

Not Restricted to Selection Research: Accounting for Indirect Range Restriction in Organizational Research

Organizational Research Methods
1-33

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1094428119859398

journals.sagepub.com/home/orm



Jeffrey A. Dahlke¹  and Brenton M. Wiernik²

Abstract

Range restriction is a common problem in organizational research and is an important statistical artifact to correct for in meta-analysis. Historically, researchers have had to rely on range-restriction corrections that only make use of range-restriction information for one variable, but it is not uncommon for researchers to have such information for both variables in a correlation (e.g., when studying the correlation between two predictor variables). Existing meta-analytic methods incorporating bivariate range-restriction corrections overlook their unique implications for estimating the sampling variance of corrected correlations and for accurately assigning weights to studies in individual-correction meta-analyses. We introduce new methods for computing individual-correction and artifact-distribution meta-analyses using the bivariate indirect range restriction (BVIRR; “Case V”) correction and describe improved methods for applying BVIRR corrections that substantially reduce bias in parameter estimation. We illustrate the effectiveness of these methods in a large-scale simulation and in meta-analyses of expatriate data. We provide R code to implement the methods described in this article; more comprehensive and robust functions for applying these methods are available in the *psychmeta* package for R.

Keywords

meta-analysis, reliability and validity, construct validation procedures

Range variation, also called selection bias or collider bias (Elwert & Winship, 2014), is a common statistical phenomenon in which the variance of a variable in a sample is unrepresentative of the variance of that variable in the desired target population (Sackett & Yang, 2000; e.g., the variance in samples of job incumbents is smaller than the variance in the job applicant population from which the incumbents were selected). Range variation is called *range restriction* when variance is smaller

¹Department of Psychology, University of Minnesota, Minneapolis, MN, USA

²Department of Psychology, University of South Florida, Tampa, FL, USA

Corresponding Author:

Jeffrey A. Dahlke, Human Resources Research Organization, 66 Canal Center Plaza, Suite 700, Alexandria, VA 22314, USA.
Email: jdahlke@humrro.org

in a sample than in the target population and *range enhancement* when variance is larger in a sample than in the target population (Schmidt & Hunter, 2015). Generally speaking, range restriction occurs when the tail(s) of a distribution is (are) underrepresented, whereas range enhancement occurs when the tail(s) of a distribution is (are) overrepresented. The “range restriction” terminology is more familiar to researchers in organizational, psychological, and educational research, so we default it in this article except where it becomes necessary to invoke range enhancement specifically, although all procedures described here apply equally well to both types of range variation. Along with sampling error and measurement error, range variation is one of the most important statistical artifacts to correct in psychometric meta-analysis to obtain unbiased estimates of the mean effect and accurately estimate the extent of effect-size heterogeneity (Hunter, Schmidt, & Le, 2006; Schmidt & Hunter, 2015). In this article, we describe the relevance of range-variation corrections for organizational research and present new correction methods when range-variation information is known for both variables.

Corrections for range restriction are common in some areas of organizational, psychological, and educational research. For example, primary studies and meta-analyses in the personnel selection and staffing literatures frequently correct for range restriction in predictor constructs (Le & Schmidt, 2006; Sackett, Lievens, Berry, & Landers, 2007). In economics and some areas of strategy, Heckman’s (1979) correction for selection bias, which is fundamentally a range-restriction correction, is highly influential and widely used. However, corrections for range restriction are rare in other literatures, even though they may be highly relevant.

For example, researchers might be interested in the relationship between leader consideration behaviors and follower satisfaction (DeRue, Nahrgang, Wellman, & Humphrey, 2011). One study using a heterogeneous sample of employees from many organizations might find a strong relationship between these variables, whereas another study using a single-organization sample finds a negligible relationship. It is possible that much of this difference in results is caused by the second sample being more restricted both on levels of leader consideration and on employee satisfaction than in the heterogeneous sample. Differences in variance between the two types of samples may cause conflicting conclusions regarding the strength of the relationship of interest; it is therefore important that researchers consider the potential influence of range-variation artifacts in addition to substantive theoretical explanations when they interpret variations in findings across studies.

Similarly, a meta-analysis of the relationship between job enrichment and employee well-being might observe that the relationship between enrichment and well-being is highly variable across studies, even after accounting for sampling error and measurement error (Humphrey, Nahrgang, & Morgeson, 2007). Researchers might hypothesize substantive moderators for this relationship, but an alternative explanation could be that the included samples differ in their mean levels and variability on job enrichment and well-being. Accounting for this differential variability could reveal there is little or no true residual variability. In this case, uncorrected range-restriction artifacts could lead to an unnecessary search for moderators and cause much wasted effort.

At the firm level, a researcher might examine the relationship between organizational culture variables and organizational financial performance. If they find no relationship, this may be because their sample of organizations is restricted on either culture or performance and thus not representative of the broader market. Accounting for the nonrepresentativeness of between-firm variability may support different conclusions than would be possible if only the unrepresentative sample data were analyzed. If unrepresentative variances are a plausible cause of the unexpected finding, more representative sampling of the broader market may be warranted to more accurately estimate the firm-level relationship between culture and performance.

Although range restriction is often thought of as an artifact that simply attenuates effect-size estimates, selection effects can also give rise to negative associations where null or positive associations should exist (or vice-versa). For example, Murray, Johnson, McGue, and Iacono (2014)

convincingly demonstrated that the reason general mental ability and conscientiousness have exhibited a negative relationship across many studies is that those studies were based on analyses of range-restricted data from college students. College students are selected based on achievement (among other things), which is positively correlated with both ability and conscientiousness. This selection process (called “conditioning on a collider”; Rohrer, 2018) can induce a negative relationship between variables with a zero or even positive true correlation. Murray et al. showed that ability–conscientiousness correlations were zero or positive in databases without achievement-related selection effects, but became negatively correlated as the researchers created range restriction in their data by only analyzing data from high-achievement individuals. This finding refutes the intelligence compensation hypothesis (i.e., the idea that negative ability–conscientiousness correlations are due to conscientiousness emerging as a way to compensate for lower ability) and instead suggests statistical artifacts were the cause of the phenomenon. Thus, failing to account for range-restriction effects in samples from selective environments runs the risk of misestimating the directions of associations and causing wasted effort in the development of a theory to explain an artifactual trend.

Correcting for range restriction can also have a role in analyses conducted as part of meta-analysis, such as heterogeneity analyses (e.g., estimating SD_{ρ} , SD_{δ} , τ , credibility intervals, and Q) and sensitivity analyses (e.g., leave-one-out meta-analyses, bootstrapped meta-analyses, and publication or small-sample bias analyses). These analyses are best performed on effect sizes corrected for artifacts, as the corrected effect sizes are better representations of the parameter distribution of interest. Conducting analyses using observed effect sizes could provide misleading indications of heterogeneity or sensitivity, as artifacts like range restriction may explain a great deal of the variation across studies (Wiernik & Dahlke, in press). Many publication bias models also assume homogeneity of the effect size parameters across studies. Range restriction introduces artifactual heterogeneity, causing analyses of observed effect sizes to violate this assumption and potentially suggesting the presence of publication bias where none exists (or vice-versa).

Additionally, comparing the results of meta-analyses of observed effect sizes to meta-analyses of artifact-corrected effect sizes provides the meta-analyst with information about the potential prevalence and impact of range-restriction effects. If the observed and corrected results are highly discrepant, the researcher has evidence to argue that selection effects are operating in his or her sample of studies and can indicate to other scholars that these artifacts should be addressed when designing sampling strategies for future studies conducted in that domain.

Approaches to Correcting for Range Restriction

Researchers are generally aware of the negative consequences of sampling biases (Rousseau & Fried, 2001), so we suspect that much of the reluctance to consider and correct for range restriction stems from a perception that the statistical information needed to apply such corrections is not widely available. However, this information may be more widely available than many researchers realize. Table 1 describes five approaches for selecting a target population to use in range-variation corrections and methods for estimating the amount of range restriction (as indexed by a u ratio; the ratio of the sample standard deviation to the target population standard deviation) from different sources of information. In the classic range-restriction scenario (Approach 1), local reference group information is available (e.g., job applicants, pre-intervention measurements, or control groups), of which the restricted sample is a subset. In this case, u ratios are computed by comparing restricted sample standard deviations or reliability coefficients to corresponding values in the unrestricted reference samples.¹ In Approach 2, researchers wish to generalize to a broader general or norm population (e.g., managers, all U.S. organizations). In this approach, u ratios are computed by comparing each sample’s SD or reliability to the corresponding value for the measure’s development or norm sample reported in

Table 1. Approaches for Selecting a Target Population and Estimating Range-Restriction u Ratios in Organizational Research.

Approach	Target Population	Source of Reference Sample Information	When Method Is Applicable	u Ratio Formulas
(1)	Population from which unrestricted samples were drawn	Local reference group (e.g., job applicants, preintervention groups)	Generalization to the unrestricted population is desired; local reference group information is available for many samples	1a) $u_{local} = SD_i / SD_{reference}$ 1b) $u_{local} = \sqrt{(1 - r_{XX_{reference}}) / (1 - r_{XX})}$
(2)	General or norm population	Measure norm or development sample information	Generalization to the norm/development population is desired; same or different measures used across samples	2a) $u_{norm} = SD_i / SD_{norm}$ 2b) $u_{norm} = \sqrt{(1 - r_{XX_{norm}}) / (1 - r_{XX})}$
(3)	Average sample included in meta-analysis	Average distribution from samples included in meta-analysis	Corrections only for differential variability are needed, no generalization to another population; same measures used across most or all samples	3a) $SD_{pooled} = \sqrt{\sum [(N_j - 1) SD_j^2] / \sum (N_j - 1)}$ $u_{pooled} = SD_i / SD_{pooled}$
(4)	Average sample included in meta-analysis	Measure norm or development sample information	Only corrections for differential variability are needed, no generalization to another population; different measures used across samples, but with comparable development or norm samples ^a	3b) $u_{pooled} = \sqrt{(1 - r_{XX_i}) / (1 - r_{XX})}$
(5)	Combined total sample for all samples included in meta-analysis	Pooled distribution from samples included in meta-analysis	Samples have different means, generalization to the total range represented across samples is desired; same measures used across most or all samples	4) Compute u_{norm} using formula 2a or formula 2b. $u_{norm} = \sum (N_j u_{normj}) / \sum N_j$ $u_{pooled} = u_{norm} - u_{norm}$ 5a) $SD_{mixture} = \sqrt{\frac{\sum [(N_j - 1) SD_j^2 + N_j (\bar{X}_j - \bar{X}_{grand})^2]}{\sum (N_j - 1)}}$ where $\bar{X}_{grand} = \sum (N_j \bar{X}_j) / \sum N_j$ $u_{mixture} = SD_i / SD_{mixture}$
5b) Compute $\bar{u}_{mixture}$ using $u_{mixture}$ values computed with formula 5a. $r_{XX_{mixture}} = 1 - \bar{u}_{mixture}^2 (1 - r_{XX})$ $u_{mixture} = \sqrt{(1 - r_{XX_{mixture}}) / (1 - r_{XX})}$				

Note: When possible, computing u ratios from standard deviations is preferable to computing from reliability coefficients.^aThis method should be used only in artifact-distribution, not individual-correction, meta-analyses with BVIRR corrections. Reliability coefficient-based formulas for u based on Le, Oh, Schmidt, and Woodlridge (2016, p. 1005).

the measure's test manual or other report (however, the researcher must be cautious and ensure that the reference sample from such a source is relevant to the population of interest in their own study). For variables such as age or organizational financial performance, population values can be drawn from the census, government labor statistics, financial statistics databases, or similar sources. In Approaches 3 and 4, a meta-analyst is concerned that differential variability across samples contributing to artifactual heterogeneity in effect sizes. They wish to remove this variability without affecting the mean effect size. If the same measures are used in all or most of the included studies (e.g., Colquitt et al., 2013; Wiernik & Kostal, 2019), u ratios can be computed by comparing each sample to the average ("pooled") within-sample SD or reliability coefficient (Approach 3). If a wide variety of measures are used across studies,² u ratios can be computed by comparing sample SD or reliability values to measure-development or norm samples (as in Approach 2), then recentring the resulting u_{norm} distribution to have a mean $u = 1.0$ (Approach 4). Finally, if samples in a meta-analysis differ in their mean levels of a variable and a meta-analyst wishes to generalize to the broader combined total sample of all studies included in a meta-analysis, u ratios can be computed by comparing each sample SD or reliability value to the corresponding total ("mixture") values that combine both within-sample and between-sample variance on variables.

Each of the five approaches described in Table 1 share a common caveat—the u ratios that result from them will only be as relevant as the samples and data from which they are derived. None of these approaches offers a magic-bullet solution for computing a u ratio if no data on a relevant target population SD or reliability coefficient is available: It is incumbent upon the researcher using any of these approaches to ensure that the SD or reliability estimates for target/reference samples represent the population to which the researcher wishes to generalize. Provided that one has vetted the relevance of the samples contributing to the denominators of one's u ratios, the five approaches from Table 1 show that range-restriction corrections can be accessible to researchers in a wide variety of primary and meta-analytic research settings. In the following sections, we present new methods for correcting the most frequently encountered form of range restriction in organizational, psychological, and educational research: indirect range restriction.

Correcting for Range Restriction

There are several equations available to correct for range restriction (see Sackett & Yang, 2000, for a review), each of which is a special case of the multivariate selection theorem presented by Aitken (1934) and Lawley (1943). Historically, most researchers have corrected correlations for range restriction using univariate-correction formulas that assume that selection effects are fully explained by only one of the variables involved in the correlation of interest. This assumption is commonly violated when selection processes differentially impact the two variables, and these univariate corrections are unnecessarily restrictive when range-restriction information is available for both variables in a correlation. In this article, we consider the unique challenges of correcting for such cases of *bivariate indirect range restriction* (BVIRR) that are not addressed by traditional psychometric meta-analytic methods, and we develop new correction methods to obtain more accurate parameter estimates in primary studies and meta-analyses. We use simulations and real-data illustrations to demonstrate the accuracy of our new methods and compare them to existing approaches.

A key task for range-restriction corrections is to accurately model the selection process(es) that caused the observed sample variance to be unrepresentative of the population. The commonly applied correction for univariate direct range restriction (UVDRR [Case II]; see Figure 1A for a schematic representation)³ assumes that selection occurs directly on one variable (X) involved in a correlation (e.g., selecting job applicants based on extraversion scores directly restricts the variance of extraversion and attenuates the extraversion–job satisfaction correlation; see Figure 1A) and

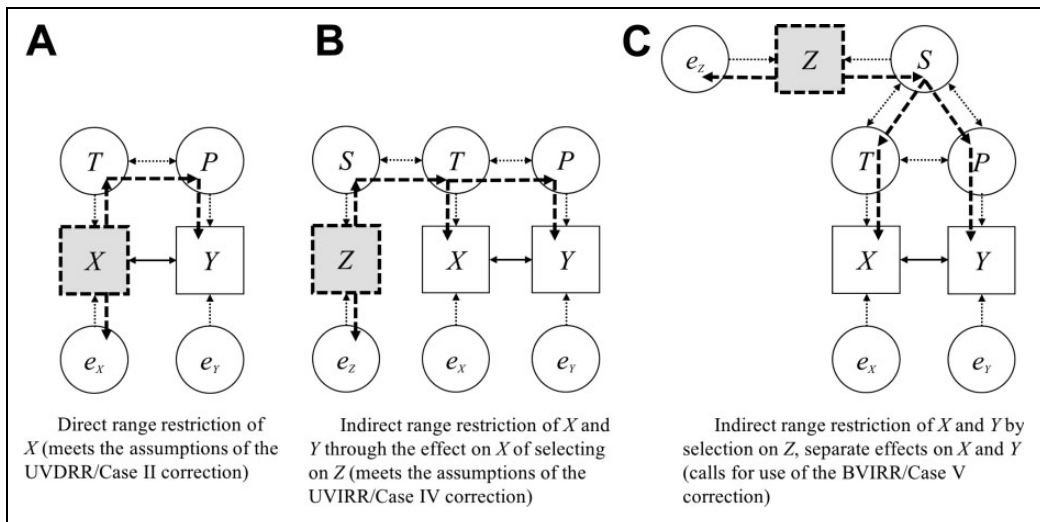


Figure 1. Schematic representations of direct and indirect range-restriction mechanisms assumed by the (A) UVDRR/Case II correction, (B) UVIRR/Case IV correction, and (C) BVIRR/Case V correction. Note that the BVIRR correction can be used in any of these three scenarios, provided that one has u ratios for both X and Y . X and Y are the manifest variables of interest in these scenarios and T and P are their respective latent construct counterparts. Z is a potentially unobserved manifest variable that represents the applicant-suitability construct S . The e_x , e_y , and e_z variables are the measurement-error variances of X , Y , and Z , respectively. Explicit selection on X causes direct range restriction in the X – Y relationship, whereas explicit selection on Z causes indirect range restriction in the X – Y relationship. A variable in a dashed, shaded box indicates a selection variable. Solid lines represent observed X – Y relationships. Dotted lines represent correlational relationships among latent variables when paths have two arrowheads and causal effects of latent variables on manifest variables when paths have one arrowhead. Heavy dashed lines show the paths through which explicit selection performed on a manifest variable affects other variables, both latent and observed. Relationships between X and Z , Y and Z , X and S , Y and S , Z and T , and Z and P exist but are not shown because Z is presumed to be an unknown or unmeasured variable that is not of substantive interest when one is studying the relationships between X and Y and between T and P .

corrects the correlation using the u ratio for X (i.e., the ratio of SD_X in the selected sample to SD_X in the target population).

Hunter et al. (2006) presented methods for the more common case of univariate *indirect* range restriction (UVIRR [Case IV]) wherein selection occurs on a third variable, but the range restriction of true-score variance for one of the variables involved in the correlation of interest fully accounts for the effect of range restriction on the other variable (see Figure 1B). Hunter et al.'s UVIRR correction is structurally very similar to the UVDRR correction, but it assumes that the effect of selection on an unmeasured third variable (Z) on one observed variable (Y) in a correlation is fully mediated by the other variable in the correlation (X) for which a u ratio is known (e.g., if selection decisions are based on interview ratings that have no direct association with job satisfaction after controlling for extraversion, correcting only for range restriction of extraversion will permit accurate estimation of the extraversion–job satisfaction correlation; Figure 1B). Indirect range restriction is transmitted through true-score correlations because error scores are assumed to be independent of other variables. Both the UVDRR and UVIRR corrections assume that selection effects are fully explained by only one of the variables in the target correlation. This assumption is commonly violated when selection processes differentially impact the two variables involved in a correlation. Violating UVIRR's mediation assumption reduces the correction's accuracy (e.g., if the interview

ratings in our previous example have a relationship with job satisfaction that is not accounted for by extraversion, the UVIRR correction will be less accurate); however, even in such cases, the UVIRR correction is more accurate than the UVDRR correction (Beatty, Barratt, Berry, & Sackett, 2014; Le & Schmidt, 2006).

Even more accurate corrections for range restriction can be made using methods that do not share UVIRR's mediation assumption. Such methods require that u ratios are available for both variables in the target correlation. Alexander, Carson, Alliger, and Carr (1987) presented a correction for direct range restriction on both variables in a correlation (i.e., *bivariate direct range restriction*; BVDRR). Schmidt and Hunter (2015) reported that this correction performs reasonably well in meta-analyses even when the selection process is indirect. Alexander (1990; based on earlier work by Bryant & Gokhale, 1972) presented a similar correction for correlations that have been restricted via selection on a single unknown third variable when u ratios are available for both variables in the correlation; we refer to this correction as the *bivariate indirect range restriction* (BVIRR) correction. This correction is potentially useful in a wide range of research and practice settings where range restriction affects both variables in a correlation through an unknown selection mechanism. Le, Oh, Schmidt, and Wooldridge (2016) combined Alexander's (1990) formula with corrections for measurement error and proposed methods for applying the BVIRR correction (which they called "Case V") in meta-analysis. Le et al. proposed the use of conventional individual-correction and interactive artifact-distribution psychometric meta-analysis procedures (see Chapters 3 and 4 of Schmidt & Hunter, 2015) with the BVIRR correction; however, Le et al.'s methods overlooked important effects of the BVIRR correction on the sampling error of corrected correlations and the unique corresponding implications of applying the BVIRR correction in primary research and meta-analyses. In this article, we present a generalized form of the BVIRR correction and describe new methods to accommodate its impacts on the sampling variance of correlations.

Correcting for Bivariate Indirect Range Restriction

When applying the BVIRR correction, one is interested in the target population-level correlation between the constructs T and P , which are represented by the imperfectly measured variables X and Y , respectively. While T and P are the constructs of interest, a third construct called S is also important because it represents a "suitability" construct or a screening process through which selection affects the variances of T and P (see Figure 1C). S need not be represented by a known or measured variable (or set of variables)—its effects on X and Y can be inferred from the observed variables' u ratios without knowing the exact nature of the selection process (the only requirement is that one must know the signs of S 's correlations with T and P). The key advantage of the BVIRR correction over the UVIRR correction is that the BVIRR correction permits selection on S to have separate direct effects on both T and P , rather than assuming the effect of selection on P is fully mediated by T (cf. Figure 1B and Figure 1C).

Based on the selection path model shown in Figure 1C, a general form of the BVIRR correction can be written as:⁴

$$\rho_{TP_a} = \frac{\rho_{XY_i} u_X u_Y + \lambda \sqrt{|1 - u_X^2| |1 - u_Y^2|}}{\sqrt{\rho_{XX_a} \rho_{YY_a}}}, \quad (1)$$

with a corresponding attenuation formula written as:

$$\rho_{XY_i} = \frac{\rho_{TP_a} \sqrt{\rho_{XX_a} \rho_{YY_a}} - \lambda \sqrt{|1 - u_X^2| |1 - u_Y^2|}}{u_X u_Y}, \quad (2)$$

Table 2. Decision Rules for Using the BVIRR Correction and for Choosing Values of λ .

Values of u Ratios	Correlations with Selection Mechanism S	Suggested Value of λ or Other Course of Action
Both < 1	ρ_{SP_a} and ρ_{ST_a} have same sign	+1
	ρ_{SP_a} and ρ_{ST_a} have different signs	-1
	ρ_{SP_a} or ρ_{ST_a} equals 0	0
	Both ρ_{SP_a} and ρ_{ST_a} equal 0	Range variation cannot be attributed to S ; identify a different selection mechanism
Both > 1	ρ_{SP_a} and ρ_{ST_a} have same sign	-1
	ρ_{SP_a} and ρ_{ST_a} have different signs	+1
	ρ_{SP_a} or ρ_{ST_a} equals 0	0
	Both ρ_{SP_a} and ρ_{ST_a} equal 0	Range variation cannot be attributed to S ; identify a different selection mechanism
One ≤ 1 and one ≥ 1	Regardless of other considerations, a multivariate correction is the best approach for dealing with this pattern of range variation. If a multivariate approach cannot be used, an approximation may be reached by using the following general-case equation to estimate λ :	
ρ_{SP_a} and ρ_{ST_a} are both nonzero	$\text{sign}[\rho_{ST_a}\rho_{SP_a}(1-u_X)(1-u_Y)]$	$\frac{\text{sign}(1-u_X)\min(u_X, \frac{1}{u_X}) + \text{sign}(1-u_Y)\min(u_Y, \frac{1}{u_Y})}{\min(u_X, \frac{1}{u_X}) + \min(u_Y, \frac{1}{u_Y})}$
ρ_{SP_a} or ρ_{ST_a} equals 0		0
Both ρ_{SP_a} and ρ_{ST_a} equal 0		Range variation cannot be attributed to S ; identify a different selection mechanism

where ρ_{TP_a} is the correlation between X and Y that has been fully corrected for measurement error and range restriction, ρ_{XY_i} is the restricted-group (i.e., range-restricted or enhanced) correlation between X and Y , ρ_{XX_a} and ρ_{YY_a} are unrestricted-group (i.e., target population) reliability values for X and Y , respectively, u_X and u_Y are observed-score u ratios (ratios of SD_i to SD_a)⁵ for X and Y , respectively, and λ is a coefficient that modulates the effect of the u ratios. If unrestricted-group reliability values are unavailable, these can be estimated from restricted-group reliabilities using the formula given by Schmidt and Hunter (2015, p. 127):

$$\rho_{XX_a} = 1 - u_X^2(1 - \rho_{XX_i}) \tag{3}$$

and

$$\rho_{YY_a} = 1 - u_Y^2(1 - \rho_{YY_i}) . \tag{4}$$

In individual studies, the parameters given in Equations 1–4 are estimated using sample statistics (e.g., r_{xy_i} in place of ρ_{XY_i}). In Equations 1 and 2, the value of λ is determined based on the signs of the correlations between S and T and between S and P and whether the variances of the X and Y variables are restricted or enhanced. A summary of decision rules for the value of λ is shown in Table 2. If both X and Y are range restricted (i.e., $u < 1$), λ equals +1 when ρ_{SP_a} and ρ_{ST_a} (i.e., unrestricted correlations of S with the constructs of interest) have the same sign and -1 when they have different signs (Alexander, 1990; Le et al., 2016). If both X and Y are range enhanced (i.e., $u > 1$), the patterns of λ values are reversed: λ equals -1 when ρ_{SP_a} and ρ_{ST_a} have the same sign and +1 when they have different signs.

In the case that one u ratio indicates range restriction and the other indicates range enhancement (either due to complex/multiple selection mechanisms or simply due to sampling error in the observed standard deviations),⁶ a multivariate range-variation correction is most appropriate (Lawley, 1943). However, if insufficient information is available to perform such a correction, an approximate correction can be made using the BVIRR correction with λ determined by the relative

magnitudes of range restriction and range enhancement for X and Y , as well as the patterns of ρ_{ST_a} and ρ_{SP_a} correlations:⁷

$$\lambda = \text{sign}[\rho_{ST_a} \rho_{SP_a} (1 - u_X)(1 - u_Y)] \frac{\text{sign}(1 - u_X) \min\left(u_X, \frac{1}{u_X}\right) + \text{sign}(1 - u_Y) \min\left(u_Y, \frac{1}{u_Y}\right)}{\min\left(u_X, \frac{1}{u_X}\right) + \min\left(u_Y, \frac{1}{u_Y}\right)}. \quad (5)$$

The result of Equation 5 ranges from -1 to $+1$ and reflects a compromise between the opposing effects of range restriction and range enhancement, tilted toward the effect with the stronger impact on the correlation. In the case that both variables are affected by the same type of range variation, Equation 5 reduces to ± 1 as described above.

A final difference between Equation 1 and previously presented BVIRR correction formulas (cf. Alexander, 1990; Le et al., 2016) is the absolute value brackets under the radical in the numerator. These ensure that the value under the radical is positive when a mixed pattern of range restriction and range enhancement is observed for X and Y . Any negative signs under the radical that are lost due to the use of absolute values are accounted for in Equation 5.

Sampling Error and Bivariate Indirect Range-Restriction Corrections

The sampling variance of the observed correlation coefficient is used to compute confidence intervals and significance tests in primary research, as well as to estimate study weights and the sampling error variance of effect sizes in meta-analysis. The sampling distribution of a nonzero Pearson correlation is asymmetric, but its overall variance can be approximated as:

$$\text{var}_e = \frac{(1 - \rho_{XY_i}^2)^2}{N - 1}, \quad (6)$$

where ρ_{XY_i} is the population correlation (typically estimated in meta-analysis by the sample size-weighted mean observed correlation) and N is the sample size of an individual study (Schmidt & Hunter, 2015, p. 101). When psychometric corrections are applied to a correlation, the sampling error variance must also be adjusted to accurately reflect the amount of sampling error associated with the corrected correlation. It is commonly stated that sampling error variance for the corrected correlation (var_e) is always larger than for the observed correlation (e.g., “corrections incur a cost”; Oswald & McCloy, 2003, p. 317). Indeed, the correction equations for measurement error and univariate range restriction (i.e., UVDRR and UVIRR) do imply that the corrections increase sampling error.⁸ However, the BVIRR equation implies that the sampling variances of correlations corrected for range restriction can be *smaller* than the sampling variances of observed correlations. Below, we describe the equation-implied corrected sampling variance, discuss why this is not the most appropriate sampling variance estimator for corrected correlations, and provide a new approach for estimating corrected sampling variances, which we evaluate via a meta-analysis simulation.

Equation-Implied Sampling Variance Estimator for BVIRR

The BVIRR correction is a linear transformation in slope-intercept form. It includes a multiplicative term:

$$\rho_{XY_i} \times \frac{u_X u_Y}{\sqrt{\rho_{XX_a} \rho_{YY_a}}}$$

and an additive term:

$$+ \frac{\lambda \sqrt{|1 - u_X^2| |1 - u_Y^2|}}{\sqrt{\rho_{XX_a} \rho_{YY_a}}} .$$

The additive component of a linear transformation only impacts a variable's mean, not its variance. Accordingly, only the multiplicative term in Equation 1 affects the sampling error variance of the corrected correlation. Multiplying a variable by a scalar (i.e., the value $u_X u_Y / \sqrt{\rho_{XX_a} \rho_{YY_a}}$ in this case) transforms its variance by the squared value of that scalar. Thus, the BVIRR equation-implied variance for the corrected correlation is:

$$\text{var}_{e_c} = \text{var}_e V^2, \quad (7)$$

where

$$V = \frac{u_X u_Y}{\sqrt{\rho_{XX_a} \rho_{YY_a}}} . \quad (8)$$

The equation-implied effect of the BVIRR correction on sampling variance is that, under conditions of range restriction and high-quality measurement, corrected correlations have *smaller* sampling variances than observed correlations because range-restricted u ratios are smaller than 1. This implication is quite strange, as corrections for range-restriction should theoretically increase sampling variances because the correction introduces additional uncertainty into the effect-size estimation process. We consider these additional sources of error momentarily.

We note that Equation 7 differs from previously proposed sampling variance formulas for BVIRR-corrected correlations. Le et al. (2016, p. 983) proposed a var_{e_c} formula analogous to those for other psychometric corrections:

$$\text{var}_{e_c} = \frac{\text{var}_e}{A^2}, \quad (9)$$

where A is the compound attenuation factor that quantifies the combined impact of range restriction and measurement error on the correlation:

$$A = \frac{\rho_{XYi}}{\rho_{TPa}} . \quad (10)$$

Equation 9 assumes that the BVIRR correction affects sampling variances in the same way as corrections for measurement error, UVDRR, and UVIRR. The ρ_{TPa} value in Equation 10 includes the BVIRR correction's additive term and is thus a mathematically inappropriate compound attenuation factor, as the additive term does not have a direct impact on variance. The problem with the attenuation factor in Equation 9 is that it cannot account for the effect of the correction on the sampling distribution of a correlation, particularly when the sampling distribution of the observed correlation distribution spans zero by a non-trivial margin.

There is a nonlinear association between observed correlations and the squared compound attenuation factors used in Equation 9 and, as shown in Figure 2, this nonlinearity becomes most pronounced when the sign of the observed correlation differs from the sign of the unrestricted true-score correlation. The parabolic shape of the associations in Figure 2 means that observed correlations with the same absolute value but different signs will have the same squared compound attenuation factor if they correspond to corrected correlations with the same absolute value. For example, observed correlations of $-.10$ and $+.10$ that both have corrected values of $.20$ will have squared compound attenuation factors equal to $.25$ even though one of the observations incurs a much larger correction than the other. This incorrectly implies that the magnitude of a correction does not directly correspond to the effect of the correction on the estimated sampling error of the corrected value or to the weight assigned to the corrected value in an individual-correction

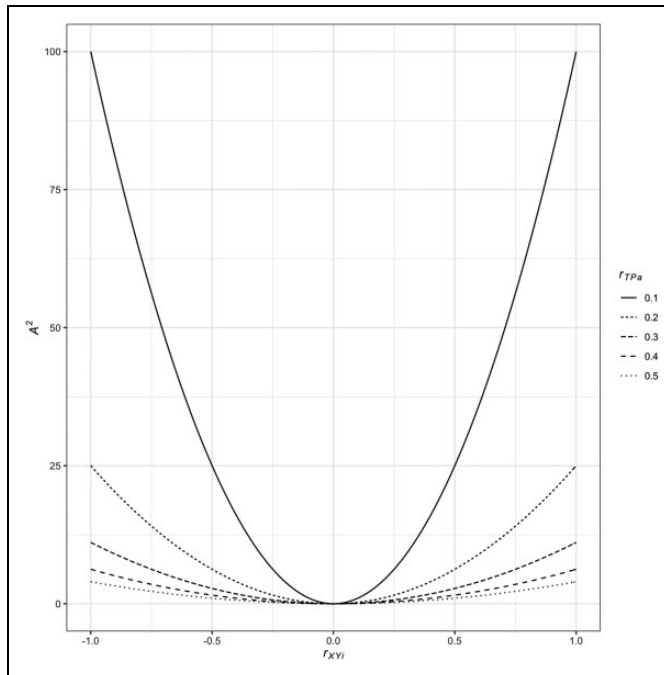


Figure 2. Association between observed correlation values and squared conventional compound attenuation factors (A^2 ; Le, Oh, Schmidt, & Wooldridge, 2016) for unrestricted true-score correlations ranging from .10 to .50.

meta-analysis. As a result, corrected correlations that have a different sign than their observed values will get too much weight in a meta-analysis, biasing both the mean ρ and SD_ρ estimates. This bias will be most pronounced in settings where the mean correlation is closer to zero, the distribution of correlations is more variable, and the mean sample size is smaller (i.e., there is more sampling error): Each of these conditions increases the probability that observed correlations will have different signs than their corresponding corrected correlations. Given that applying corrections contributes new sources of uncertainty to a statistical estimate, the quadratic trends in Figure 2 offer an implausible description of the effects of the BVIRR correction on the precision of corrected estimates. The nonlinearity shown in Figure 2 also affects the compound attenuation factors of other corrections (e.g., UVDRR, UVIRR, measurement error), but the attenuation factors are reasonable approximations for those corrections because those corrections are not able to alter the sign of an effect size.

Apart from the issues caused by the parabolic association between observed correlations and squared compound attenuation factors, another problem with the traditional usage of compound attenuation factors is that they fail to account for how the sampling error of artifacts impacts the uncertainty of corrected correlations. Neither Equation 7 nor Equation 9 accounts for sampling error in the artifacts used to make corrections, which is particularly problematic for BVIRR because of the correction's additive term. In the following section, we describe how to obtain a more accurate estimate of the corrected sampling variance that accounts for uncertainty in the artifact values.

A More Accurate Sampling Variance Estimator for BVIRR

A full account of the effects of the BVIRR correction on correlations' sampling variances must include the sampling error of the artifacts used to make the corrections. This is because observed

artifact values are statistics, not parameters, and are thus estimated with error (cf. Raju, Burke, Normand, & Langlois, 1991). We used the delta method to estimate the sampling variance of BVIRR-corrected correlations, incorporating sampling error in reliability coefficients and u values. The delta method is a widely used technique for approximating the variances of linear and nonlinear functions of multiple variables (Jones & Waller, 2013; Oehlert, 1992). The delta method essentially finds a linear regression function (called a “Taylor Series approximation;” for an introduction, see Stein & Barcellos, 1992, p. 624) that closely resembles the actual function. This linear function is then used to represent the variance of the function as a weighted sum of the input variables’ variances (e.g., the sampling error variances of the observed correlation, the reliability coefficients, and the u ratios). A Taylor series approximation (TSA) of a function’s variance is like linear regression, but instead of computing the regression weights from data, the regression weights are computed using the partial derivatives of the function, differentiating by each input variable. The partial derivatives of a function are simply regression weights that describe how each input is linearly related to the function’s output, with the other inputs held constant. Just as regression weights can be used to compute the variance of fitted criterion values from the variance-covariance matrix of predictor variables, the partial derivatives of a function can be used to estimate the variance of the function’s outputs by computing a linear combination of the inputs’ variances.

Our TSA formula for the sampling variance of a corrected correlation relies on principles of error propagation (Meyer, 1975), used widely in the physical sciences to estimate error in complex measurements, to estimate the combined effects of uncertainty in artifacts and the observed correlation on the total uncertainty in the corrected correlation. Details of this procedure are given in Online Appendices C and D. The TSA formula for the sampling error of a BVIRR-corrected correlation is:

$$\text{var}_{ec} = SE_{\rho_{TPa}}^2 \approx \beta_1^2 SE_{q_{X_a}}^2 + \beta_2^2 SE_{q_{Y_a}}^2 + \beta_3^2 SE_{u_X}^2 + \beta_4^2 SE_{u_Y}^2 + \beta_5^2 SE_{\rho_{XY_i}}^2, \quad (11)$$

where SE is a standard error and $\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 are first-order partial derivatives of Equation 1 with respect to $q_{X_a}, q_{Y_a}, u_X, u_Y,$ and ρ_{XY_i} , respectively. The q_{X_a} and q_{Y_a} terms represent measure quality indices, which are the square roots of the ρ_{XX_a} and ρ_{YY_a} reliability coefficients, respectively, and are interpretable as correlations between true scores and observed scores. By using the squared values of these partial derivatives as weights in a linear combination of sampling-variance estimates, one can more completely account for the sampling error in corrected correlations than is possible with Equations 7 or 9.

We note that Equation 11 assumes that each source of variance is independent, even though they are likely to be correlated to some degree in reality. There are currently no accurate estimators of how the sampling distributions of u ratios, reliabilities, and correlations relate to each other, particularly given the many methods to compute reliability estimates and the variability in whether u ratios are computed using unrestricted standard deviations from the local context (Approach 1) or an external study (e.g., norms from test manuals; Approaches 2–5). We attempted to derive stable estimators of the correlations among the sampling distributions of correlations, reliability indices, and u -ratios via the delta method and the method of moments for products; we found that these estimates were inconsistent in their accuracy due to the difficulty of accurately defining correlations among variables that have been exponentiated (e.g., using square roots, squared terms, or inversions). Fortunately, as demonstrated in our later simulation, we have found that Equation 11, even with its independence assumptions, functions rather well in practice; however, we nonetheless encourage future researchers to derive stable, accurate estimators of the covariances among artifacts’ sampling distributions.

Correcting for Bivariate Indirect Range Restriction in Individual-Correction Meta-Analysis

The more accurate estimate of BVIRR-corrected correlations' sampling variances provided by Equation 11 has implications for individual-correction meta-analyses. Existing procedures for meta-analyses using bivariate range-restriction corrections have ignored artifact sampling error (e.g., Le et al., 2016; Schmidt & Hunter, 2015). Le et al. (2016) outlined a procedure for using the BVIRR correction in individual-correction meta-analyses that parallels procedures for other psychometric corrections. In their procedure, correlations are individually corrected for BVIRR and then meta-analyzed using study weights based on sample sizes and compound attenuation factors (see our Equation 10 for the attenuation factors):

$$w_j = N_j A_j^2 . \quad (12)$$

Study weights in psychometric meta-analysis are intended to reflect the inverse of an effect size's sampling variance (using the mean effect size to estimate the population parameter; Schmidt & Hunter, 2015). As discussed earlier, the BVIRR compound attenuation factor is not directly related to the corrected correlation's sampling variance, so it is unlikely that the weights in Equation 12 would be optimal for individual-correction meta-analyses incorporating BVIRR. More accurate study weights might be based on the equation-implied BVIRR-corrected sampling variance (see Equation 7):

$$w_j = \frac{N_j}{V_j^2} , \quad (13)$$

where dividing the sample size by the squared V coefficient produces a weight analogous to the inverse equation-implied corrected sampling variance. Even more accurate weights may be computed as a function of a pseudo-compound attenuation factor defined by the ratio of the error variance of the observed correlation to the TSA-based error variance of the corrected correlation that incorporates the TSA estimate of sampling error variance in the artifacts:

$$w_j = N_j \frac{\text{var}_{e_j}}{\text{var}_{ec_j}} . \quad (14)$$

where var_{ec_j} is estimated using Equation 11. In Equation 14, we use the pseudo-compound attenuation factor to determine weights rather than simply taking the inverse of the corrected error variance (i.e., $w_j = 1/\text{var}_{ec_j}$) so that the weights will be compatible with the individual-correction weights computed for correlations corrected using other types of artifact corrections (e.g., corrections for UVIRR, UVDRR, or measurement error alone); this approach will allow different corrections to occur within the same meta-analysis (e.g., if some correlations are corrected for BVIRR while others are corrected for UVDRR or UVIRR). Given that they account for sampling error in the artifact estimates, we expect that the TSA-based weights in Equation 14 will be superior for estimating the average corrected sampling variance in a meta-analysis and thus provide more accurate SD_p estimates. It is not immediately clear whether the weights in Equation 13 or Equation 14 will provide more accurate estimates of the mean corrected correlation. The question of which type of weight is best is an empirical one, so we will compare the efficacy of these competing weights in a meta-analysis simulation after describing how BVIRR can be implemented using artifact distributions.

For the remainder of the article, we refer to using the weights from Equation 12 and the sampling variance estimates from Equation 9 as the IC_A method, we refer to using the weights from Equation 13 and the sampling variance estimates from Equation 7 as the IC_V method, and we refer to using the weights from Equation 14 and the sampling variance estimates from Equation 11 as the IC_{TSA}

method. In all cases, the individually corrected mean correlation is estimated as:

$$\widehat{\rho}_{TP_a} = \frac{\sum_{j=1}^k w_j r_{TP_{aj}}}{\sum_{j=1}^k w_j}, \quad (15)$$

where k is the number of studies. The observed standard deviation of individually corrected mean correlations is computed as,

$$SD_{r_{TP_a}} = \sqrt{\frac{\sum_{j=1}^k w_j (r_{TP_{aj}} - \widehat{\rho}_{TP_a})^2}{\sum_{j=1}^k w_j}}, \quad (16)$$

and the residual SD_{ρ} from an individual-correction meta-analysis is estimated as,

$$\widehat{SD}_{\rho_{TP_a}} = \sqrt{SD_{r_{TP_a}}^2 - \frac{\sum_{j=1}^k w_j var_{e_{C_j}}}{\sum_{j=1}^k w_j}}. \quad (17)$$

Note that if the residual variance term under the radical is negative, $\widehat{SD}_{\rho_{TP_a}}$ is set to zero.

Correcting for Bivariate Indirect Range Restriction in Artifact-Distribution Meta-Analysis

Just as artifact sampling error is important when correlations are corrected individually for BVIRR, accounting for artifact sampling error can also improve estimates in artifact-distribution meta-analyses. Reliability coefficients and u ratios are statistics, not parameters, which means that the variance of an artifact distribution used in a meta-analysis is a function of both the true variance of the artifacts as well as the sampling variance of the artifacts. Using observed artifact distributions in meta-analyses overestimates artifact variance and underestimates SD_{ρ} because the sampling error of the artifacts artificially inflates estimates of the amount of variance in correlations that is attributable to measurement error and range restriction. Failure to account for the sampling error in artifact distributions is particularly noticeable for BVIRR corrections compared to other psychometric corrections because the BVIRR correction includes an additive term that is entirely determined by artifact values. Interactive method BVIRR meta-analyses (as presented by Le et al., 2016) subtract the $\lambda\sqrt{|1 - u_X^2||1 - u_Y^2|}$ term from the mean corrected correlation (using the attenuation formula in Equation 2) when estimating the variance attributable to psychometric artifacts, allowing the error variance of artifacts, especially u ratios, to have an exaggerated effect on estimates of SD_{ρ} .

In a simulation study of their interactive method artifact-distribution BVIRR meta-analysis procedure, Le et al. (2016) noted that the method underestimated SD_{ρ} . They attributed this negative bias to correlations among artifacts (p. 987) and recommended dividing the variance attributable to artifacts by 2 before subtracting this term from the variance of observed correlations to estimate SD_{ρ} . While the artifacts may certainly covary, the negative bias more likely stems from artifacts' sampling error variances. Dividing the variance in correlations attributable to artifacts in half may yield a better estimate of artifactual variance in some cases, but it is inadequate as a general method for estimating SD_{ρ} because it does not account for the fact that artifacts' sampling variances are correlated with the sample sizes of the studies from which the artifacts were obtained. Halving the estimate of artifact-induced variance will only work in certain circumstances; in other cases, it will underestimate artifactual variance when meta-analyses have large average sample sizes and it will overestimate artifactual variance when meta-analyses have small average sample sizes. We therefore developed solutions using alternative SD_{ρ} estimators.

A Taylor Series Approximation Estimator for SD_ρ in BVIRR Artifact-Distribution Meta-Analyses

We address the impact of artifact sampling error on BVIRR SD_ρ estimates by introducing a new method to estimate SD_ρ using a Taylor series approximation (TSA). Our TSA method relies on a linear approximation of the effects of correlations' sampling error variances and the variances of artifacts on observed correlations to estimate SD_ρ . TSA procedures have been found to be highly accurate for meta-analyses involving other range-restriction corrections (e.g., Le & Schmidt, 2006; Raju & Burke, 1983). For example, Le and Schmidt (2006) evaluated the accuracy of Hunter et al.'s (2006) TSA procedure for the UVIRR (Case IV) correction and showed that the TSA procedure generally produced accurate results. We have found via simulation that TSA methods converge well with the results of interactive methods, but are more intuitive, less computationally intensive, and allow researchers to incorporate artifact information from more types of sources by virtue of only requiring the mean and variance of each artifact distribution.

The TSA BVIRR method for estimating SD_ρ involves computing a weighted linear combination of the variances of the u ratios and the measure quality indices for X and Y to estimate the amount of variance in distributions of observed correlations that is predictable from artifacts. This predicted artifact variance and the predicted sampling error variance are subtracted from the observed variance of correlations; the remaining residual variance can then be corrected for measurement error and range restriction to estimate the variance of the corrected correlations:

$$\widehat{var}_{\rho_{TP_a}} \approx \left[var_{r_{XY_i}} - var_e - (b_1^2 var_{q_{X_a}} + b_2^2 var_{q_{Y_a}} + b_3^2 var_{u_X} + b_4^2 var_{u_Y}) \right] / b_5^2 \quad (18)$$

In Equation 18, $var_{r_{XY_i}}$ is the sample size-weighted variance of observed correlations, var_e is the predicted sampling error variance of observed correlations computed using Equation 6 with the sample size-weighted mean observed correlation and the mean sample size, and $var_{q_{X_a}}$, $var_{q_{Y_a}}$, var_{u_X} , and var_{u_Y} are the sample size-weighted variances of the distributions of q_{X_a} , q_{Y_a} , u_X , and u_Y sample values, respectively. The b_1 , b_2 , b_3 , b_4 , and b_5 terms are the first-order partial derivatives of Equation 2 with respect to q_{X_a} , q_{Y_a} , u_X , u_Y and ρ_{TP_a} , respectively. The portion of Equation 18 in square brackets represents the residual variance (\widehat{var}_{res}) of observed correlations after accounting for variance attributable to sampling error and artifacts; dividing this value by b_5^2 converts it to the true-score metric using the mean values of all artifacts. Full details of this TSA procedure are given in Online Appendix E. Both this TSA method and Le et al.'s (2016) interactive method assume that q_{X_a} (i.e., $\sqrt{\rho_{XX_a}}$), q_{Y_a} (i.e., $\sqrt{\rho_{YY_a}}$), u_X , u_Y , and ρ_{TP_a} are all independent. We refer to the practice of computing TSA artifact-distribution meta-analyses using observed artifact variances as the AD_{TSA} method.

Equation 18 can be modified to account for the negative-biasing effects of artifact sampling error on SD_ρ estimates. To achieve this, one must first conduct a bare-bones meta-analysis of each artifact and calculate the expected variance in the artifact due to sampling error (SE^2). This sampling error variance is then subtracted from the observed artifact variance to estimate the true (nonsampling) random-effects artifact variance (sampling error variance estimators for each artifact are given in Online Appendix C). For example, for u_X :

$$var'_{u_X} \approx var_{u_X} - SE_{u_X}^2, \quad (19)$$

where var_{u_X} is the observed variance of u_X , $SE_{u_X}^2$ is the predicted sampling error variance of u_X (computed via a bare-bones meta-analysis of u_X), and var'_{u_X} is the estimated true (i.e., residualized) random-effects variance of u_X . One can then substitute these residualized artifact variance estimates

into Equation 18 to yield:

$$\widehat{var}_{\rho_{TP_a}} \approx \left[var_{r_{XY_i}} - var_e - (b_1^2 var'_{q_{X_a}} + b_2^2 var'_{q_{Y_a}} + b_3^2 var'_{u_X} + b_4^2 var'_{u_Y}) \right] / b_5^2. \quad (20)$$

By estimating and removing artifacts' sampling variances from the observed artifact variances, this residualized TSA estimator gives a better estimate of the true amount of artifactual variance than is possible using either the interactive method (Le et al., 2016) or the unadjusted TSA method (Equation 18). We refer to the practice of computing TSA artifact-distribution meta-analyses using residualized artifact variances as the AD_{TSA_res} method.

An Interactive Artifact-Distribution Estimator for SD_ρ Using Shrunk Artifact Distributions

Just as Equation 20 uses residualized artifact variances to estimate SD_ρ in a Taylor series approximation artifact-distribution meta-analysis, it is possible to use true-score estimation techniques from classical test theory to remove the influence of error variance from distributions of artifact values. The classical test theory formula to estimate true scores is:

$$X_{true'} = \bar{X}_{obs} + (X_{obs} - \bar{X}_{obs}) \sqrt{\frac{var_{X_{obs}} - var_{X_{error}}}{var_{X_{obs}}}}, \quad (21)$$

where X_{obs} is a vector of observed scores, X_{error} represents estimated errors of measurement, $X_{true'}$ is a vector of estimated true scores, and the ratio in which $var_{X_{obs}} - var_{X_{error}}$ is divided by $var_{X_{obs}}$ is the definition of a reliability coefficient. Regressing (i.e., "shrinking") the $X_{obs} - \bar{X}_{obs}$ deviation scores toward zero using the measure quality index (i.e., the square root of the reliability coefficient) of the distribution as a regression weight reduces the variance of the distribution to reflect the amount of systematic (i.e., "true") variance in the distribution. Adding back the observed mean to the shrunken deviation scores yields a usable vector of estimated true scores. This linear transformation rescales the variance to account for the effects of error on observed scores, but it does not change the rank order of scores, nor does it change the overall shape of the distribution.

One can use the logic of the true-score estimation formula in Equation 21 to compute new artifact distributions consisting of artifact values that have been shrunken toward the mean. Once again using the u_X distribution as an example, a shrunken artifact distribution can be computed as:

$$u'_X = \bar{u}_X + (u_X - \bar{u}_X) \sqrt{\frac{var_{u_X} - SE_{u_X}^2}{var_{u_X}}}, \quad (22)$$

where u_X is a vector of observed u ratios, var_{u_X} is the variance of observed u ratios, SE_{u_X} is the standard error of the u -ratio distribution, and u'_X is a vector of shrunken values. After computing a shrunken distribution for each type of artifact, the new artifact distributions can be used in an interactive meta-analysis in the same way that one would use observed artifact distributions. We refer to Le et al.'s (2016) original practice of computing interactive artifact-distribution meta-analyses using vectors of observed artifacts values as the AD_{int} method. We refer to the practice of computing interactive artifact-distribution meta-analyses using shrunken (residualized) vectors of artifact values as the AD_{int_res} method.

Accuracy of Bivariate Indirect Range Restriction Correction Methods in Meta-Analysis

Having demonstrated the mathematical considerations involved in estimating sampling variances, artifact variances, and study weights when the BVIRR correction is used in meta-analysis, we now present a series of simulations to evaluate the relative accuracy of several BVIRR meta-analytic

Table 3. Parameter Values and Distributions Used in the Simulation

Parameter	Values/Distribution
k	10, 20, 50, 100
N	Gamma distribution with shape = .64 and scale = 165
$\bar{\rho}_{TP_a}$.1, .3, .5
$SD_{\rho_{TP_a}}$.00, .05, .10, .15, .20
$\bar{\rho}_{SP_a}$ and $\bar{\rho}_{ST_a}$.4
$SD_{\rho_{SP_a}}$ and $SD_{\rho_{ST_a}}$.2
$\bar{\rho}_{XX_a}$ and $\bar{\rho}_{YY_a}$.8
$SD_{\rho_{XX_a}}$ and $SD_{\rho_{YY_a}}$.05
SR	.1 (p = .028), .2 (p = .066), .3 (p = .124), .4 (p = .180), .5 (p = .204), .6 (p = .180), .7 (p = .124), .8 (p = .066), .9 (p = .028)

Note: N = sample size after selection; $\bar{\rho}_{TP_a}$ = mean unrestricted true-score correlation between X and Y; $\bar{\rho}_{SP_a}$ = mean unrestricted true-score correlation between selection variable and Y; $\bar{\rho}_{ST_a}$ = mean unrestricted true-score correlation between selection variable and X; $\bar{\rho}_{XX_a}$ = mean reliability of X; $\bar{\rho}_{YY_a}$ = mean reliability of Y; $SD_{\rho_{TP_a}}$ = standard deviation of unrestricted true-score correlations between selection variable and Y; $SD_{\rho_{SP_a}}$ = standard deviation of unrestricted true-score correlations between selection variable and X; $SD_{\rho_{ST_a}}$ = standard deviation of unrestricted true-score correlations between selection variable and X; $SD_{\rho_{XX_a}}$ = standard deviation of the reliability of X; $SD_{\rho_{YY_a}}$ = standard deviation of the reliability of Y; SR = selection ratio applied to the selection variable; p = weight determining the likelihood of a given SR being drawn from a random distribution. Sample sizes were sampled from the indicated gamma distribution with the constraint that sample sizes had to be at least 30 to be used so as to exclude studies smaller than those that typically appear in the published literature. The correlation parameters used to simulate a given sample were drawn from normal distributions with the constraint that they had to be valid correlations and form a positive definite correlation matrix. The reliability parameters used to simulate a given sample were drawn from beta distributions with the desired mean and standard deviation.

methods. We compare the accuracy of the three sets of study weights introduced earlier (see Equations 12, 13, and 14) in individual-correction meta-analyses. These simulations represent the first evaluation of the accuracy of BVIRR individual-correction methods (cf. Le et al., 2016, who only evaluated the accuracy of their interactive artifact-distribution procedure). We also compare the accuracy of the new TSA estimators against the interactive method estimators for SD_{ρ} in artifact-distribution meta-analysis. Based on the results of these simulations, we consider whether each estimator produces reasonable estimates and offer recommendations for future meta-analyses applying BVIRR corrections.

Method

We used simulation parameters that resemble the types of correlations, reliabilities, and sample sizes commonly observed in organizational and psychological literatures. We designed our simulations to introduce sampling error in correlations as well as artifacts to produce a high-fidelity representation of real primary studies. We used the `simulate_r_database` function in the *psychmeta R* package (Dahlke & Wiernik, 2017/2019) to generate all of our simulated samples with sampling error affecting all correlations and artifact estimates. Below, we give an overview of the simulation’s parameters and procedures.

Distributions of Correlation Parameters

We used a variety of mean and SD parameters to define the correlations among P , T , and S (see Table 3). The parameter distributions of correlations between T and P used in this simulation represent both fixed-effect and random-effects scenarios, with the mean correlations ranging from .10 to .50 and SD_{ρ} values ranging from .00 to .20. We chose to use a mean correlation of .40 for the

relationships between S and P and between S and T because this is a commonly observed magnitude of true-score correlation and creates a scenario in which selection on S could meaningfully impact the correlation between T and P . Preliminary simulations in which we varied both the mean and variance of the correlations involving S produced little variation in outcomes, so for the present study we only simulated conditions in which ρ_{ST_a} and ρ_{SP_a} had means of .40 and standard deviations of .20 because this represented a challenging set of circumstances for the meta-analytic methods to overcome. Greater average range restriction from selecting on S (a function of the mean correlations with S) and greater variability in the impact of selecting on S (a function of the SD of correlations with S) creates more variable artifact distributions and less consistent effects of artifacts across studies.

Distributions of Artifact Parameters

The constructs of interest, T and P , are measured with error, so reliability parameters were randomly drawn from the distribution described in Table 3 to attenuate the correlations involving T and P prior to generating data and inducing range restriction. We note that Le et al. defined separate parameter distributions for ρ_{XX_a} and ρ_{YY_a} , such that the ρ_{YY_a} distribution included the low levels of reliability typically observed for job performance criteria. However, we chose to use the same reliability distribution parameters for Y as for X so as to model potential applications of BVIRR in meta-analyses in areas of organizational research other than personnel selection. To induce indirect range restriction in the relationship between X and Y , we used selection ratios (proportion of the applicant sample selected) ranging from .10 to .90 to perform selection on a suitability construct S . Although selection is always performed on measured variables in practice, our choice to select on a construct does not harm the fidelity of our simulation, as selecting on a measured variable can only affect the variance of other variables through selection's effects on true scores (see Figure 1C) and the reliability of S therefore has little effect on the extent of range restriction in X and Y .

Sample Sizes of Primary Studies (N)

Sample sizes (N) need to be defined to model the statistical artifact of sampling error. We sampled N from a gamma distribution (shape = .64 and scale = 165), which produces a skewed distribution of sample sizes similar to the distributions commonly observed in published meta-analyses (the mean sample size using this distribution was 155). In generating random sample sizes, we only retained sample sizes of 30 or larger so that our simulated samples would be similar to those encountered in many domains of organizational and psychological research. Our distribution of sample sizes reflects the number of cases *after* selection, so we had to simulate $N_a = N / SR$ cases in each sample (where SR is a selection ratio indicating the proportion of applicants selected) to ensure that there would be N cases left over after performing selection.

Number of Studies in Meta-Analyses (k)

We computed meta-analyses with 10, 20, 50, and 100 studies. These second-order sample sizes span a range of ks commonly encountered in organizational research, although meta-analyses with more than 100 studies are not uncommon. These ks allowed us to examine the asymptotic accuracy of BVIRR as a function of the size of the meta-analysis in which it was used.

Proportion of Studies Providing Artifacts

For artifact-distribution meta-analyses, we manipulated whether 20% or 50% of studies provided artifact information by randomly deleting artifact information from our simulated databases. When

artifact deletion is used to create missingness in a random fashion, it makes no difference whether the artifacts are deleted at the level of the study or the individual artifact observation (Le et al., 2016). We chose to delete all artifacts associated with randomly selected studies.

Simulation Procedure

The combinations of parameters described above resulted in a $3 (\bar{\rho}_{TP_a}) \times 5 (SD_{\rho_{TP_a}}) \times 4 (k)$ design for our simulation (plus a 2-level missingness manipulation for artifact-distribution methods). For each of the 60 possible combinations of the $\bar{\rho}_{TP_a}$, $SD_{\rho_{TP_a}}$, and k parameters, we simulated meta-analytic databases from sets of parameter values that were randomly drawn from the parameter distributions associated the simulation condition; all other parameters were randomly selected from the same distributions across all conditions. In each simulated primary study, a 3×3 correlation matrix containing ρ_{TP_a} , ρ_{SP_a} , and ρ_{ST_a} was constructed, a random multivariate data set was generated from the population matrix with measurement error affecting the T and P constructs, and selection was performed on the S construct using a randomly chosen selection ratio. The observed r_{XY_i} correlation and the range-restricted reliabilities and standard deviations of X and Y were then computed. The range-restricted standard deviations served as u ratios, as the population standard deviation toward which all studies were corrected was unity. We used a consistent referent standard deviation to define u ratios because it is a common practice for meta-analysts to correct observed correlations toward a standard deviation reported in a test manual or in a large-scale, representative study (Approaches 2–5; cf. Ones & Viswesvaran, 2003; Sackett & Ostgaard, 1994).

For each of our conditions, we meta-analyzed 1,000 sets of simulated studies. For each random sample of studies, we computed meta-analyses using seven methods: IC_A , IC_V , IC_{TSA} , AD_{TSA} , AD_{TSA_res} , AD_{Int} , and AD_{Int_res} . All comparative results were computed by analyzing exactly the same data with competing meta-analytic methods. We computed the mean and standard deviation (i.e., standard error) of all $\bar{\rho}_{TP_a}$ and $SD_{\rho_{TP_a}}$ estimates across conditions and plotted these means and standard errors for intuitive presentation of results. For the sake of parsimony, we present results for meta-analyses with k of 10 and 50 and with $SD_{\rho_{TP_a}}$ parameters of 0, .1, and .2. Tables of numeric results from all simulation conditions are available in Online Appendix F and a full set of figures showing results from all simulation conditions is available in Online Appendix G.

Results

Mean ρ Estimates

Simulation results for mean ρ estimates are shown in Figure 3. The first notable trend in the estimates of mean correlations is that the conventional method for computing individual-correction meta-analyses (IC_A) using the $N \cdot A^2$ weights recommended by Le et al. (2016) does a very poor job of identifying the correct mean parameter value with the BVIRR correction. The IC_A method gave the most dramatic underestimates of the mean correlation when the true mean correlation was small (e.g., .10) and, although it performed better with larger correlations, the IC_A method's mean correlation estimates were also less accurate when the true variance of correlation parameters was large (e.g., $SD_{\rho} = .20$). Interestingly, the IC_A method performed *worse* when more studies were included in a meta-analysis, which indicates that it is an inconsistent estimator.

All the other meta-analytic methods performed quite well in estimating the true mean correlation. The TSA method of determining corrected sampling errors and study weights in individual-correction meta-analyses (IC_{TSA}) yielded slightly more variable (i.e., less consistent) estimates of the mean correlation than the equation-implied method based on the V coefficient (IC_V). The difference in accuracy and consistency between the IC_{TSA} and IC_V methods was small enough that neither is clearly superior to the other; we will therefore base our recommendations about which method to use on the

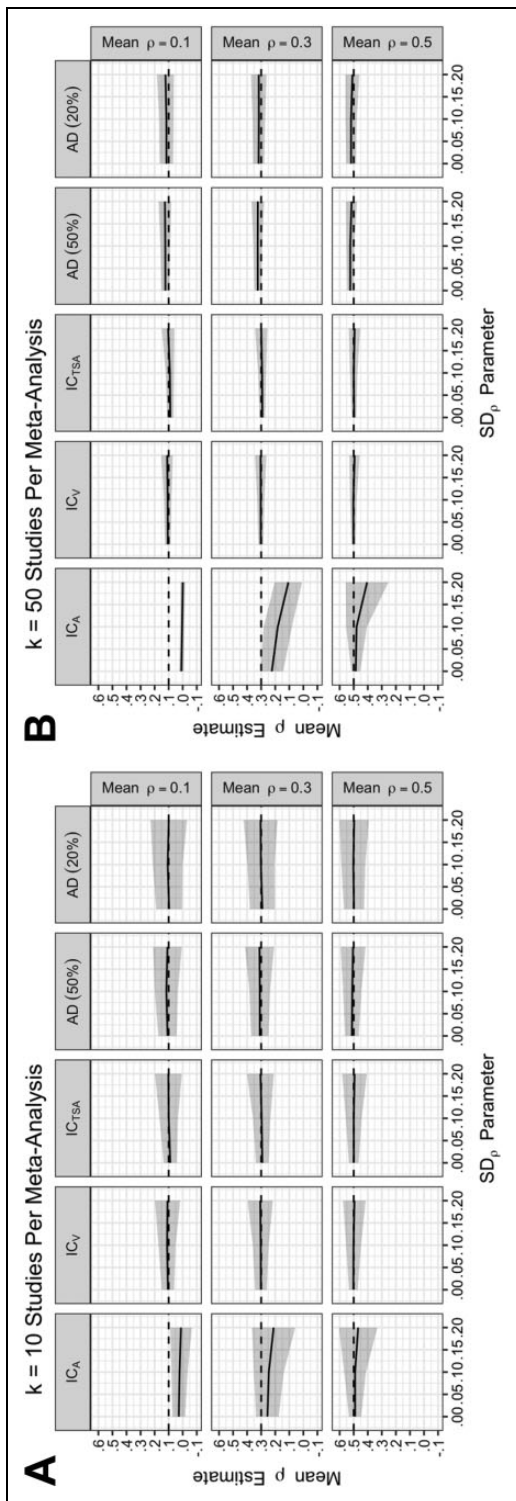


Figure 3. Mean ρ estimates by meta-analytic method and number of studies meta-analyzed. IC_A = traditional compound attenuation factor-based individual-correction method; IC_V = individual-correction method using equation-implied error variances and weights; IC_{TSA} = individual-correction method using Taylor series approximation to estimate corrected error variances and weights; $AD(50\%)$ = artifact distribution method with 50% of artifacts available; $AD(20\%)$ = artifact distribution method with 20% of artifacts available. AD results were identical across different AD computation methods, as interactive and Taylor series approximation methods both use mean artifacts to estimate mean ρ values. Dashed lines indicate true parameter values, solid lines indicate mean parameter estimates, and ribbons around solid lines indicate the parameter estimates values within 1 standard error of the mean parameter estimate (standard error was computed as the standard deviation of statistics from 1,000 replications).

accuracy of their SD_ρ estimates. The artifact-distribution methods did quite well at recovering the mean correlation, even when only 20% of artifacts were available. The interactive and TSA artifact-distribution methods both used mean artifact values to estimate mean corrected correlations, so only the TSA-based artifact-distribution mean ρ estimates are displayed in Figure 3.

SD_ρ Estimates

Figure 4 and Figure 5 depict the accuracy of SD_ρ estimates from competing meta-analytic methods for second-order sample sizes of 10 and 50 studies, respectively. The first clear trend from these results is that the IC_A method was quite inaccurate at estimating SD_ρ . As the true SD_ρ increased, a linear association between the true mean ρ and the estimated SD_ρ emerged, indicating that the IC_A method has a strong and systematic bias. This bias, combined with the IC_A method's bias for estimating mean ρ and its weak mathematical foundation, is such that we cannot under any circumstance advocate using this method. The other individual-correction methods were more accurate at estimating SD_ρ than the IC_A method, but not without their own weaker biases. The IC_V method showed a stronger positive bias than the IC_{TSA} method across all second-order sample sizes. Considering the superior accuracy of the IC_{TSA} method for estimating SD_ρ , we recommend this method over the other individual-correction methods.

In terms of artifact-distribution methods, we found that the interactive method using observed artifact distributions (AD_{Int}) and the TSA method using observed artifact variances (AD_{TSA}) produced SD_ρ estimates with distinct negative biases, while the TSA method using residualized artifact variances (AD_{TSA_res}) and the interactive method using shrunken artifact distributions (AD_{Int_res}) produced SD_ρ estimates that were generally quite accurate. The biases of the AD_{Int} and AD_{TSA} methods were larger when the SD_ρ parameter was smaller because artifactual variance represented a larger proportion of the total variance in these conditions; these methods do not account for the sampling error of artifacts and therefore misidentify that error variance as contributing to genuine artifactual variance in the distribution of correlations. By accounting for the predicted sampling error in distributions of u ratios and measure quality indices, the AD_{TSA_res} and AD_{Int_res} methods more accurately estimated the SD_ρ parameter and reduced the disproportionate influence of artifact distributions on SD_ρ estimates when true parameter variance was small.

Considering the excellent performance of the AD_{TSA_res} and AD_{Int_res} methods with incomplete artifact information, we applied the AD_{TSA_res} method using *complete* artifact information and compared the results of those analyses against the results obtained with the IC_{TSA} method. Our summaries of the mean ρ and SD_ρ results plotted in Figure 6 reveal that the AD_{TSA_res} method was, in fact, more accurate than the IC_{TSA} method for estimating both parameters.

Discussion

We designed a simulation to evaluate the accuracy with which several methods for applying the BVIRR correction in meta-analyses recovered correlations' mean and standard deviation parameters. We found that the conventional IC_A method was ill-equipped to accurately estimate the means or standard deviations of the true parameter distributions. Our simulated meta-analyses of varying second-order sample sizes showed that the IC_A method performed worse at estimating mean ρ and SD_ρ in larger- k meta-analyses than in smaller- k meta-analyses. This asymptotic *inaccuracy* indicates that the IC_A method is not a consistent statistical estimator. Both of our newly proposed individual-correction methods proved superior to the IC_A method when applied with the BVIRR correction, but of these two we found that the IC_{TSA} method was more balanced at accurately estimating both the mean and the SD of the parameter distribution. The IC_{TSA} method accounts for the sampling error of artifact values, thereby providing a more complete estimate of the uncertainty

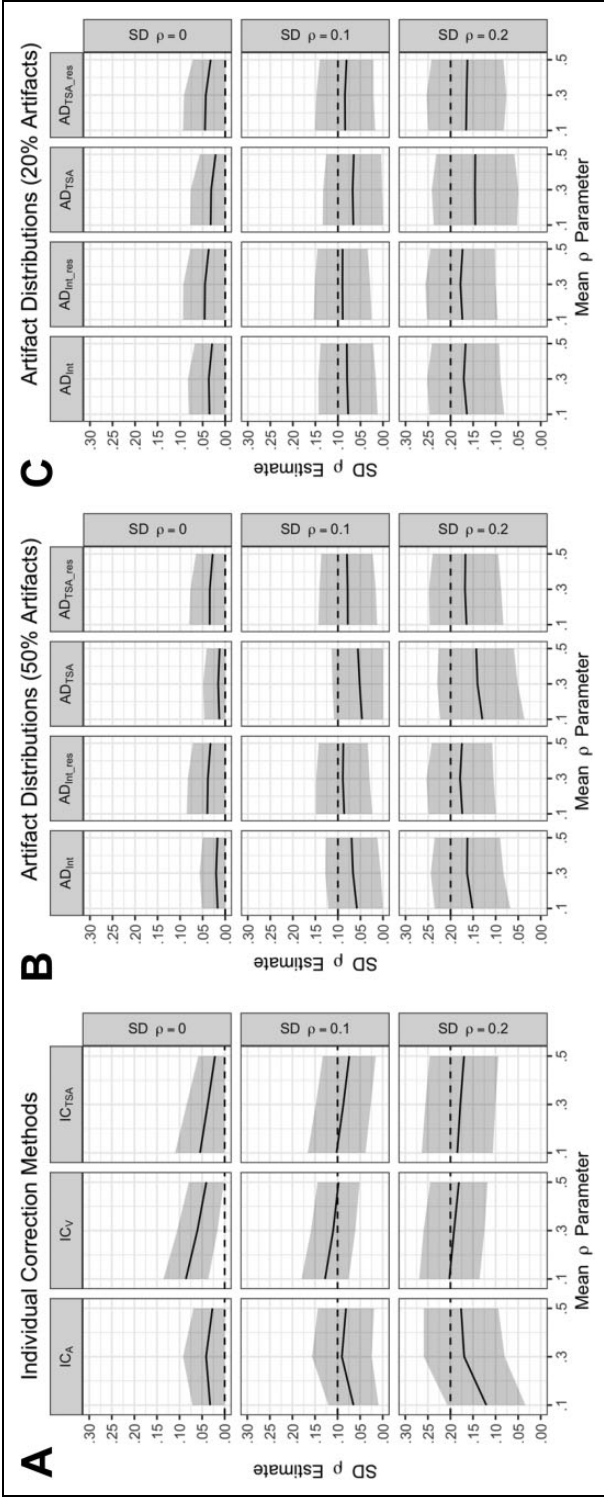


Figure 4. SD_p estimates by meta-analytic method for meta-analysis of 10 studies ($k = 10$). IC_A = traditional compound attenuation factor-based individual-correction method; IC_V = individual-correction method using equation-implied error variances and weights; IC_{TSA} = individual-correction method using Taylor series approximation to estimate corrected error variances and weights; AD_{Int} = interactive artifact distribution method; AD_{Int_res} = interactive artifact distribution method with shrunken artifact distributions; AD_{TSA} = Taylor series approximation artifact distribution method using observed artifact variances; AD_{TSA_res} = Taylor series approximation artifact distribution method using artifact variances residualized to remove the influence of predicted sampling error. Dashed lines indicate true parameter values, solid lines indicate mean parameter estimates, and ribbons around solid lines indicate the parameter estimates values within 1 standard error of the mean parameter estimate (standard error was computed as the standard deviation of statistics from 1,000 replications).

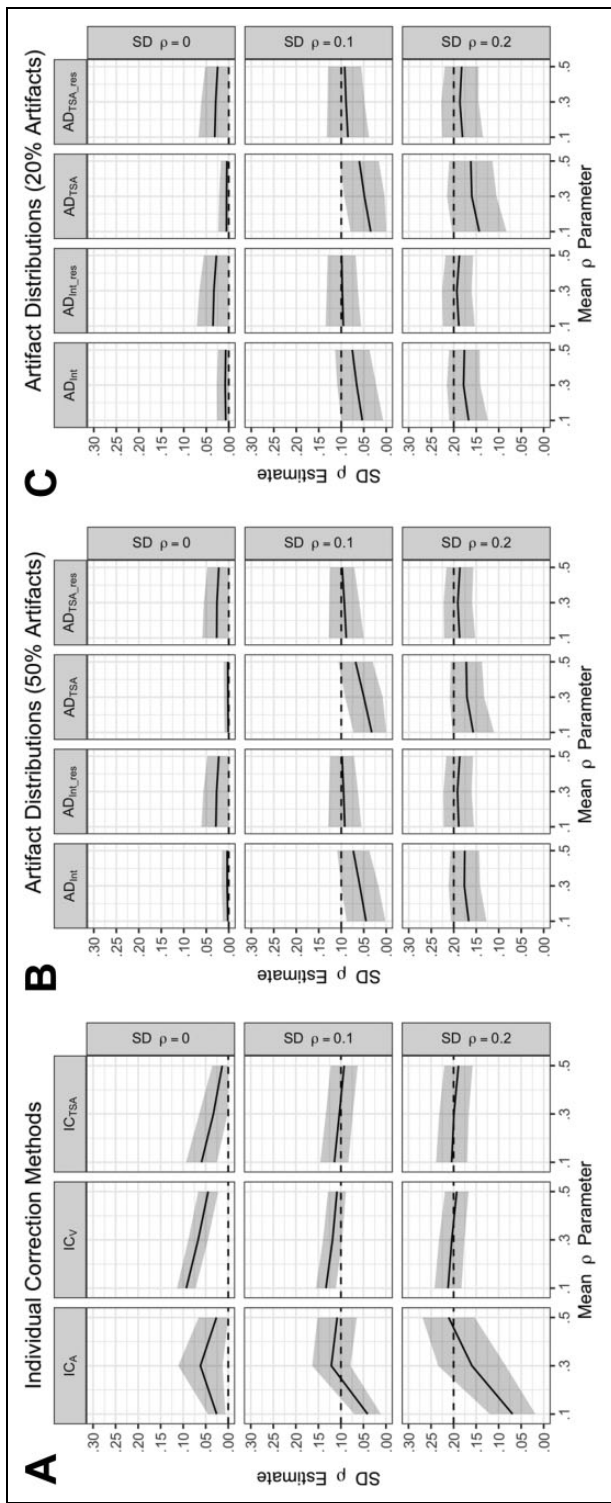


Figure 5. SD_p estimates by meta-analytic method for meta-analysis of 50 studies ($k = 50$). IC_A = traditional compound attenuation factor-based individual-correction method; IC_V = individual-correction method using equation-implied error variances and weights; IC_{TSA} = individual-correction method using Taylor series approximation to estimate corrected error variances and weights; AD_{Int} = interactive artifact distribution method; AD_{Int_res} = interactive artifact distribution method with shrunken artifact distributions; AD_{TSA} = Taylor series approximation artifact distribution method using observed artifact variances; AD_{TSA_res} = Taylor series approximation artifact distribution method using artifact variances residualized to remove the influence of predicted sampling error. Dashed lines indicate true parameter values, solid lines indicate mean parameter estimates, and ribbons around solid lines indicate the parameter estimates values within 1 standard error of the mean parameter estimate (standard error was computed as the standard deviation of statistics from 1,000 replications).

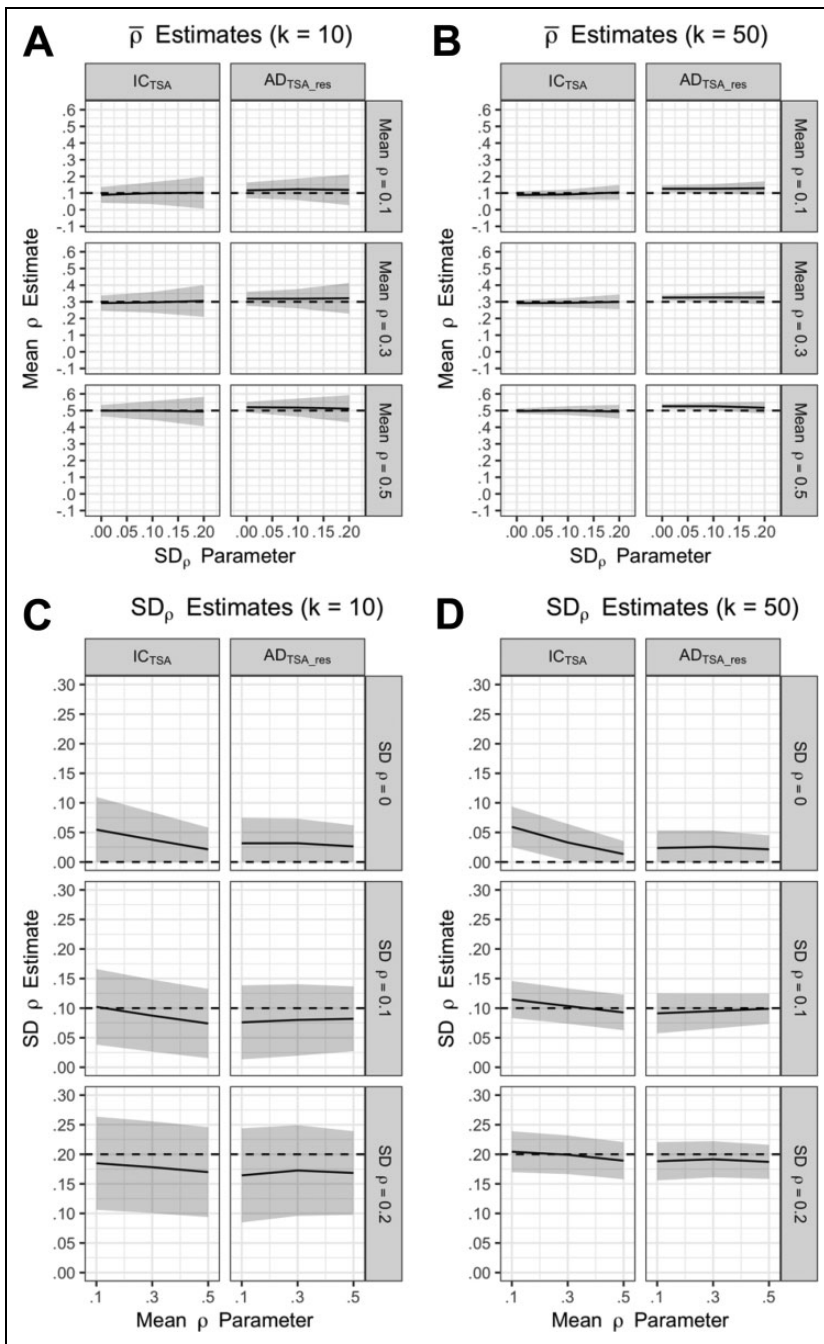


Figure 6. Mean ρ and SD ρ estimates for IC_{TSA} (individual-correction method using Taylor series approximation to estimate corrected error variances and weights) and AD_{TSA_res} (Taylor series approximation artifact distribution method using artifact variances residualized to remove the influence of predicted sampling error; using 100% of artifact information) methods by number of studies meta-analyzed.

in the corrected correlations and enabling more appropriate weights to be assigned to corrected correlations.

In analyses using the IC_A method, we found patterns of bias that were wildly inconsistent across mean ρ , SD_ρ , and k parameters, as well as across meta-analyses with different mean sample sizes. The other individual-correction methods were reliable estimators of the mean correlation, and the artifact distribution methods performed the most consistently of all of the methods when estimating mean ρ . The IC_A method demonstrated patterns of results that are consistent with the issues raised earlier regarding the trends depicted in Figure 2 and support our assertion that the method's poor statistical foundations make it inadequate for operational use.

With regard to artifact-distribution methods, we found that sampling error in statistical artifacts caused negatively biased estimates of SD_ρ with the AD_{Int} and AD_{TSA} methods, which both rely solely on observed artifact distributions. Removing sampling variance from artifact distributions before using them to estimate SD_ρ in the AD_{TSA_res} and AD_{Int_res} methods provided the most accurate artifact-distribution estimates of SD_ρ . In fact, the SD_ρ estimates from these methods were even more accurate than the estimates produced by any of the individual-correction meta-analysis estimators. We compared the AD_{TSA_res} method (using 100% of artifact information) to the IC_{TSA} method and found that applying the AD_{TSA_res} method was more accurate than making individual corrections (see Figure 6). Given the accuracy of the AD_{TSA_res} artifact-distribution methods at estimating both mean ρ and SD_ρ and the fact that even our best individual-correction method for estimating SD_ρ exhibited varying degrees of bias across different mean ρ parameters, we recommend using the AD_{TSA_res} method even when artifact information is available from all studies. Even though the AD_{TSA_res} and AD_{Int_res} methods performed similarly in our simulations, we prefer to recommend the AD_{TSA_res} method because it is more computationally efficient. In addition, the AD_{TSA_res} method requires only artifact means and variances (rather than vectors of individual artifact estimates), allowing greater flexibility when applying the BVIRR correction because researchers can combine observed artifact distributions with the descriptive statistics of artifact distributions from relevant populations reported in previous meta-analyses.

Application of the BVIRR Correction to Expatriate Adjustment Meta-Analyses

The above simulation illustrated that our new BVIRR methods are more accurate than competing methods. To demonstrate how the BVIRR correction can be useful in analyses of real effect sizes, we reanalyzed meta-analyses of relationships between Big Five personality traits and expatriate adjustment reported by Harari, Reaves, Beane, Laginess, and Viswesvaran (2018). We chose to reanalyze these data for three key reasons. First, the expatriate literature does not have a history of applying corrections for range restriction, and this example demonstrates how range-restriction corrections can be applied in settings other than personnel selection research. Second, the studies included in these meta-analyses used a variety of adjustment and personality measures, often with varying numbers of items, and frequently did not report standard deviations. This meta-analysis is thus typical of many literatures in organizational research and allows us to demonstrate the reliability-based u ratio methods described in Table 1. Third, the expatriate literature offers an opportunity to showcase how range-restriction corrections can be applied in a way that accounts for the effects of range variation on SD_ρ without changing mean ρ , much like how researchers who use validity generalization in personnel selection are interested in accounting for artifactual variation due to predictor measurement error but are not interested in correcting the mean effect size for predictor measurement error. The expatriate literature lacks a clear or distinct target population, so range restriction's impact on mean effect sizes is not of substantive interest, but correcting

for the impact of range variation on SD_p could add value by accounting for more artifactual variance in meta-analyses.

Methods

Meta-Analytic Sample

Data for these meta-analyses were drawn from Harari et al. (2018). These authors reported a meta-analysis of relations between the Big Five personality traits and multiple dimensions of expatriate adjustment—the degree to which expatriate employees feel adapted to their new location in terms of everyday errands and tasks (“general adjustment”), interacting with host-country nationals (“interactional adjustment”), their new work roles (“work adjustment”), and overall adjustment (a composite of the other three dimensions; Black & Gregersen, 1991). Harari et al. found small to moderate relationships between each of the Big Five traits and each dimension of expatriate adjustment ($\bar{\rho}$ ranged .13–.30 across traits and adjustment dimensions). However, most correlations were associated with substantial residual variability in correlations across samples (mean $SD_p = .11$, range .00–.17), suggesting the presence of moderators. Harari et al. corrected for measurement error in both personality and adjustment, but did not correct for differential variability (range restriction/enhancement) across samples. It is plausible that some of the included samples are more homogenous in terms of personality. Similarly, it is plausible that some contexts may be easier to adjust to, leading to less variability on adjustment than in other contexts. Accordingly, it is possible that some (or much) of the estimated residual variability is due to differential variability on personality and/or adjustment. We tested this hypothesis by reanalyzing their meta-analyses of adjustment with the three traits most commonly discussed in relation to adjustment—emotional stability, openness, and extraversion (Albrecht, Ones, & Sinangil, 2018; Ones, Sinangil, & Wiernik, 2018).⁹

Analyses

In our reanalyses, we corrected for both measurement error and range variation in both variables. Reliability coefficients were reported in each study. The appropriate unrestricted reference population for expatriate adjustment–personality relations is unclear (e.g., the general population, domestic managers, expatriates across all contexts; Kostal, Wiernik, Albrecht, & Ones, 2018), so we chose to correct only for artifactual variability, to not adjust $\bar{\rho}$, and to compute u ratios using Approach 4 in Table 1. We compared each sample to either norm samples for its personality and adjustment measures or the samples in which the scales were originally developed. If a study reported an SD and used the full set of items for the scale with the original number of response scale points, we computed a u ratio using SD s and formula 2a in Table 1. If a study did not report an SD , used an abbreviated scale, or used a different number of response scale points as were used in the norm/development sample, we computed a u ratio using reliability coefficients and formula 2b in Table 1.¹⁰ We centered the u ratio distribution for each construct around 1.0 using formula 3 in Table 1. Based on the simulation results, we conducted meta-analyses correcting for BVIRR using the residualized artifact distribution method, and we compared these results to those obtained correcting only for measurement error. By design, the mean u ratio was 1.0 for each variable, with the random-effects SD s of the u ratio distributions equal to .243 for emotional stability, .200 for openness, .275 for extraversion, .210 for overall adjustment, .179 for general adjustment, .219 for interaction adjustment, and .174 for work adjustment.

Results

Meta-analytic results are shown in Table 4. When correcting for measurement error alone, all personality–adjustment relations showed substantial residual variability (SD_p ranged .11–.18, mean

Table 4. Artifact-Distribution Results for Personality-Expatriate Adjustment Meta-Analyses.

Personality Trait	Adjustment Dimension	k	N	\bar{r}	SD_r	SD_{res}	$\bar{\rho}$	SD_{r_c}	SD_{ρ}	% Decr.	95% CI	80% CV
Results from correcting for measurement error alone												
Openness	Overall	27	5288	.19	.15	.14	.24	.19	.17		(.16, .31)	(.02, .46)
	General	15	3758	.15	.11	.09	.19	.14	.12		(.11, .27)	(.04, .35)
	Interaction	14	3743	.18	.11	.09	.22	.14	.12		(.14, .30)	(.06, .38)
	Work	14	3982	.18	.15	.14	.23	.19	.17		(.12, .33)	(-.01, .46)
Extraversion	Overall	31	4513	.25	.14	.11	.30	.17	.14		(.24, .37)	(.12, .49)
	General	17	2793	.23	.11	.09	.29	.15	.11		(.21, .36)	(.14, .44)
	Interaction	18	2977	.20	.12	.10	.24	.15	.11		(.17, .32)	(.09, .40)
	Work	14	2378	.22	.16	.14	.27	.20	.18		(.15, .38)	(.03, .50)
Emotional stability	Overall	26	4115	.24	.14	.12	.29	.17	.14		(.23, .36)	(.11, .48)
	General	14	2584	.19	.12	.09	.24	.15	.12		(.15, .32)	(.08, .40)
	Interaction	14	2634	.16	.12	.10	.20	.15	.12		(.11, .28)	(.04, .36)
	Work	14	2881	.25	.14	.13	.30	.18	.16		(.20, .40)	(.09, .51)
Results from correcting for measurement error and bivariate indirect range restriction (BVIRR)												
Openness	Overall	27	5288	.19	.15	.13	.24	.19	.15	12%	(.16, .31)	(.04, .44)
	General	15	3758	.15	.11	.08	.19	.14	.10	17%	(.11, .27)	(.05, .33)
	Interaction	14	3743	.18	.11	.08	.22	.14	.10	17%	(.14, .30)	(.09, .35)
	Work	14	3982	.18	.15	.13	.23	.19	.16	6%	(.12, .33)	(.01, .45)
Extraversion	Overall	31	4513	.25	.14	.08	.30	.17	.09	36%	(.24, .37)	(.18, .43)
	General	17	2793	.23	.11	.05	.29	.15	.06	45%	(.21, .36)	(.21, .37)
	Interaction	18	2977	.20	.12	.07	.24	.15	.08	27%	(.17, .32)	(.14, .35)
	Work	14	2378	.22	.16	.12	.27	.20	.15	17%	(.15, .38)	(.06, .47)
Emotional stability	Overall	26	4115	.24	.14	.09	.29	.17	.11	21%	(.23, .36)	(.15, .44)
	General	14	2584	.19	.12	.07	.24	.15	.09	25%	(.15, .32)	(.11, .37)
	Interaction	14	2634	.16	.12	.08	.20	.15	.10	17%	(.11, .28)	(.06, .33)
	Work	14	2881	.25	.14	.11	.30	.18	.13	19%	(.20, .40)	(.12, .48)

Note: k = number of studies contributing to meta-analysis; N = total sample size; \bar{r} = mean observed correlation; SD_r = observed standard deviation of r ; SD_{res} = residual standard deviation of r ; $\bar{\rho}$ = mean true-score correlation; SD_{r_c} = observed standard deviation of corrected correlations (r_c); SD_{ρ} = residual standard deviation of ρ ; % decr. = percentage decrease in SD_{ρ} after accounting for differential variability; CI = confidence interval around $\bar{\rho}$; CV = credibility interval around $\bar{\rho}$; Correlations corrected using artifact distributions. BVIRR corrections adjusted only residual variability, not mean correlations.

= .14). When differential variability across samples was also accounted for using the BVIRR correction, SD_p estimates decreased by an average of 21% (range 6%–45%; mean variance decrease = 37%, range 11%–70%), with SD_p now ranging .06–.15 (mean = .11). Correcting for BVIRR had the largest impact for extraversion. For example, when correcting for measurement error only, the credibility interval for extraversion–general adjustment correlations ranged from small to quite large values (.14–.44). In contrast, when indirect range variation was accounted for, the credibility interval spanned consistently moderate values (.21–.37), indicating less potential influence for moderators (cf. Wiernik, Kostal, Wilmot, Dilchert, & Ones, 2017). Impacts for the other two traits were smaller.

Discussion

Our reanalysis of Harari et al.'s (2018) data shows how range-restriction corrections can help to account for additional sources of artifactual variance even when the influence of range restriction on the mean effect size is not of substantive interest. By accounting for the variation in effect sizes attributable to range restriction, we arrived at SD_p estimates that were, on average, 21% smaller than when range variation was ignored. The effect of range-restriction artifacts on SD_p also impacts credibility intervals, allowing for a clearer understanding of the true heterogeneity of effect sizes and better estimates of the generalizability of relationships.

General Discussion

We presented several new approaches for applying the bivariate indirect range-restriction (BVIRR) correction in primary research and meta-analyses. We presented a generalized BVIRR formula that can be used to correct for either range restriction or range enhancement (and allows an approximate correction for mixtures thereof), a new Taylor series approximation procedure for estimating the sampling error of BVIRR-corrected correlations, new methods for assigning study weights in BVIRR individual-correction meta-analyses, and a new Taylor series approximation method for estimating SD_p in BVIRR artifact-distribution meta-analyses. We demonstrated the accuracy of our BVIRR methods via simulation and used meta-analyses of expatriate adjustment to illustrate how our methods can be applied to real data, even when correcting for range restriction is not an obvious course of action.

The issues addressed in this article converge on the general conclusion that the BVIRR correction will function very differently than univariate range-restriction corrections (i.e., either UVDRR/Case II or UVIRR/Case IV) or corrections for measurement-error alone when applied in psychometric meta-analysis. The BVIRR correction includes an additive term, which not only permits BVIRR-corrected correlations to change signs, but also contributes large negative bias to SD_p estimates in artifact-distribution meta-analysis unless steps are taken to account for sampling error in artifact distributions. The BVIRR correction's additive term also means that traditional methods of assigning study weights in individual-correction meta-analysis produce inaccurate parameter estimates because traditional weights only work with multiplicative corrections for which meaningful compound attenuation factors exist. We presented solutions to these issues and illustrated that our refinements to BVIRR meta-analytic methods dramatically improve the recovery of parameter values. The new methods in this article are broadly relevant to researchers and practitioners who wish to correct for indirect range restriction, whether in primary analyses or meta-analyses.

We have shown that accounting for sampling error in artifact distributions dramatically increases the accuracy of SD_p estimates in artifact-distribution meta-analyses. As SD_p is the most important indicator of effect-size heterogeneity in the Hunter-Schmidt method of meta-analysis, this finding has important implications for the accuracy of inferences regarding the generalizability of effects.

Without accounting for artifact sampling error, researchers using the BVIRR correction have a much greater risk of errantly concluding that an effect generalizes (i.e., has a homogenous distribution) and prematurely foreclosing on searches for moderators (cf. job satisfaction's correlations with neuroticism, extraversion, and openness reported by Le et al., 2016). Artifact-distribution approaches are the most commonly used psychometric meta-analytic methods because artifact information is inconsistently reported in primary studies; thus, our improvements to both interactive and TSA-based artifact-distribution meta-analytic methods have important implications for how BVIRR will be used in meta-analyses. The greater accuracy of artifact-distribution versus individual-correction methods for BVIRR meta-analyses also counters common beliefs among researchers that individual-correction methods are inherently superior. We recommend using our residualized TSA artifact-distribution method in BVIRR meta-analyses, even in the rare case that complete artifact information is available for all studies, and particularly in lieu of ad hoc procedures for using individual-correction methods with missing data (e.g., mean substitution of missing reliabilities and/or u ratios or setting missing values to unity). It is also worth noting that TSA-based artifact-distribution methods are more flexible than interactive methods because they only require one to know the means and variances of artifacts. This means that researchers can easily combine their observed artifact distributions with the means and variances of artifact distributions reported in prior meta-analyses or in technical manuals (assuming, of course, that these artifact distributions represent relevant populations of participants and measures) and can compute artifact-distribution meta-analyses using the means and variances of mixture distributions that result from combining all relevant artifact distributions.

Although the BVIRR correction is particularly vulnerable to the biasing effects of sampling error in artifact distributions, we recommend that future research evaluate the effects of sampling error in artifact-distribution meta-analyses using other corrections. Similar gains in precision may be realized by applying the logic of our procedures to other artifact corrections. Our recommendations and refinements are supported by simulations and substantive examples that clearly illustrated the substantial bias that can result from treating BVIRR like other artifact corrections by applying conventional meta-analytic procedures. To aid in using these methods, functions for applying the BVIRR correction procedures described in this article to primary analyses and meta-analyses are implemented in the *psychmeta* R package (Dahlke & Wiernik, 2018, 2017/2019). Simple R functions for computing BVIRR-corrected correlations and sampling error variances are also available in Online Appendix H. We strongly encourage researchers to use the BVIRR correction when possible and we emphasize the importance of correctly modeling BVIRR's effects on correlations in both primary studies and meta-analyses.

Authors' Note

Jeffrey A. Dahlke is currently at HumRRO but completed this research while pursuing his PhD in industrial-organizational psychology at the University of Minnesota.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Jeffrey A. Dahlke  <https://orcid.org/0000-0003-1024-4562>

Supplemental Material

Supplemental material for this article is available online at <http://journals.sagepub.com/doi/suppl/10.1177/1094428119859398> and <https://osf.io/sh9dn/>

Notes

1. If some samples do not report unrestricted group information, u values computed using Approach 2 are often substituted, with relatively minor impacts on results (Ones & Viswesvaran, 2003; Sackett & Ostgaard, 1994).
2. Or if studies frequently change the number of items or response scale points used on a measure.
3. As the number of range-variation corrections in use increases, the traditional Roman-numeral-based nomenclature (Cases I, II, III, IV, V) becomes unwieldy. In this article, we introduce a more descriptive system of abbreviations to succinctly identify the nature of range variation and the information used to make corrections (UVDRR [Case II], UVIRR [Case IV], BVIRR [Case V], BVDRR). In this naming scheme, UV means univariate, BV means bivariate, DRR means direct range restriction, and IRR means indirect range restriction.
4. Following tradition in the range-restriction literature, we use the subscript i to refer to values for the selected (“incumbent”) group and the subscript a to refer to values for the unselected (target population, “applicant”) group.
5. Le, Oh, Schmidt, and Wooldridge (2016) presented mathematically equivalent equations using estimated true score u ratios. However, true-score u ratios are unnecessary in the BVIRR correction (their requirement in the UVIRR correction stems from UVIRR’s mediation assumption; cf. Alexander, 1990). We present several mathematically equivalent forms of Equation 1 in Online Appendix A. Eliminating the use of true-score u ratios simplifies the correction and avoids the common problem that true-score u ratios are undefined for particular combinations of reliability estimates and observed-score u ratios.
6. Although one certainly hopes that observed u ratios reflect the actual effects of selection on the variables of interest, u ratios are affected by substantial sampling error. This sampling error means that one can obtain u ratios larger than 1 when range restriction has actually occurred, or one can obtain u ratios smaller than 1 when range enhancement has actually occurred. This uncertainty can be accommodated by meta-analytic methods described later. When using the BVIRR correction in primary studies, there is no way to correct individual u ratios for sampling error, so one must simply assume that observed u ratios are meaningful and use them for corrections; that said, suspicion of one’s u ratios due to small sample size should beget suspicion of the corrected correlation.
7. A derivation of Equation 5 and a description of the logic of the equation are given in Online Appendix B.
8. Corrections for univariate range *enhancement*, however, imply that var_{e_c} can be *smaller than* var_e . This is because range enhancement represents a specific form of oversampling (at the tails of a distribution) and applying a correction reduces the variance for the sake of estimating a more representative effect size. Since range-restriction corrections involve *extrapolating* variance information whereas range-enhancement corrections involve *sacrificing* variance information, these corrections have opposite effects on the sampling variances of the corrected statistics.
9. Thank you to Michael Harari for kindly sharing the data to conduct these reanalyses. Additional results are available from Michael Harari at mharari@fau.edu.
10. If an abbreviated scale was used, we adjusted r_{XX_i} upward using the Spearman-Brown prophecy formula to reflect the reliability of the full-length scale. For studies that developed their own scales, u was set equal to 1.0.

References

- Aitken, A. C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 4(2), 106-110. <https://doi.org/10/bmqrbz>

- Albrecht, A.-G., Ones, D. S., & Sinangil, H. K. (2018). Expatriate management. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Eds.), *The SAGE handbook of industrial, work and organizational psychology* (2nd ed., Vol. 3, pp. 429-465). London, UK: Sage. <https://doi.org/10/c4xf>
- Alexander, R. A. (1990). Correction formulas for correlations restricted by selection on an unmeasured variable. *Journal of Educational Measurement*, 27(2), 187-189. <https://doi.org/10/cjrr9c>
- Alexander, R. A., Carson, K. P., Alliger, G. M., & Carr, L. (1987). Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Educational and Psychological Measurement*, 47(2), 309-315. <https://doi.org/10/bc6dzj>
- Beatty, A. S., Barratt, C. L., Berry, C. M., & Sackett, P. R. (2014). Testing the generalizability of indirect range restriction corrections. *Journal of Applied Psychology*, 99(4), 587-598. <https://doi.org/10.1037/a0036361>
- Black, J. S., & Gregersen, H. B. (1991). The other half of the picture: Antecedents of spouse cross-cultural adjustment. *Journal of International Business Studies*, 22(3), 461-477. <https://doi.org/10/b2nbxx>
- Bryant, N. D., & Gokhale, S. (1972). Correcting correlations for restrictions in range due to selection on an unmeasured variable. *Educational and Psychological Measurement*, 32(2), 305-310. <https://doi.org/10/fc4533>
- Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Conlon, D. E., & Wesson, M. J. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*, 98(2), 199-236. <https://doi.org/10/q6c>
- Dahlke, J. A., & Wiernik, B. M. (2018). *psychmeta*: An R package for psychometric meta-analysis. *Applied Psychological Measurement*, 43(5), 415-416. <https://doi.org/10/gfgt9t>
- Dahlke, J. A., & Wiernik, B. M. (2019). *psychmeta*: Psychometric meta-analysis toolkit [R Package]. Retrieved from <https://cran.r-project.org/package=psychmeta> (Original work published 2017)
- DeRue, D. S., Nahrgang, J. D., Wellman, N. E. D., & Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology*, 64(1), 7-52. <https://doi.org/10/fwzt2t>
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40(1), 31-53. <https://doi.org/10/gdmb59>
- Harari, M. B., Reaves, A. C., Beane, D. A., Laginess, A. J., & Viswesvaran, C. (2018). Personality and expatriate adjustment: A meta-analysis. *Journal of Occupational and Organizational Psychology*, 91(3), 486-517. <https://doi.org/10/gdmb6c>
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161. <https://doi.org/10/c62z76>
- Humphrey, S. E., Nahrgang, J. D., & Morgeson, F. P. (2007). Integrating motivational, social, and contextual work design features: A meta-analytic summary and theoretical extension of the work design literature. *Journal of Applied Psychology*, 92(5), 1332-1356. <https://doi.org/10/dch52t>
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594-612. <https://doi.org/10/bt4t68>
- Jones, J. A., & Waller, N. G. (2013). Computing confidence intervals for standardized regression coefficients. *Psychological Methods*, 18(4), 435-453. <https://doi.org/10/gckfx4>
- Kostal, J. W., Wiernik, B. M., Albrecht, A.-G., & Ones, D. S. (2018). Expatriate personality: Facet-level comparisons with domestic counterparts. In B. M. Wiernik, H. Rieger, & D. S. Ones (Eds.), *Managing expatriates: Success factors in private and public domains* (pp. 33-52). Opladen, Germany: Budrich. <https://doi.org/10/c4xg>
- Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, 62(1), 28-30. <https://doi.org/10/ckc2>

- Le, H., Oh, I.-S., Schmidt, F. L., & Wooldridge, C. D. (2016). Correction for range restriction in meta-analysis revisited: Improvements and implications for organizational research. *Personnel Psychology, 69*(4), 975-1008. <https://doi.org/10/gckf2p>
- Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new meta-analytic procedure. *Psychological Methods, 11*(4), 416-438. <https://doi.org/10/fv7g5s>
- Meyer, S. L. (1975). *Data analysis for scientists and engineers*. New York, NY: John Wiley.
- Murray, A. L., Johnson, W., McGue, M., & Iacono, W. G. (2014). How are conscientiousness and cognitive ability related to one another? A re-examination of the intelligence compensation hypothesis. *Personality and Individual Differences, 70*, 17-22. <https://doi.org/10/f6gvt6>
- Oehlert, G. W. (1992). A note on the delta method. *American Statistician, 46*(1), 27-29. <https://doi.org/10/d27qpx>
- Ones, D. S., Sinangil, H. K., & Wiernik, B. M. (2018). Validity of Big Five personality traits for expatriate success: Results from Turkey. In B. M. Wiernik, H. Rüger, & D. S. Ones (Eds.), *Managing expatriates: Success factors in private and public domains* (pp. 83-102). Opladen, Germany: Budrich. <https://doi.org/10/c4xh>
- Ones, D. S., & Viswesvaran, C. (2003). Job-specific applicant pools and national norms for personality scales: Implications for range-restriction corrections in validation research. *Journal of Applied Psychology, 88*(3), 570-577. <https://doi.org/10/c2rvwx>
- Oswald, F. L., & McCloy, R. A. (2003). Meta-analysis and the art of the average. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 311-338). Mahwah, NJ: Lawrence Erlbaum. <https://doi.org/10/ckc3>
- Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68*(3), 382-395. <https://doi.org/10/fq5thj>
- Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*(3), 432-446. <https://doi.org/10/dcrgkf>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27-42. <https://doi.org/10/gcvj3r>
- Rousseau, D. M., & Fried, Y. (2001). Location, location, location: Contextualizing organizational research. *Journal of Organizational Behavior, 22*(1), 1-13. <https://doi.org/10/dcvv7s>
- Sackett, P. R., Lievens, F., Berry, C. M., & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology, 92*(2), 538-544. <https://doi.org/10/ch6qk2>
- Sackett, P. R., & Ostgaard, D. J. (1994). Job-specific applicant pools and national norms for cognitive ability tests: Implications for range restriction corrections in validation research. *Journal of Applied Psychology, 79*(5), 680-684. <https://doi.org/10/fkpnv2>
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112-118. <https://doi.org/10/c6npmd>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage. <https://doi.org/10/b6mg>
- Stein, S. K., & Barcellos, A. (1992). *Calculus and analytic geometry* (5th ed.). New York, NY: McGraw-Hill.
- Wiernik, B. M., & Dahlke, J. A. (in press). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artefacts. *Advances in Methods and Practices in Psychological Science*.
- Wiernik, B. M., & Kostal, J. W. (2019). Protean and boundaryless career orientations: A critical review and meta-analysis. *Journal of Counseling Psychology, 66*(3), 280-307. <https://doi.org/10/gfxrgp>
- Wiernik, B. M., Kostal, J. W., Wilmot, M. P., Dilchert, S., & Ones, D. S. (2017). Empirical benchmarks for interpreting effect size variability in meta-analysis. *Industrial and Organizational Psychology, 10*(3), 472-479. <https://doi.org/10/ccnv>

Author Biographies

Jeffrey A. Dahlke is a research scientist at the Human Resources Research Organization (HumRRO). His research interests include predictive bias, adverse impact, validity degradation, predictor-weighting strategies, and meta-analytic methods. He is the creator and lead developer of the *psychmeta* package for *R*.

Brenton M. Wiernik is an assistant professor of industrial-organizational psychology at the University of South Florida. His research focuses on applied individual differences assessment, quantitative research methods, and meta-analysis. He is a developer of the *psychmeta R* package for psychometric meta-analysis.