). The factor of correction is .57/.30 = 1.90, but because of the nonlinearity of the range correction, the standard error is increased by less than this factor. The standard error of the corrected estimate is $V^{1/2}(r_c) = (.11375)(1.90)(.864) = .18587$, where $V^{1/2}(r_c)$ is the standard error of the corrected validity estimate, .11375 is the standard error of the uncorrected estimate of .30, 1.90 is the correction factor (i.e., .57/.30), and .864 is the adjustment factor for range restriction calculated via Equation 4,

$$a = 1/[(U_x^2 - 1)r^2 + 1], \quad (4)$$

where a = the adjustment factor, $U_x = 1/u_x$ (here, 1/.603), and r = the observed (uncorrected) correlation (here, .30). This factor adjusts for the fact that the range restriction correction, unlike the measurement error correction, is nonlinear and so increases the standard error somewhat less than indicated in the correction factor. For a complete explanation, see Hunter and Schmidt (2004, pp. 109–112; Equation 4 here is Equation 3.21 in Hunter & Schmidt, 2004). We now have all the information we need to apply the Bayesian procedure. We have the prior distribution, with a mean of .47 and a $\hat{\sigma}_\rho$ of .22, and we have the likelihood ratio, with a mean of .57 and a $\hat{\sigma}_\rho$ of .18587 (.19, rounded).

Applying Equations 2 and 3, we obtain for the mean and standard deviation, respectively, of the new posterior distribution, ${}_{post}\bar{\rho} = .53$ and ${}_{post}\hat{\sigma}_\rho = .14$. From this, it can be seen that the single new study had an important effect on our estimates: It raised the mean from .47 to .53 (a 13% increase) and reduced the $\hat{\sigma}_\rho$ from .22 to .14 (a 36% decrease). Hence, it is clear that the new study was weighted quite heavily.

The statistic used in the example above is the correlation coefficient. Some authors (e.g., Hedges, 1988) recommend use of the Fisher's (Fz) transformation of the correlation coefficient (followed by back-transformation of final estimates to r) on grounds that Fz more closely approximates a normal distribution. However, this difference in distribution shape is quite small and, in any event, appears only with large sample sizes (Hunter & Schmidt, 2004; Hunter, Schmidt, & Coggin, 1996). In addition, use of the Fisher's z transformation appears to produce an upward bias in the mean correlation in random-effects models (Field, 2001, 2005; Hall & Brannick, 2002; Hunter et al., 1996; Schulze, 2004).

This Bayesian approach to incorporating new information (new studies) into prior knowledge (the results of an existing meta-analysis) has been advocated for two somewhat different purposes. The first is simply to update the prior distribution by incorporating the new study. This usage was advocated by Schmidt and Hunter (1977) and Schmidt et al. (1979), although not by Brannick (2001) or Brannick et al. (2002). Until a new study appears, the mean and standard deviation from the existing meta-analysis describe the posterior distribution—that is, the distribution that incorporates all existing knowledge. When the new study appears, this posterior distribution becomes the prior distribution and is combined with the results of the new study (i.e., the likelihood ratio from that study) to produce a new, revised posterior distribution. This new posterior distribution is taken as summarizing all available knowledge about this parameter.

The second purpose for which Bayesian methods have been advocated is to provide estimates for a particular study or setting (Brannick, 2001; Brannick & Hall, 2003; Brannick et al., 2002; Hedges, 1988; Raudenbush & Bryk, 1985; Schmidt & Hunter, 1977; Schmidt et al., 1979). In this usage, the posterior $\hat{\bar{\rho}}$ of .53 and ${}_{post}\hat{\sigma}_\rho$ of .14 in our example above would be taken not as a statement of updated general knowledge but as an improved estimate of what the true state of nature is in the new study (or the setting or situation represented by that study). That is, it is taken as a more accurate estimate for the individual study. Use of Bayesian analysis in this manner is based on the explicit or implicit assumption that it is likely that actual parameter values (independent of artifacts that distort obtained values) will differ to an important extent by study or setting, making it meaningful to attempt to obtain the most accurate estimates for each individual study or setting. We later present evidence questioning this assumption.

Conditions Necessary for the Medical Model and the Above-Mentioned Bayesian Model

The conditions necessary to use the medical model[3] are fairly basic: The new study (or studies) must be included in the updated meta-analysis on the same basis as the studies included in the previous meta-analysis. Thus, it can be seen that use of the medical model to update a meta-analysis requires access to the study-level data for the studies used in the previous meta-analysis (in addition to the data from the new study).

The conditions necessary for use of the Bayesian model discussed here, as stated or implied by Brannick (2001) and Schmidt and Hunter (1977), are more complicated. First, the prior distribution and the likelihood ratio must be in the same statistical metric. For example, if the meta-analysis has corrected for downward biases caused by measurement error and range restriction, it would not be appropriate to combine this prior distribution with a likelihood ratio consisting of the observed (uncorrected) correlation in the new study and its standard error. In this connection, it can be noted that in the Bayesian analysis presented by Brannick (2001) neither the prior nor the likelihood ratio was corrected for these two artifacts (the prior was corrected only for sampling error variance), whereas Brannick et al. (2002) presented an improved model in which both distributions were corrected for these artifacts. In both cases, the two distributions meet the requirement of being in the same metric. A procedure in which no artifact corrections are made is of limited research value. Typically, the requirement of the same metric means that the estimate from the new study must also be corrected for artifacts (Hedges, 1988). It is important to note that this condition does not mean that if the likelihood ratio reflects sampling error variance, the prior distribution must also reflect sampling error variance. As noted earlier, under general Bayesian theory, the prior distribution can be any encoding of prior information deemed the best summary of prior knowledge. Although it is not the case here, the prior information can even be based on subjective judgment, and it is clear that subjective judgments do not include sampling error. If it is possible to eliminate sampling error variance from the variance of the prior (as it is in meta-analysis), this can be viewed as being required to

[3] In referring to this approach as the medical model, we do not intend to imply that medical researchers and biostatisticians never use Bayesian methods. Obviously they do (e.g., see Berry & Stangl, 1996). However, the procedure we are addressing here is not Bayesian.

make the prior the most accurate possible summary of prior knowledge. This is, in fact, the second condition.

Second, these methods assume that the variance of the prior distribution must not include sampling error. Inclusion of sampling error would mean that the prior does not in fact summarize existing knowledge, because sampling error variance is artifactual variance and, in a meta-analysis, can be removed from the prior to provide a more accurate estimate of ${}_{prior}\hat{\sigma}_\rho$ (or ${}_{prior}\hat{\sigma}_\delta$, if the *d* statistic is used), thus more accurately reflecting attained knowledge. This does not imply that the estimates of $\bar{\rho}$ and σ_ρ describing the prior distribution are parameter values and are estimated perfectly; it means simply that the variance produced by sampling error in the distribution of correlations has been subtracted out in the estimation of ${}_{prior}\hat{\sigma}_\rho$. Brannick (2001), Brannick et al. (2002), Schmidt and Hunter (1977), and Schmidt et al. (1979) all advocated that sampling error be removed from the prior distribution used (see Footnote 1). As we show later, this assumption is one of the major reasons why the results from the medical model differ from the results from Brannick et al.'s and Schmidt et al.'s Bayesian procedure.

Third, the prior distribution must have a nonzero standard deviation. As can be seen by examination of Equations 2 and 3, if the prior distribution in the Bayesian analysis has ${}_{prior}\hat{\sigma}_\rho = 0$, the posterior distribution is the same as the prior distribution, and the new study always obtains an effective weight of zero. This condition rules out the use of the Bayesian model in cases in which the prior random-effects meta-analysis found that artifacts accounted for all the variance of the observed correlations or *d* values, leading to the result that ${}_{prior}\hat{\sigma}_\rho = 0$. In such cases, which occur with some frequency, this Bayesian procedure would assign a zero weight to any new study (and to a sequence of such studies). The medical model, however, would assign nonzero weights—namely, the same weights as the new studies would have had if they had been included in the original meta-analysis. The alternative Bayesian procedure described later would perform similarly to the medical model. In passing, we again note that an empirically determined estimate of zero for the variance of the population parameters produced by a random-effects model is not equivalent to the a priori assumption of a zero value that characterizes the fixed-effects model in meta-analysis.

Using the Schmidt–Hunter Bayesian Procedure to Update the Posterior Distribution

As noted earlier, one purpose for which the Bayesian procedure discussed here has been advocated is to update the posterior distribution by incorporating the new study. The resulting new posterior distribution is then said to capture everything known about the relationship up to that point. As most treatments of Bayesian statistics make clear, there is nothing wrong with this process in principle and as a general statistical method. In fact, in the next section of this article we present an alternative Bayesian procedure in which this updating process works quite well. However, in the specific Bayesian procedure being examined here, in which the prior distribution is described by the mean and variance from a previous meta-analysis, it is very often the case that Bayesian analysis privileges the new study in a way that may not seem plausible or rational. That is, the effective weight given to the new study, relative to the studies included in the existing meta-analysis, appears unreasonably large. For example, the meta-analysis might be based on 200 studies, with a total sample size of 185,000, and the new study might be based on $N = 85$. Yet the single new study may be given a weight equal to that given to all 200 previous studies combined.

Consider the example presented earlier. In Pearlman et al. (1980), the meta-analysis examining the validity of perceptual speed measures for predicting job performance in clerical jobs was based on 882 studies with a total sample size of 70,935. The new study was based on $N = 65$. Yet in the Bayesian model, addition of this small new study raised the mean estimate from .47 to .53 and reduced the standard deviation estimate from .22 to .14. Hence, it is clear that the new study was weighted quite heavily. The weights applied to the prior distribution and to the likelihood ratio are the inverse of their variances. The weight given to the prior is 1/.0484, which is 20.66. The weight given to the single study is 1/.0361, which is 27.70. The relative weights are more informative: The single new study was given a weight 1.34 times greater than the weight given to the prior. Alternatively, the prior distribution, based on 882 studies with a total sample size of 70,935, received only 74% as much weight as a single new study, based on only $N = 65$. Hence, the new study was privileged to an extreme degree. If one looks at it another way, the Bayesian analysis says that 57% of the available total information is contained in the single new study based on a sample size of 65, whereas the prior distribution contains only 43% of the total information.

What would these percentages be if the medical model were used? The weight given to the new study in the medical model would be approximately 65/71,000 = .0009, whereas the weight given to the other 882 studies would be 70,935/71,000 = .9991. That is, instead of getting only 74% as much weight as the new study, the prior would get 1,110 times as much weight. Hence, these two methods lead to grossly different relative weights.

This example is unusual only in that most meta-analyses are based on fewer than 882 studies and their total sample sizes are smaller than 70,935. For example, in Pearlman et al. (1980), the prior for general mental ability measures was based on 194 studies with a total sample size of 17,539; its mean was .52 and its $\hat{\sigma}_\rho$ was .24 (see Table 7 in Pearlman et al., 1980), a $\hat{\sigma}_\rho$ value similar to the .22 in our example. It is important to note that the weight given in this Bayesian procedure to the prior meta-analysis results does not depend in any way on the number of studies or the total sample size; it depends only on $\hat{\sigma}_\rho$. For example, a meta-analysis based on only 6 studies with a total sample size of 500 that reports a $\hat{\sigma}_\rho$ value of .10 will get more weight than one that is based on 500 studies and a total sample size of 42,000 but that finds a $\hat{\sigma}_\rho$ of, say, .22, as in our example. The $\hat{\sigma}_\rho$ value of .22 in our example is quite typical. The average $\hat{\sigma}_\rho$ value in Pearlman et al. for job performance measures for pooled clerical job families was .199. This value is very similar to the .186 value for our study based on $N = 65$. Hence, on average in this Bayesian procedure, these meta-analysis results would receive a weight approximately equal to that given to a study based on only 65 people. Many similar examples can be found in the meta-analysis literature.

Another criticism can be made of this method: Because the likelihood ratio contains sampling error variance, the posterior distribution will also reflect sampling error variance, and therefore this posterior should not be interpreted as an estimate of the

variance of $\hat{\sigma}_\rho^2$, because the variance of ρ is defined as independent of sampling error variance; that is, it is defined as excluding variance due to sampling error. Another point is also relevant here. It is clear from an examination of Equation 3 that the variance of the posterior distribution is always smaller than the variance of the prior distribution. Logically, the addition of a new study could increase, as well as decrease, the estimate of the variance of $\hat{\sigma}_\rho^2$. This is, in fact, the case for the medical model and for the alternative Bayesian procedure discussed in the next section but not for this Bayesian procedure. This fact further reinforces the conclusion that the variance of the posterior in this procedure cannot be taken as an estimate of $\hat{\sigma}_\rho^2$. In the next section of this article, a Bayesian procedure is presented that, unlike the procedure described here, assigns essentially the same weight to a new study that is assigned by the medical model. However, in that procedure it is not the variance of the posterior distribution that is interpreted as the estimate of $\hat{\sigma}_\rho^2$. Instead, $\hat{\sigma}_\rho^2$ is estimated via a different equation from the one used to estimate the variance of the posterior distribution. It appears that in any Bayesian procedure used to update a meta-analysis, the variance of the posterior cannot be interpreted as $\hat{\sigma}_\rho^2$. Because that is, in fact, done in the procedure being examined here, this is a problem with the procedure.

A modification of this Bayesian procedure that computes the sampling error variance for the likelihood ratio on the basis of the random-effects model equation for sampling error variance would reduce but not eliminate this problem. The random-effects sampling error variance (Hedges & Vevea, 1998; Hunter & Schmidt, 2000; National Research Council, 1992) is the sum of the simple sampling error variance used in our example (Equation 1) and the variance of ρ:

$$V(r) = \frac{(1 - r^2)^2}{N - 1} + \sigma_\rho^2, \tag{5}$$

where $V(r)$ is the random-effects sampling error variance for study i, $\frac{(1 - r^2)^2}{N - 1}$ is the simple (fixed-effects) sampling error variance for a study, and σ_ρ^2 is the variance of the population-level correlations.

In our example, this is $(.18587)^2 + (.22)^2 = .08295$, yielding a standard error of $\sqrt{.08295} = .2880$. The larger random-effects standard error (i.e., standard deviation of the likelihood ratio) reflects the fact that the simple sampling error is not the only source of uncertainty in the estimate produced by the new study. In addition to the simple sampling error variance, there is uncertainty about the location of our study parameter in the distribution of ρ (Hedges & Vevea, 1998; Hunter & Schmidt, 2000). If we reapply Equations 2 and 3 using this standard error for the new study, the new posterior distribution is ${}_{post}\bar{\rho} = .51$ (vs. the earlier .53) and ${}_{post}\hat{\sigma}_\rho = .18$ (vs. the earlier .14).

The increase in the mean and the decrease in the standard deviation are both less than in the earlier analysis because of the change in the relative weights. The weight on the prior is now greater than the weight on the new study (20.66 vs. 12.055). The prior now receives 71% more weight than the new study. Still, the new study, based on $N = 65$, receives 58% as much weight as the prior, based on $N = 70{,}935$. Although this relative weighting is an improvement, the new study still gets a much greater weight than it would get in the medical model. In addition, it should be emphasized that use of the random-effects sampling error variance for the likelihood ratio was not advocated by Brannick (2001), Brannick et al. (2002), or Schmidt and Hunter (1977; see Footnote 2) and hence is not part of the procedure being discussed here.

Values for $\hat{\sigma}_\rho$ vary across meta-analyses, but values in the vicinity of .20 are not uncommon in large published meta-analyses. One way to describe this situation is to say that the relatively large standard deviations of the meta-analysis-derived prior distributions cause them to receive limited weight in this Bayesian procedure. One might ask why these standard deviation values are not smaller. The answer is that in most meta-analyses there are at least 10 artifacts that cause between-studies variance in study results but cannot be corrected for (Hunter & Schmidt, 2004, pp. 195–201; Schmidt & Hunter, 2003; Schmidt et al., 1993). In the single new study, such artifacts distort the observed estimate, but because there is only one study, they do not increase the variance of the likelihood ratio (as computed in the Brannick et al., 2002, and Schmidt & Hunter, 1977, procedures). For example, errors in data recording inflate between-studies variance in reported outcomes, but a data error in the single new study (e.g., a digit reversal in the reported correlation) does not increase the estimated variance of the likelihood ratio and hence does not decrease the weight given to the study results. Another example is the artifact of criterion contamination. In the meta-analysis results, variation across studies in criterion contamination causes variance in the meta-analysis distribution that (typically) cannot be corrected for. Conversely, in a single study criterion contamination will inflate or deflate the validity estimate (depending on whether the contamination is predictor correlated or not), but it will have no effect on the estimate of the standard deviation of the likelihood ratio.

An Alternative Bayesian Procedure for Updating Meta-Analysis

This section describes an alternative Bayesian procedure for updating a meta-analysis. Use of this procedure requires estimates of (a) the mean of population validities (or effect sizes), (b) the sampling error variance of the mean estimate, and (c) an estimate of the variance of population validities. Let these three estimates be denoted by $\hat{\bar{\rho}}$, $V(\hat{\bar{\rho}})$, and $\hat{\sigma}_\rho^2$, respectively. To denote that these estimates represent the prior distribution, we add a prescript to these three symbols: ${}_{prior}\hat{\bar{\rho}}$, $V({}_{prior}\hat{\bar{\rho}})$, and ${}_{prior}\hat{\sigma}_\rho^2$. Let there be $k - 1$ studies in the prior meta-analysis, and let the correlation (or effect size) and its sampling variance from a new study (Study k) be denoted by r_k and $V(r_k)$, respectively. In the alternative Bayesian procedure, the updated Bayesian meta-analytic estimates (based on all k studies) can be obtained as follows:

$${}_{post}\hat{\bar{\rho}} = \frac{[V^{-1}({}_{prior}\hat{\bar{\rho}})]({}_{prior}\hat{\bar{\rho}}) + [V^{-1}(r_k)](r_k)}{V^{-1}({}_{prior}\hat{\bar{\rho}}) + V^{-1}(r_k)} = \frac{\frac{{}_{prior}\hat{\bar{\rho}}}{V({}_{prior}\hat{\bar{\rho}})} + \frac{r_k}{V(r_k)}}{\frac{1}{V({}_{prior}\hat{\bar{\rho}})} + \frac{1}{V(r_k)}}. \tag{6}$$

$$V({}_{post}\hat{\bar{\rho}}) = [V^{-1}({}_{prior}\hat{\bar{\rho}}) + V^{-1}(r_k)]^{-1} = \left[\frac{1}{V({}_{prior}\hat{\bar{\rho}})} + \frac{1}{V(r_k)}\right]^{-1} = \frac{V({}_{prior}\hat{\bar{\rho}})V(r_k)}{V({}_{prior}\hat{\bar{\rho}}) + V(r_k)} \tag{7}$$

$$ {}_{post}\hat{\sigma}^2_{\rho} = \frac{(k-1)({}_{prior}\hat{\sigma}^2_{\rho}) + [k^2 V({}_{post}\hat{\bar{\rho}}) - (k-1)^2 V({}_{prior}\hat{\bar{\rho}})] - V(r_k)}{k}. \quad (8) $$

Equations 6 and 7 provide estimates of the mean of the population validities (ρs) and the sampling error variance of this estimated mean, respectively. They are very similar to Equations 2 and 3 in format, but with an important distinction. The weight for the prior estimate of the mean of ρs is the inverse of its sampling error variance, not the inverse of the variance of ρs. As shown in Appendix A, these estimates of the mean of ρs and its sampling variance are nearly identical to estimates that one would obtain in the medical model in the random-effects case. (These estimates would be exactly identical in the fixed-effects case, but the fixed-effects case is rarely appropriate in meta-analysis; National Research Council, 1992; and in the fixed-effects case both the prior and the posterior variance of the estimate of the mean ρs—$V({}_{prior}\hat{\bar{\rho}})$ and $V({}_{post}\hat{\bar{\rho}})$—would be zero by definition in Equations 6 and 7. Also, Equation 8 would be zero by definition and not needed.) It is important to note that in Equations 6 and 7, the $V(r_k)$ term refers to random-effects sampling error variance as defined in Equation 5. However, in Equation 8, the term $V(r_k)$ by definition refers to the simple (fixed-effects) sampling error variance as defined in Equation 1. (Justification for Equation 8 as an estimate of the variance of ρs is given in Appendix B.) That is, in this random-effects model, the sampling variance of r_k in Equations 6 and 7 is defined as

$$ V(r_k) = \frac{(1 - r_k^2)^2}{N_k - 1} + \sigma^2_{\rho}, \quad (9) $$

where N_k is the number of participants in Primary Study k (Equation 9 is the same as Equation 5). As shown in Equation 1, in the fixed-effects model the variance of r_k consists of only the first term on the right-hand side of Equation 9. Because an estimate of the variance of ρs is typically available from the prior meta-analysis (based on $k - 1$ primary studies), one can use that estimate of the variance of ρs to compute the variance of r_k in this random-effects model. Given that such an estimate of the variance of ρs is based on $k - 1$ primary studies rather than all k studies, one should expect to find small differences between the estimates (Equations 6, 7, and 8) based on the alternative Bayesian procedure and the estimates based on the medical model.

Equations 6, 7, and 8 are presented for bare-bones meta-analysis (i.e., analyses with no correction for measurement error, range restriction, dichotomization, or other artifacts). A bare-bones application of this procedure requires only knowledge of the correlation values and their samples sizes. For a complete meta-analysis (one including corrections for artifacts) more information is required, as discussed in Appendix A. However, application of the medical model requires the same items of information for the studies in the prior meta-analysis.

Table 1 shows the data for 10 primary studies, and Table 2 displays the results from a meta-analysis using both the medical model and the alternative Bayesian approach. For simplicity, the analysis presented is a bare-bones analysis (no corrections for measurement error or range restriction were made). The procedures described in Appendixes A and B were used to estimate the needed posterior parameters. For the prior information, the maximum likelihood iterative procedures described in Raju and Drasgow (2003) were used. These methods are consistent with the description in Appendix A. In this example, the medical model and the alternative Bayesian procedure yield almost identical estimates of the parameters. Estimates from the medical model and the alternative Bayesian model will not necessarily always be this close. Additional empirical research would be helpful to get a better assessment of the degree to which these methods provide similar results.

Table 1
Data for 10 Primary Studies

Study number	N	r
1	60	.44
2	75	.20
3	85	.60
4	110	.32
5	50	.41
6	90	.25
7	100	.12
8	65	.35
9	80	.35
10	65	.19

A reviewer inquired as to why it is important for a Bayesian procedure to yield results similar or identical to those from the medical model. A desirable property in a Bayesian procedure is invariance of results with respect to the order of introduction of data. That is, it is desirable that the final estimate of the parameter mean and standard deviation be independent of the order in which different studies were (or are) entered into the analysis (Lee, 1989; Novick & Jackson, 1974). Obviously, the medical model, although not a Bayesian model, has this property. The Bayesian method we introduce here also has this property. However, the older Bayesian procedure does not have this property. For example, in our earlier example, the final results will change—perhaps dramatically—if the new study, with $N = 65$, had instead been included in the original meta-analysis and a different study from the meta-analysis had been the new study. Hence, our new Bayesian model possesses this Bayesian desideratum, whereas the older Bayesian model does not. This consideration relates to questions about the meaning of the concept of information value of a prior Bayesian distribution. As noted in an earlier section, it can be argued that the mean ρ and SD_{ρ} provides a prior distribution with more information value than mean ρ and the sampling error variance of mean ρ, because the former is free of sampling error variance and the latter is not. However, as we have seen, the use of the former as the prior leads to a violation of a basic Bayesian standard.

Using the Brannick–Hunter–Schmidt Empirical Bayesian Procedure to Estimate Parameter Values in Specific Studies

We return now to the older Bayesian procedure. Brannick (2001), Brannick and Hall (2003), Brannick et al. (2002), Hedges (1988), Schmidt and Hunter (1977), and Schmidt et al. (1979) have

Table 2
Validity Generalization Estimates From the Medical Model and the Alternative Bayesian Approach

Condition	$\hat{\bar{\rho}}$	$V(\hat{\bar{\rho}})$	$\hat{\sigma}^2_\rho$
MA with the first 9 studies (prior)	.346	.002	.012
MA with all 10 studies (medical model)	.333	.002	.012
Bayesian estimation for all 10 studies	.333	.002	.011

Note. MA = meta-analysis.

recommended that Bayesian methods be used to enhance the accuracy of estimates of the true validity, correlation, or effect size in individual studies. Each individual study produces its estimate of the population effect size, and that value could be used to estimate the actual population value underlying that study. The advocates of this Bayesian procedure argue, however, that researchers can improve this estimate by combining it with previous knowledge about the distribution of such estimates across other studies—that is, by combining it with the posterior meta-analytic distribution of ρ or δ, which becomes the prior distribution for the Bayesian analysis. The formal calculations for doing this are the same as for the first Bayesian procedure; that is, we again apply Equations 2 and 3.

A problem here is that this usage of Bayesian procedure requires the a priori assumption that parameter values (i.e., population correlations) differ importantly across studies beyond the variation caused by artifacts and hypothesized moderators. That is, this is not an assumption about the effects of artifacts or hypothesized moderators. For example, if the hypothesized moderator of job complexity is supported empirically (as, indeed, it is; e.g., see Hunter & Hunter, 1984), then the complexity level of the job in the new study determines which of the complexity-level meta-analysis distributions is to be used as the prior distribution in the Bayesian analysis. For example, if the job in the new study has Complexity Level 3 (medium complexity), then the meta-analysis result for medium-complexity jobs is used as the prior distribution. This meta-analysis will have been corrected for some but not all variance-producing artifacts. In the context of personnel selection, the method advocated by Brannick (2001), Brannick and Hall (2003), Brannick et al. (2002), and Schmidt and Hunter (1977) requires the a priori assumption that situational specificity is real and important. Researchers typically make this assumption made without considering (weighting in) the prior probability that the assumption is correct. In fact, there is empirical evidence (discussed below) indicating that this probability is low, and herein lies the crux of the problem. A fundamental principle of Bayesian analysis is that statistical procedures should incorporate all available prior information into the procedure. The fact that classical statistics does not do this is the basis of the Bayesian criticism of classical statistics (Lee, 1989; Novick & Jackson, 1974). Therefore, from a Bayesian point of view, it is inappropriate to assume a priori the truth of the situational specificity hypothesis (i.e., study-specific parameters) without examining and incorporating the evidence for its prior probability.

What is the prior probability of this assumption? Most of the evidence bearing on the tenability of the assumption of study-specific population (parameter) values comes from research on the validity of personnel selection methods. This is because this area, unlike others, had a long history of informal and uncritical acceptance of the hypothesis of situational specificity, which eventually stimulated empirical tests of this assumption (Schmidt, Hunter, Pearlman, & Hirsh, 1985; Schmidt et al., 1993). Within the personnel selection area, the strongest evidence comes from the study of cognitive ability (general mental ability) tests and specific aptitude tests (e.g., verbal, quantitative, and spatial aptitude). This evidence was presented in detail in Schmidt et al. (1993) and Schmidt and Hunter (2003); here we merely list the lines of evidence. First, well-controlled large-scale consortium studies, superior methodologically to other studies, find no evidence for situational specificity. Second, studies conduced repeatedly over time in exactly the same setting and job find the same amount of variability in observed validities as is found across organizations, contradicting predictions of situational specificity. Third, 10 different sources of artifactual variance in validities exist that are not corrected for in most published meta-analyses, inflating estimates of σ^2_ρ. Fourth, a large-sample meta-analysis found that when 4 of these sources of artifactual variance were eliminated, estimates of σ^2_ρ decreased to nearly zero.

In light of these multiple converging lines of disconfirming empirical evidence, probably the most reasonable conclusion for ability and aptitude test validity is either that there is no situational specificity or that it is so small as to be negligible. Either of these indicates a low prior probability for the assumption required by the use of Bayesian methods to refine single-study estimates of situation-specific validities. During the review process for this article, the question was raised regarding whether accepting a zero or essentially zero value for the σ^2_ρ on the basis of these findings is equivalent to use of the fixed-effects model in meta-analysis. The fixed-effects model assumes a priori that $\sigma^2_\rho = 0$, so all estimates of σ^2_ρ are constrained to be zero by fiat. The research referred to here used a random-effects model to estimate σ^2_ρ empirically and then, on the basis of the empirical findings, concluded that the best estimate was zero or near zero. Hence, this is quite different from assuming a priori that the value is zero.

Fewer data are available on other types of selection methods (Schmidt & Hunter, 1998), so the evidence against situational specificity is not as ironclad. However, a key finding has been that the $\hat{\sigma}_\rho$ values for such selection methods as employment interviews (McDaniel et al., 1994), assessment centers (Gaugler, Rosenthal, Thornton, & Benson, 1987), ratings of education and job experience (McDaniel, Schmidt, & Hunter, 1985), biographical data (Carlson, Scullen, Schmidt, Rothstein, & Erwin, 1999; Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990), integrity tests (Ones et al., 1993), job knowledge tests (Dye, Reck, & Murphy, 1993), and college grades (Roth, BeVier, Switzer, & Shippmann, 1996) have been no larger than those for ability and aptitude tests, for which the evidence indicates that the remaining variance is probably due to artifacts not corrected for. This fact suggests that if the same analyses were done on these predictors as have been conducted for aptitude and ability measures, the findings would again support the conclusion that there is little or no variation across studies in validities. These considerations cast doubt on the assumption that there are substantial study-specific differences in population correlations for these job performance predictors once artifactual sources of variation and theoretically hypothesized moderators are accounted for. This is the

assumption that must be accepted to justify the use of Bayesian analysis to estimate validities in specific settings.

Other research areas have not traditionally been characterized by belief in study-specific population parameters, which suggests a low prior probability that such variation exists in these domains. In addition, the disconfirming findings in the one area (personnel selection) in which the assumption of substantial situational specificity has been extensively examined empirically suggest a low probability for the truth of this assumption. In any event, the evidence that is needed to support this assumption does not exist, rendering the assumption questionable. A feeling that this assumption is intuitively plausible is not a substitute for evidence in support of it.

We saw earlier that when one uses the Bayesian application suggested by Schmidt and Hunter (1977) and Schmidt et al. (1979) simply to update the data by incorporating a new study, the method appears to greatly privilege the new study, with the new small-sample study often receiving as much weight as all previous studies combined. The same thing occurs with the second usage of Bayesian methods, and, in fact, our earlier example could be reinterpreted as an analysis intended to estimate a study-specific population parameter. In this application the privileging might be justified if, in fact, there were evidence that each individual study represented an importantly different population parameter from those underlying other studies. In the second type of Bayesian application, the problem is not so much that the method privileges the new study but rather that it does this on the basis of the assumption of study-specific parameters—an assumption that appears to have a low prior probability of holding true. Hence, we have the irony that this ostensibly Bayesian procedure itself violates the Bayesian principle that available information on relevant prior probabilities should be weighted into all statistical procedures. This problem occurs not because the procedure is Bayesian but because it is incompletely Bayesian[4].

In an earlier section, we saw that use of the random-effects sampling error variance estimate for the new study reduced somewhat the apparent overweighting of the new study used to update the meta-analysis. Although use of the random-effects sampling error variance estimate would have the same effect in the present case, its use would not be conceptually appropriate. In this usage, the objective is to estimate the specific value of ρ thought to characterize the particular situation represented by the study. This means that only the simple sampling error variance indexes the uncertainty in the new study estimate, so it is inappropriate to use the random-effects sampling error variance as the uncertainty index for the situational study.

Given that the Bayesian procedure advocated by Brannick et al. (2002) and Schmidt et al. (1979) is not appropriate, at least in the area of personnel selection, what procedures can be used to estimate local parameters? There are several possibilities. First, one could apply the medical model. The medical model would avoid the inappropriate—or at least highly questionable—assumption that the new study represents an importantly different population value by rerunning the meta-analysis with the new study included. The new study would be weighted in the same manner as the previous studies, and the resulting estimated meta-analytic distribution of ρ or δ would be taken as summarizing existing knowledge about the relationship. This is the appropriate weighting of the new study if there is little or no situational specificity, which, on the basis of the available evidence, is more probable than the opposite. Second, one could use the alternative Bayesian procedure presented earlier to combine the new study with the previous meta-analysis findings. As shown earlier, this procedure would be expected to produce results very similar to those of the medical model. Finally, there is the option of using validity generalization. That is, do not conduct a local study; generalize validity to the new setting on the basis of the results of the previous meta-analysis. This procedure would be expected to produce results similar to those of the medical model and the alternative Bayesian procedure.

Summary

As research conclusions increasingly come to be based on meta-analyses, the question of how to update meta-analytic results as new studies become available assumes increased importance. The traditional procedure for doing this, called the medical model because of its extensive use in medical research, calls for adding new studies to the existing database and recalculating the meta-analysis. This article has examined the Bayesian procedures recommended by Schmidt and Hunter (1977) and Schmidt et al. (1979) as alternative methods for this purpose and has shown them to be problematic. This Bayesian procedure for updating a meta-analysis assigns extremely large weights to new studies, relative to the body of existing studies, and there appears to be no basis for this extreme weighting. The use of this Bayesian procedure to estimate study-specific parameters, recommended by Brannick (2001), Brannick et al. (2002), Schmidt and Hunter (1977), and Schmidt et al. (1979), also does this, but that application of the procedure is based on an a priori assumption that appears to have a low prior probability of holding, which makes the method questionable and contrary to Bayesian principles. We therefore conclude that the medical model and the alternative Bayesian procedure proposed here should be the methods of choice for updating meta-analytic findings in most research areas.

[4] A reviewer pointed out that this same argument could be made against the use of credibility intervals in meta-analysis. Because they are based on the estimated distribution of population values, credibility intervals can be viewed as a Bayesian application. Whether they are so viewed or not, it is clear that overestimation of σ_ρ^2 leads to underestimation of lower credibility bounds and overestimation of upper credibility bounds. This fact has been pointed out in the literature (e.g., Hunter & Schmidt, 2004, pp. 66–69). However, if there are areas in which evidence is found that population parameters in fact differ so substantially that such variance dwarfs variance in the prior due to artifacts not corrected for, an argument can be made for use of this empirical Bayes approach. But the important thing is that there be such evidence.

References

Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chambers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Journal of the American Medical Association, 268,* 240–248.

Berry, D. A., & Stangl, D. K. (1996). *Bayesian biostatistics.* New York: Marcel Dekker.

Bobko, P., & Reick, A. (1980). Large sample estimators for standard errors of junctions of correlation coefficients. *Applied Psychological Measurement, 4,* 385–398.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis.* New York: Wiley.

Brannick, M. T. (2001). Implications of empirical Bayes meta-analysis for test validation. *Journal of Applied Psychology, 86,* 468–480.

Brannick, M. T., & Hall, S. M. (2003). Validity generalization from a Bayesian perspective. In K. Murphy (Ed.), *Validity generalization: A critical review*. Mahwah, NJ: Erlbaum.

Brannick, M. T., Hall, S. M., & Liu, Y. (2002, April). Empirical Bayes local validation estimates with artifact corrections. In S. Maxwell (Chair), *Rethinking artifact corrections in meta-analysis: Innovations and extensions.* Symposium conducted at the meeting of the Society for Industrial and Organizational Psychology, Toronto, Ontario, Canada.

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

Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H. R., & Erwin, F. (1999). Generalizable biographical data validity: Is multi-organizational development and keying necessary? *Personnel Psychology, 52,* 731–756.

Cureton, E. E. (1936). On certain estimated correlation functions and their standard errors. *Journal of Experimental Education, 4,* 252–264.

Dye, D., Reck, M., & Murphy, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment, 1,* 153–157.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods, 6,* 161–180.

Field, A. P. (2005). Is the meta-analysis of correlation coefficient accurate when population correlations vary? *Psychological Methods, 10,* 444–467.

Forsyth, R. A., & Feldt, L. S. (1969). An investigation of empirical sampling distribution of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement, 29,* 61–71.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Benson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72,* 493–511.

Hall, S. M., & Brannick, M. T. (2002). Comparison of two random effects methods of meta-analysis. *Journal of Applied Psychology, 87,* 377–389.

Hedges, L. V. (1988). Meta-analysis of test validity studies: Some new approaches. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 191–212). Hillsdale, NJ: Erlbaum.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Hunt, M. (1997). *How science takes stock.* New York: Russell Sage Foundation.

Hunter, J. E., & Hunter, R. F. (1984). The validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–99.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (1st ed.). Newbury Park, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275–292.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Coggin, T. D. (1996). *Meta-analysis of correlations: Bias in the correlation coefficient and the Fisher z transformation.* Unpublished manuscript, University of Iowa.

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

Iffaldano, M. T., & Muchinsky, P. M. (1985). Job satisfaction and job performance: A meta-analysis. *Psychological Bulletin, 97,* 251–273.

Judge, T. A., Thorensen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin, 127,* 376–401.

Kelly, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.

Lee, P. M. (1989). *Bayesian statistics: An introduction.* New York: Halsted Press.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1985). A meta-analysis of the validity of training and experience ratings in personnel selection. *Personnel Psychology, 41,* 283–314.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599–616.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research.* Washington, DC: National Academy of Sciences Press.

Novick, M. R., & Jackson, P. H. (1974). *Statistical methods for educational and psychological research.* New York: McGraw-Hill.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection. *Journal of Applied Psychology Monograph, 78,* 679–703.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training criteria in clerical occupations. *Journal of Applied Psychology, 65,* 373–407.

Petty, M. M., McGee, G. W., & Cavender, J. W. (1984). A meta-analysis of the relationship between individual job satisfaction and individual job performance. *Academy of Management Review, 9,* 712–721.

Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Methods, 27,* 52–72.

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

Raju, N. S., & Drasgow, F. (2003). Maximum likelihood estimation in validity generalization. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 263–285). Mahwah, NJ: Erlbaum.

Raudenbush, S. W., & Bryk, S. A. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10,* 75–98.

Roth, P. L., BeVier, C., Switzer, F. S., III, & Shippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81,* 548–556.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75,* 175–184.

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

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

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., & Hunter, J. E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975–2002. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31–66). Hillsdale, NJ: Erlbaum.

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 model. *Personnel Psychology, 32,* 257–281.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. A. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78,* 3–13.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches.* New York: Hogrefe & Huber.

Appendix A

An Alternative Bayesian Procedure for Updating Meta-Analysis

This appendix presents an alternative Bayesian approach for both performing and updating a meta-analysis. Let k be the number of (primary) validation studies. Let the individual correlations be denoted by r_i and their sampling variances by $V(r_i)$. Within the medical approach, an estimate of the mean of ρs ($\hat{\bar{\rho}}$) can be expressed as

$$ {}_{(k)}\hat{\bar{\rho}} = \frac{\dfrac{r_1}{V(r_1)} + \dfrac{r_2}{V(r_2)} + \ldots + \dfrac{r_k}{V(r_k)}}{\dfrac{1}{V(r_1)} + \dfrac{1}{V(r_2)} + \ldots + \dfrac{1}{V(r_k)}} = \frac{\displaystyle\sum_{i=1}^{k}\left(\frac{r_i}{V(r_i)}\right)}{\displaystyle\sum_{i=1}^{k}\left(\frac{1}{V(r_i)}\right)} \quad \text{(A1)} $$

The prescript (k) in the above equation is used to denote that an estimate of the mean of ρs is based on k primary studies. An estimate of the sampling variance of ${}_{(k)}\hat{\bar{\rho}}$ may be expressed as

$$ \hat{V}({}_{(k)}\hat{\bar{\rho}}) = \left[\sum_{i=1}^{k}\left(\frac{1}{V(r_i)}\right)\right]^{-1} \quad \text{(A2)} $$

This estimate of $\bar{\rho}$ is the one that is obtained by computing the meta-analysis using the data from all k studies. The variance estimate in Equation A2 is not an estimate of $\hat{\sigma}^2_{\rho}$ but instead is the squared standard error of the estimate of $\bar{\rho}$. Because we are interested in a random-effects model, this is the random-effects standard error. That is, $V(r_i)$ in Equation A2 is defined as in Equations 5 and 9 in the text.

Bayesian Procedure

Let us say that that an estimate of the prior distribution is based on the first $k - 1$ studies. Therefore, an estimate of the prior mean of ρs based on the first $k - 1$ studies can be expressed as

$$ {}_{(k-1)}\hat{\bar{\rho}} = \frac{\dfrac{r_1}{V(r_1)} + \dfrac{r_2}{V(r_2)} + \ldots + \dfrac{r_{k-1}}{V(r_{k-1})}}{\dfrac{1}{V(r_1)} + \dfrac{1}{V(r_2)} + \ldots + \dfrac{1}{V(r_{k-1})}} = \frac{\displaystyle\sum_{i=1}^{k-1}\left(\frac{r_i}{V(r_i)}\right)}{\displaystyle\sum_{i=1}^{k-1}\left(\frac{1}{V(r_i)}\right)}. \quad \text{(A3)} $$

with the estimate of the sampling variance of the mean estimate being given by

$$ \hat{V}({}_{(k-1)}\hat{\bar{\rho}}) = \left[\sum_{i=1}^{k-1}\left(\frac{1}{V(r_i)}\right)\right]^{-1}. \quad \text{(A4)} $$

Let us now say that we have a new validation study, Study k, with a validity coefficient denoted by r_k and its sampling variance by $V(r_k)$. We would like to update our prior to get the posterior mean estimate and its sampling variance. Using Equations 6 and 7 in the text, we can express the posterior mean estimate (without any prescript) as

$$ \hat{M}_{\rho} = \frac{\dfrac{{}_{(k-1)}\hat{\bar{\rho}}}{\mathrm{V}({}_{(k-1)}\hat{\bar{\rho}})} + \dfrac{r_k}{V(r_k)}}{\dfrac{1}{V({}_{(k-1)}\hat{\bar{\rho}})} + \dfrac{1}{V(r_k)}}. \quad \text{(A5)} $$

Using Equation A4, we can rewrite Equation A5 as

$$ \hat{M}_{\rho} = \frac{{}_{(k-1)}\hat{\bar{\rho}}\left[\displaystyle\sum_{i=1}^{k-1}\frac{1}{V(r_i)}\right] + \dfrac{r_k}{V(r_k)}}{\left[\displaystyle\sum_{i=1}^{k-1}\frac{1}{V(r_i)}\right] + \dfrac{1}{V(r_k)}}. \quad \text{(A6)} $$

Furthermore, in view of Equation A3,

$$ {}_{(k-1)}\hat{\bar{\rho}}\left[\sum_{i=1}^{k-1}\frac{1}{V(r_i)}\right] = \sum_{i=1}^{k-1}\left(\frac{r_i}{V(r_i)}\right), \quad \text{(A7)} $$

Substituting Equation A7 into Equation A6, one obtains

$$ \hat{M}_{\rho} = \frac{\left[\displaystyle\sum_{i=1}^{k-1}\frac{r_i}{V(r_i)}\right] + \dfrac{r_k}{V(r_k)}}{\left[\displaystyle\sum_{i=1}^{k-1}\frac{1}{V(r_i)}\right] + \dfrac{1}{V(r_k)}} = \frac{\left[\displaystyle\sum_{i=1}^{k}\frac{r_i}{V(r_i)}\right]}{\left[\displaystyle\sum_{i=1}^{k}\frac{1}{V(r_i)}\right]} \quad \text{(A8)} $$

which is identical to Equation A1. Hence, the $\bar{\rho}$ estimate obtained by updating the meta-analysis via this Bayesian procedure is the same $\bar{\rho}$ estimate that one obtains when one reruns the meta-analysis after adding the new study—that is, the same mean produced by the medical model.

Weighting

In the process of obtaining an estimate of the mean of ρs, each observed correlation is weighted inversely by its (random-effects) sampling variance (the sampling variance in Equations 5 and 9 in the text). This type of weighting is commonly used in random-effects models (e.g., see Hedges & Vevea, 1998), and it is consistent with Equations 6 and 7 in this article. However, weighting by study sample size would produce very similar final results, as shown by Field (2005). In an extensive simulation study, Field compared weighting by the inverse of random-effects sampling error variance (i.e., by the inverse of Equations 5 or 9) with weighting by sample size and found that these study weights produced almost identical results.

In weighting the prior mean, the (inverse of) sampling variance of the estimate of the prior mean is used rather than the (inverse of) variance of ρs. The basis for this is that the prior mean is an

(Appendixes continue)

estimate and therefore has its own sampling error. This weighting produces essentially the same mean final (posterior) estimate of $\bar{\rho}$ as the medical model.

The above equations hold for the fixed-effects model as well as the random-effects model. The difference is that in the fixed-effects model, the sampling variances do not include the variance of ρs. However, the fixed-effects model is rarely, if ever, appropriate in meta-analysis (National Research Council, 1992).

In the above derivation, it is assumed that the observed correlations are not corrected for unreliability and range restriction. Therefore, these equations deal only with the so-called bare-bones meta-analysis, the results of which are usually of limited informational value. The procedure can be expanded, however, to apply to corrected correlations. Doing this requires correcting each correlation in the previous meta-analysis individually and properly computing the sampling error variance of each corrected correlation. Making such corrections requires, for each correlation in the previous meta-analysis, knowledge of the reliability of both variables and the degree of range restriction, in addition to the information required for a bare-bones application (sample sizes and correlations). However, the application of the medical model also requires these items of information when correlations (or *d* values) are corrected individually. This procedure can also be extended to applications in which corrections for artifacts are made via artifact distributions, but that extension is beyond the scope of this article.

Appendix B

An Estimate of the Posterior Variance of ρs

The procedure described in Appendix A provides an estimate of $\bar{\rho}$ but not an estimate of σ^2_ρ. The following equations develop a procedure for estimating σ^2_ρ. In the following equations, sampling variances of the form $V(\hat{M}_\rho)$ are random-effects sampling error variances (see Equations 5 and 9 in the text), whereas sampling error variances of the form $V(r_i)$ are fixed-effects (simple) sampling error variances (see Equation 1 in the text).

Let ${}_{post}\hat{\bar{\rho}}$ represent the posterior estimate of the mean of ρs and let $V({}_{post}\hat{\bar{\rho}})$ be the (random-effects) sampling error variance of this estimate of the mean. This variance, by definition, can be expressed as

$$V({}_{post}\hat{\bar{\rho}}) = \frac{{}_{post}\sigma^2_\rho + Ave_k[V(r_i)]}{k}, \tag{B1}$$

where k is the total number of primary studies, $V(r_i)$ is the simple (fixed-effects) sampling variance of r_i, and $Ave(_k)$ is the average based on all k studies. The above equation can be rewritten as

$${}_{post}\sigma^2_\rho = (k)V({}_{post}\hat{\bar{\rho}}) - Ave_{(k)}[V(r_i)]. \tag{B2}$$

We now recall that the variance of the estimate of the prior mean of ρs, which is known, can be expressed as

$$V({}_{prior}\hat{\bar{\rho}}) = \frac{{}_{prior}\sigma^2_\rho + Ave_{(k-1)}[V(r_i)]}{k-1}, \tag{B3}$$

which is based on $k - 1$ studies. Also, all elements on the right-hand side of Equation B3 are known. Rewriting Equation B3, one gets

$$Ave_{(k-1)}(V(r_i)) = (k-1)V({}_{prior}\hat{\bar{\rho}}) - {}_{prior}\sigma^2_\rho. \tag{B4}$$

Now,

$$Ave_{(k)}[V(r_i)] = \frac{(k-1)Ave_{(k-1)}[V(r_i)] + V(r_k)}{k}. \tag{B5}$$

Using Equation B5, one can rewrite Equation B2 as

$${}_{post}\sigma^2_\rho = (k)V({}_{post}\hat{\bar{\rho}}) - \frac{(k-1)Ave_{(k-1)}[V(r_i)] + V(r_k)}{k}. \tag{B6}$$

In view of Equation B4, Equation B6 can be rewritten as

$${}_{post}\sigma^2_\rho = \frac{(k-1)({}_{prior}\sigma^2_\rho) + [k^2V({}_{post}\hat{\bar{\rho}})) - (k-1)^2V({}_{prior}\hat{\bar{\rho}})] - V(r_k)}{k} \tag{B7}$$

According to Equation B7, the posterior variance of ρs depends on the prior variance of ρs, prior and posterior variances of the estimates of the mean of ρs, and the sampling variance of the newly added validity coefficients. Equations B6 and B7 provide an estimate of the posterior variance of ρ which is the same (or very nearly the same) value that would be obtained from the medical model. This estimate of the posterior variance of ρ is invariant with respect to which study is the new study, just as is the medical model. As with the equations in Appendix A, these equations are presented for the case of bare-bones meta-analysis (no corrections for artifacts) but can be extended to meta-analyses that correct for the biasing effects of artifacts (complete meta-analyses).

Received December 1, 2004
Revision received March 2, 2006
Accepted March 27, 2006 ■