2019

Pooling Correlation Matrices Corrected for Selection Bias: Implications for Meta-analysis

Kenneth Matthews
University of Central Florida

POOLING CORRELATION MATRICES CORRECTED FOR SELECTION BIAS: IMPLICATIONS FOR META-ANALYSIS

by

KENNETH J. MATTHEWS
B.A. Stetson University, 2005
M.A. University of South Florida, 2008

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Learning Sciences and Educational Research
in the College of Community Innovation and Education
at the University of Central Florida
Orlando, Florida

Summer Term
2019

Major Professor: Stephen A. Sivo

ABSTRACT

Selection effects systematically attenuate correlations and must be considered when performing meta-analyses. No research domain is immune to selection effects, evident whenever self-selection or attrition take place. In educational research, selection effects are unavoidable in studies of postsecondary admissions, placement testing, or teacher selection. While methods to correct for selection bias are well documented for univariate meta-analyses, they have gone unexamined in multivariate meta-analyses, which synthesize more than one correlation from each study (i.e., a correlation matrix). Multivariate meta-analyses of correlations provide opportunities to explore complex relationships and correcting for selection effects improves the summary effect estimates. I used Monte Carlo simulations to test two methods of correcting selection effects and evaluate a method for pooling the corrected matrices. First, I examined the performance of Thorndike's corrections (for both explicit and incidental selection) and Lawley's multivariate correction for selection on correlation matrices when explicit selection takes place on a single variable. Simulation conditions included a wide range of selection ratios, samples sizes, and population correlations. The results indicated that univariate and multivariate correction methods perform equivalently. I provide practical guidelines for choosing between the two methods. In a second Monte Carlo simulation, I examined the confidence interval coverage rates of a Robust Variance Estimation (RVE) procedure when it is used to pool correlation matrices corrected for selection effects under a random-effects model. The RVE procedure empirically estimates the standard errors of the corrected correlations and has the advantage of having no distributional assumptions. Simulation conditions included τ^2 ratio, within-study sample size, number of studies, and selection ratio. The results were mixed, with RVE performing well under higher selection ratios and larger unrestricted sample sizes. RVE performed consistently across values of τ^2. I recommend applications of the results, especially for educational research, and opportunities for future research.

May whatever good comes from this work most of all benefit the education of students through which their lives may be enriched and opportunities expanded.

ACKNOWLEDGMENTS

With some reflection, we can trace every accomplishment back to the kindness of others. While personal effort is important, we really can do very little without the contributions of others. Some provide us with training and knowledge, some with encouragement, some with an ear, and still others with the simplicity of care. I owe a great deal of thanks to many people who have contributed to where I stand today including family, friends, colleagues, and teachers.

At the top of my list of thanks are my parents, Beverly and Gary Matthews, who never ceased to encourage my learning and, more importantly, to acknowledge my potential for as long as I have memories to draw upon. While he is not here to share in this moment, my father would be proud—and probably relieved that I did not stop at studying philosophy. My mother continues to provide endless encouragement. Whether I am running off to live in a monastery or leaving a promising career to become a full-time student, she has always provided me the space to find my own way and there is no greater gift from a parent and, really, to the research enterprise since independent thought and exploration are at its heart.

I want to thank Dr. Malcolm Butler who, despite the fact that I invited him to my committee late in the game, made important contributions to this work by always returning my focus to the practical applications of my research. I have had the privilege to work with my other committee members for several years. Their contributions extend beyond specific feedback about this work to years of modeling scholarship that has shaped how I approach research. Dr. Haiyan Bai has modeled the spirit of pushing just a little further. Dr. David Boote has modeled the spirit of independent thought and healthy skepticism. Dr. Debbie Hahs-Vaughn has modeled the spirit of productivity and goal-driven work necessary in academia. Finally, Dr. Stephen Sivo has modeled the spirit of exploration and curiosity that I have always treasured in learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Meta-analysis is a method for statistically combining the quantitative results of multiple studies (Cooper, Hedges, & Valentine, 2009; Glass, 1976; Schmidt & Hunter, 2015). While researchers can synthesize any parameter (i.e., averaged), they typically focus on effects sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hafdahl, 2007). Meta-analyses provide a summary of quantitative research findings that could otherwise be misleading if one reviewed them qualitatively (Chalmers, 2007; Glass, 1976; Schmidt, 1992; Schmidt & Hunter, 2015). Implemented well, the methods that underlie meta-analysis permit us to accumulate the results of sometimes apparently contradictory studies to arrive at a more robust conclusion about the effect of interest. Thus, they provide a distinct advantage over purely narrative systematic reviews. Presumably, as a result of this advantage, meta-analyses have become ubiquitous. The search results for the word *meta-analysis* in the abstracts of works indexed in ERIC, MEDLINE, and PsycINFO evidence the growth in publications that are themselves meta-analyses or are related to them (e.g., methodological works). Figure 1 illustrates that since 2008, the annual number of studies related to meta-analysis indexed in these databases has nearly quintupled from just over 4,000 to nearly 20,000.

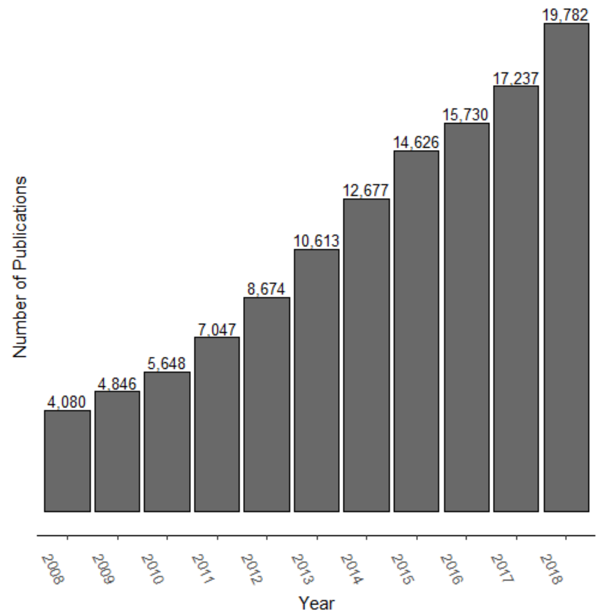

Figure 1. Annual count of meta-analysis studies indexed in ERIC, MEDLINE, and PsycInfo.

Meta-analyses have also become highly influential. In medicine, for example, Patsopoulous, Analatos, and Ioannidis (2005) found that after controlling for year and journal of publication, meta-analyses had the highest citation count among medical publications published between 1991 and 2001. Furthermore, their lead over citations of studies using randomized control trials—the second most cited design—grew during the same period. They have been used by policymakers to inform their decisions (Borenstein et al., 2009; Schmidt & Hunter, 2015) and the founding of the Cochrane Collaboration in 1994 and Campbell Collaboration in 1999 provide further evidence of the influence of meta-analyses in research and policy. The mission of both organizations is to influence policy-making by supporting and cataloging high-quality meta-analyses (Campbell Collaboration, 2019; Cochrane Collaboration, 2019).

Meta-analysis promises to bring greater certainty to sprawling, heterogeneous bodies of literature by statistically combining parameter estimates—usually effect sizes—reported in a series of related studies (Cooper et al., 2009; Glass, 1976; Hedges & Olkin, 1985). Today, methods exist

that permit us to synthesize the parameters arising from intricate patterns or relationships. We can use meta-analysis methods not merely to summarize single relationships but as a basis to test theories (Becker, 2009; M. W.-L. Cheung & Chan, 2005; Viswesvaran & Ones, 1995). It is with these methods that meta-analysis provides its most significant promise for the research enterprise—to engender greater confidence in our understanding of theoretical relationships.

We can use meta-analytic correlations as input for a multiple regression analysis, path analyses, and SEM. This, in turn, permits us to study the partial effects of the predictors, indirect effects, and model comparisons (Becker, 2009). As quantitative summaries of the extant literature, the correlations forming the matrix theoretically provide the best basis available for testing a theoretical model. This is because the correlations themselves are estimated across numerous studies, reducing sampling error in the estimated correlations (Schmidt & Hunter, 2015).

The key to pooling correlation matrices to form the input matrix for such analyses is accounting for the fact that the correlations within each study-level matrix arise from the same sample and are, thus, themselves correlated (Steiger, 1980). This dependency needs to be modeled to avoid biases in the standard errors of the pooled estimate (Raudenbush, Becker, & Kalaian, 1988; Shadish, 1996). The methods for pooling correlation matrices have all been developed in what I refer to as the generalized meta-analysis literature. Becker (1992) first introduced a method of pooling using generalized least squares (GLS). Since that time, a great deal of research in this body of literature has been devoted to multivariate methods of meta-analysis applicable to pooling correlation matrices or any other set of effect sizes arising from the same sample (Aloe & Becker, 2012; Becker & Fahrbach, 1994; M. W.-L. Cheung, 2008; Viswesvaran & Ones, 1995). Multivariate meta-analyses refer to those times when multiple effects are the outcomes of interest (Becker, 2007; Riley, 2009; Wei & Higgins, 2012). These can be contrasted with univariate meta-analyses, which have only a single effect as the outcome of interest.

While pooling correlation matrices permits advanced analyses (e.g., path analysis, SEM), when we wish to synthesize correlation coefficients, we need to consider their accuracy as estimates of the target population parameters. It is well known from psychometric literature that correlation coefficients are biased estimates when they are affected by what Schmidt and Hunter (1977) refer to as artifacts. They identified 11 such artifacts when they developed their unique meta-analysis framework, called psychometric meta-analysis.

Selection bias is one artifact that attenuates correlation estimates (Sackett & Yang, 2000). It is present when observations on a variable are systematically excluded as a result of some process that results in non-random sampling (Schmidt & Hunter, 2015). No domain of research is immune to selection bias, but they are particularly important in research on selection processes such as postsecondary admissions (M. J. Allen & Yen, 1979; Crocker & Algina, 2008; Schmidt & Hunter, 2015), personnel selection such as teacher recruitment (Kimbrel, 2019), and placement testing (Waschull, 2018) where selection effects are inherently present. Schmidt and Hunter (2015, p. 92) state that measurement error and selection effects are among the most well-studied artifacts in psychometric literature providing some indication that they present the most common problems in applied research. There are only two studies of corrections for measurement error applied in multivariate meta-analysis (S. F. Cheung, Chan, & Sun, 2018; Schram, 1995) and there is no prior research on the integration of corrections for selection bias and multivariate meta-analysis.

The psychometric meta-analysis literature details corrections for selection bias so that summary effect estimates are closer to their actual population values, but it is an exclusively univariate methodology (Schmidt & Hunter, 2015). In other words, the psychometric meta-analysis synthesizes only a single effect at a time even when multiple effects arise from the same study. As a result, the psychometric meta-analysis literature does not provide research on the application of artifact corrections to pooling correlation matrices—only to individual correlations. On the other hand, the generalized meta-analysis literature that does offer pooling methods virtually ignores the application of statistical artifact corrections in the multivariate

setting. As a result, the extent to which the psychometric meta-analysis methods are appropriate in the multivariate context is unclear.

Purpose

With only two studies (S. F. Cheung et al., 2018; Schram, 1995) to my knowledge, there is a dearth of research on the application of artifact corrections described in psychometric meta-analysis to multivariate meta-analysis described in generalized meta-analysis and none that examine the application of corrections for selection effects specifically although they are known to improve correlation estimates (Duan & Dunlap, 1997; Greener & Osburn, 1979; Schmidt & Hunter, 2015; Thorndike, 1949). Two bodies of literature have developed along different trajectories with different emphases leaving this gap in the methodological research. Importantly, we cannot assume that procedures that work in the univariate context where each study supplies only one correlation also work in the multivariate setting where studies provide more than one correlation (Hafdahl, 2007). Given the ubiquity and importance of meta-analytic studies, it is worthwhile to open a research agenda that seeks to integrate generalized and psychometric meta-analysis procedures.

The two primary goals of a meta-analysis are to compute a weighted mean effect as an estimate of the population effects and to estimate the variance of effects in the population (Borenstein et al., 2009; Schmidt & Hunter, 2015). In a univariate meta-analysis, we estimate a single effect and its variance. In a multivariate meta-analysis, we estimate multiple effects and their variances. Pooling correlation matrices (i.e., multiple correlations per study) corrected for selection bias calls for us to consider conditions unique to the multivariate setting. These conditions occur at two stages in the meta-analysis. First, we correct the correlation matrix from each study for selection effects. Then, we must pool the corrected matrices to provide a mean correlation matrix that serves as an estimate of the population matrix and we must estimate the

variances in the effects.

At the correction stage of the meta-analysis, I was concerned with the performance of the methods available for correcting each study matrix for selection bias. Correlation matrices have specific properties such as being positive definite and each correlation within the matrix having restraints on its possible values given the values of the other correlations within the matrix (Hubert, 1972; Olkin, 1981; Stanley & Wang, 1969). Standard psychometric meta-analysis correction methods only adjust one correlation at a time (i.e., element-wise) independently of the other elements in the matrix, ignoring these essential properties of the correlation matrix. As a result, applying standard univariate corrections to a correlation matrix introduces the potential that the resulting matrix will not be a valid correlation matrix (Schram, 1995). Although a multivariate correction for selection effects exists, methodologists in the psychometric meta-analysis literature have not given it any consideration because psychometric meta-analysis is a completely univariate framework (Schmidt & Hunter, 2015) and despite that fact that it is applied in practice (Klieger, Cline, Holtzman, Minsky, & Lorenz, 2014; Powers, 2004; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015). Therefore, it is unclear how well the univariate and multivariate methods perform and whether the multivariate procedure provides adequate advantages to justify the greater computational complexity and informational requirements. I investigate this problem in the first stage of this research.

In the second stage of the research, I considered the second stage of the meta-analysis and the fact that synthesizing correlation matrices that we have corrected for selection effects implies that we harvest correlations calculated from common samples. As a result, these corrected correlations are themselves correlated (Becker, 1992; Raudenbush et al., 1988; Steiger, 1980). We need to account for this correlation among correlations by properly estimating the within-study covariance matrix so that we arrive at proper estimates of the total variance in study effects (Becker, 2009; Gleser & Olkin, 2009; Raudenbush et al., 1988). It is well known that the standard errors of corrected correlations are larger than their uncorrected counterparts (Raju & Brand,

2003; Schmidt & Hunter, 2015). It is reasonable, therefore, to assume that their covariances are also larger. While there are closed solutions for estimates of the standard errors of corrected correlations, there are no published closed solution estimators of their covariances. Thus, we lack a method to accurately estimate the within-study covariance matrix necessary to pool the correlated corrected correlations properly. Robust variance estimation (RVE) (Hedges, Tipton, & Johnson, 2010) provides an alternative to directly estimating the within-study covariance. RVE empirically estimates within-study covariances to model correlated effects without knowledge of the covariance structure. It results in accurate estimates (Hedges et al., 2010; Park & Beretvas, 2018) but has not been studied with correlations, let alone those corrected for selection bias.

In summary, correcting correlation matrices for selection effects introduces concerns with (a) the accuracy and validity of the resulting matrix after correction and (b) pooling the correlation matrices in a way that properly models the covariances of the effects. I investigated four research questions in response:

RQ1. As indicated by absolute median bias, how well do univariate, element-wise corrections for selection recover the unrestricted correlations compared to a multivariate, simultaneous correction when they are both applied to a matrix of correlations?

RQ2. How frequently do univariate, element-wise corrections for selection produce inadmissible vectors of correlations compared to a multivariate, simultaneous correction when applied to a matrix of correlations?

RQ3. How frequently do univariate, element-wise corrections for selection result in a non-positive definite correlation matrix compared to a multivariate, simultaneous correction when applied to a matrix of correlations?

RQ4. What are the coverage rates of the confidence intervals produced by robust variance estimation when pooling correlations corrected for selection effects?

Significance and Applications

From a methodological standpoint, this research has the potential to make several contributions. In the broadest terms, the integration of psychometric and generalized meta-analysis will provide a basis for a unified framework of meta-analysis. The current literature is segregated with artifact corrections being the near-exclusive domain of psychometric meta-analysis and multivariate methods being the exclusive domain of generalized meta-analysis. Both bodies of literature contribute unique components for accurately estimating parameters of interest in a meta-analysis. Psychometric meta-analysis provides better estimates of the summary effects in the presence of artifacts, and generalized meta-analysis provides more accurate standard errors in the presence of correlated effects. Practitioners ought to have the best of both worlds. Indeed, researchers sometimes hybridize the frameworks without the benefit of thorough methodological research to guide meta-analysts (e.g. Kriegbaum, Becker, & Spinath, 2018; Roh, Chun, Ryou, & Son, 2018).

In terms of more specific practical applications, the artifacts that Schmidt and Hunter identified can affect correlations in any research domain. Thus, the corrections ought to be considered in every discipline and under any framework that seeks unbiased estimates of correlation coefficients. Schmidt and Hunter first worked primarily in personnel selection research investigating the predictive validity of tests used to make hiring decisions. A parallel field in education is research on postsecondary admissions. In both of these areas of research, selection effects are present by definition and corrections must be applied to draw correct inferences.

A great deal of evidence supports the predictive validity of cognitively loaded admissions tests such as the SAT, ACT, and GRE. On the other hand, there are concerns about the sole use of these tests to make admissions decisions because they tend to admit higher proportions of Asian and White students compared to Black or Latino students (Credé & Kuncel, 2008; Kuncel,

Hezlett, & Ones, 2001; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt et al., 2009). As a result, research incorporating non-cognitive measures and other data into admissions decisions is ongoing. Schmitt et al. (2009) evaluated the predictive validity of both cognitive test scores (SAT and ACT) and non-cognitive scores on several measures of academic performance in a single sample. One of their significant conclusions was that incorporating non-cognitive measures into the selection process would improve the proportion of Black and Latino students admitted, especially at schools considered to be more selective. Of those studies cited that are meta-analyses (Credé & Kuncel, 2008; Kuncel et al., 2001), all of them utilized univariate corrections for selection bias in their studies.

In another study, Zwick and Himelfarb (2011) found that incorporating high school SES removed the over-prediction of the first-year GPA for Black and Latino students, who disproportionately come from low SES schools. While this would suggest that a smaller proportion of those students would be admitted because their predicted GPAs are adjusted downward, the new information could be used to identify admitted students who might need early interventions for success. Assuming no effective interventions are available, a denial of admission when someone is unlikely to succeed ultimately saves time and resources. Students who do not succeed, accumulate debt, permanently hurt their GPAs and might view themselves as incapable of succeeding in college when, in fact, they were not adequately prepared. Students who find themselves in this situation may experience long-term consequences for the student's livelihood and social mobility.

While this study was being planned and carried out, the College Board announced that it will begin to assign an adversity score[1] to accompany SAT scores (Hartocollis, 2019). The new index is provided through an Environmental Context Dashboard (ECD), which provides schools with information about a student's neighborhood and school such as the average income for

[1]The words *adversity score* do not appear on the College Board website, but in media reports.

families in the area and the opportunities such as AP offerings by the school (The College Board, 2019). The College Board touts it as a method by which the cognitive SAT score can be put into context (The College Board, 2019). For example, if a student applies from a school and neighborhood with a higher index on the ECD, an admissions officer might interpret less-than-ideal scores as success in the face of adversity. Some 50 institutions have piloted the use of this score in their admissions decisions, and some have used it to change their decisions (Jaschik, 2019).

Given the gravity of admissions decisions and the interest in making them equitable, validity studies in this domain that incorporate multiple predictors are necessary and pooling their results in a comprehensive meta-analysis will necessitate the integration of both generalized and psychometric meta-analysis methods. Enhancing available methods by improving both mean validity coefficient estimates and more reliable hypothesis tests with correct standard errors should benefit both students and institutions by permitting more reliable the predictions of students' success. I discuss applications to teacher selection and placement testing as well in Chapter 5. In the chapter that follows, I expand upon the meta-analysis literature to demonstrate the origin of the gap I have identified. I then discuss selection effects in research, their correction, and the obstacles to integrating them into the multivariate meta-analysis context.

CHAPTER 2: LITERATURE REVIEW

Researchers in both personnel selection and education have used correlation matrices derived from meta-analytic estimates as input for path models to test theory since the 1980's. The earliest examples come from personnel selection research. Hunter (1983) examined the relationships between cognitive ability, job knowledge and performance, and supervisor ratings. Premack and Hunter (1988) performed 14 meta-analyses on the relationships among six variables to test their theoretical model of unionization decisions. They assembled the 14 correlations into a single matrix and tested their theory with a path model. In education, most early examples are found as illustrations in methodological pieces by Becker and, later, Becker and Schram (Becker, 1992; Becker & Schram, 1994) with an applied example in Whiteside and Becker (2000) who examined multiple outcomes for children of divorced families including the occurrence of behavioral problems, social skills, and internalizing and externalizing symptoms. Their path model is the most complex among these examples with a number of indirect effects being tested. It was only through this more complex meta-analysis that they uncovered important relationships that did not appear in a simple bivariate analysis. For example, they showed that prior narrative reviews were incorrect in concluding there were only weak relationships between father visitation and the children's outcomes (see also Becker, 2009).

Testing models rather than simple relationships offers a variety of benefits. Consider a simple model to illustrate. Assume we are interested in predicting graduate school GPA (GGPA) from undergraduate GPA (UGPA), GRE composite (GRE), and socio-economic status (SES). If we simply synthesize the six correlations among the four variables, we are left with a set of zero-order correlations. These can be informative for describing the general relationships among the variables but provides no insight into the simultaneous effects of UGPA, GRE score, and SES on GGPA. Therefore, we learn little that helps us to predict or explain GGPA effectively. If, however, we use the meta-analytic correlations as input for a multiple regression analysis, we can

study the partial effects of the predictors. This is an improvement over the zero-order correlations because each predictor-outcome relationship is controlled for the other covariates in the model. In addition to examining partial effects, we can test for indirect effects. By testing for moderators and mediators, we enhance our ability to fully explain an outcome. For example, we might model SES as a mediator of the relationship between GRE scores and GGPA, a moderator, or both (Baron & Kenny, 1986). Other advantages include the ability to compare different models and to discover those relationships that have yet to be tested in the literature (Becker, 2009). Meta-analytic correlation matrices offer a basis for performing secondary analyses such as a path analysis or SEM. As quantitative summaries of the extant literature, the correlations forming the matrix theoretically provide the best basis available for testing a theoretical model. This is because the correlations themselves are estimated across numerous studies, reducing sampling error in the estimated correlations (Schmidt & Hunter, 2015).

Authors have used several names to refer to the use of meta-analytic correlation matrices as the basis for model testing. Becker, whose research has focused on GLS methods (Becker, 1992), refers to this approach as *model-driven meta-analysis* (Becker, 2001, 2009), while Lipsey (1997) referred to it as *linked meta-analysis*. Others refer more directly to the statistical methods used to estimate and subsequently analyze the matrices. *Meta-analytic structural equation modeling* (MASEM) and *two-step meta-analytic structural equation modeling* (TSSEM) are commonly used in the literature (M. W.-L. Cheung & Chan, 2005). MASEM and TSSEM are variants of model-driven meta-analysis that use SEM methods to both pool correlation matrices and perform the secondary analysis. Finally, others use what I refer to as an *assembly approach* first proposed by Viswesvaran and Ones (1995) to construct the input matrix, pulling together meta-analytic correlations from any number of sources and assembling them into a single matrix. The assembly method is an ad hoc procedure, not a statistical pooling method. Those working exclusively in the psychometric meta-analysis framework have this assembly method as their only alternative since, as I show later, multivariate meta-analyses are not well considered by Schmidt and Hunter

(2015). Throughout this work, I adopt the more general term *model-driven meta-analysis* (MDM) to refer to pooling correlation matrices. The term captures the intent of all of the approaches and is, therefore, more general.

Although rigorous methods for pooling the corrected study-level correlation matrices are available, no work has been done to examine pooling methods when the correlations have been corrected for selection effects. In fact, there is virtually no research on pooling correlation matrices that have been corrected for various well-known attenuating factors. Hunter, Schmidt, and Jackson (1982) identified 11 biasing factors that they refer to as statistical artifacts. Some of these include measurement error (unreliability), range restriction (caused by a selection process), and artificial dichotomization of continuous variables. In the presence of these artifacts, correlations involving the variables they affect will be biased estimates of their population parameters (Schmidt & Hunter, 2015).

I could locate only two studies that examined the multivariate synthesis of correlations corrected for artifacts. Schram (1995) examined univariate and multivariate corrections for measurement error and found that the univariate corrections provided good estimates of the population values, but, as expected, the standard errors and covariances among the correlations were underestimated by univariate meta-analysis procedures, sometimes by as much as 50%. She also found that improper correlation matrices occurred more frequently when univariate corrections were applied compared to multivariate corrections.

S. F. Cheung et al. (2018) examined a method for synthesizing dependent correlations attenuated due to measurement error or dichotomization based on what they refer to as a sample-wise procedure for accommodating dependence (S. F. Cheung & Chan, 2004, 2008). In their simulation study, they varied the degree of dependence among the within-study correlations, heterogeneity between studies, degree of attenuation, within-sample study size, and number of total studies. Like Schram (1995), they found that estimates of the population correlations were typically accurate using typical psychometric meta-analysis corrections. They found that their

method worked well when corrections for attenuation were made and poorly without the corrections, which is to be expected since we already know that uncorrected correlations are biased. One exception was when there was no between-study heterogeneity in the effects. None of the methods they used performed well in that case, but it was not clear whether or not the problem was with their procedure or because they applied a generalized meta-analysis to the corrected correlations.

S. F. Cheung et al. (2018) illustrated their method using a data set with 12 studies originally reported in S. F. Cheung and Chan (2014). They randomly generated attenuation factors and applied them to the data. They found that the heterogeneity between studies was not significant after they corrected for attenuation. Despite the limitations in the illustration—only 12 studies, thus low power, and no reported source for the 12 studies—it is worth noting that their finding is exactly what would be predicted by psychometric meta-analysis. Hunter and Schmidt's motivation for their work was their hypothesis that the fluctuation of validity coefficients from study to study could be explained, nearly completely, by the fact that the variation was introduced by statistical artifacts.

I attribute the lack of integrative research to the fact that methods of multivariate meta-analysis and methods to correct for artifacts in meta-analysis have developed in two distinct bodies of literature. While methodologists are certainly aware of both literatures, little has been done to integrate them. I now provide a sketch of the literature as it appears today. My purpose is not simply to summarize, but to provide the reader with some understanding of the meta-analysis landscape and the genesis of this gap. Stated differently, there is no good argument against further research into the multivariate synthesis of corrected correlations. Rather, it seems simply not to have been done.

Two Meta-analysis Literatures

Stated simply, the purpose of meta-analysis is to statistically combine parameter estimates from a series of studies and examine, or possibly explain, their variability (Borenstein et al., 2009; Cooper et al., 2009; Hedges & Olkin, 1985; Schmidt & Hunter, 2015). Typically, the estimand is a population effect size, estimated by a weighted mean of observed effects. The variance of the effects is then used to characterize the estimate via confidence and/or credibility intervals and to assess whether the observed variance is greater than that expected due to sampling error alone. When the variance is greater than what would be expected by sampling error alone, it is taken as evidence that some other variables are moderating the observed effects. These basic steps characterize all forms of modern meta-analysis, which are broadly divisible into three camps: (a) what I refer to as generalized meta-analysis stemming primarily from the seminal work of Hedges and Olkin (1985), (b) psychometric meta-analysis, founded in the seminal works of Schmidt and Hunter (1977) and Hunter et al. (1982), and (c) individual patient data meta-analysis (IPD), which has undergone rapid development in medicine (Cooper & Patall, 2009; Kontopantelis, 2018; Riley, Lambert, & Abo-Zaid, 2010; Stewart & Tierney, 2002).

I use *generalized meta-analysis* to refer to the fact that its methodological literature has as its underlying motivation the development of generalized methods applicable to all effect sizes (Tipton, Pustejovsky, & Ahmadi, 2019). On the other hand, Schmidt and Hunter's psychometric meta-analysis is almost exclusively concerned with artifact corrections and focuses primarily on correlation coefficients. For the purposes of this research, I set aside IPD techniques. With IPD type meta-analyses, raw data sets from primary studies are combined and analyzed as one single set of data. IPD approaches offer opportunities not afforded to the traditional meta-analyst (Stewart & Tierney, 2002), but they are rare outside of medicine so I focus on meta-analysis techniques without the benefit of access to primary study data, which is still the norm in most disciplines.

Today, generalized and psychometric meta-analysis are the dominant frameworks within

which nearly all meta-analysis methods can be classified. At its inception, however, several variants of meta-analysis were being developed simultaneously. While one can barely read a methodological meta-analysis article or text without coming across some form of the phrase *Glass coined the term meta-analysis*, its methods were actually being independently developed at the same time by Glass, Rosenthal, and Schmidt and Hunter (Shadish & Lecy, 2015). Each was motivated by a similar problem in their respective disciplines—an overwhelming body of literature at conflict with itself in the form of divergent conclusions regarding similar hypotheses. It was through their pursuit of meta-analysis that these methodologists brought clarity to substantive problems in ways that changed the landscape of their disciplines and the social science research enterprise as a whole. With the ubiquity of meta-analyses in publication today, it might be difficult to appreciate the potential it has. Early examples help to highlight the fact that significant contributions can be made by well done meta-analytic studies and provide evidence that further development of this methodology is certainly worthwhile. Examining the early development of the literature also helps to suggest why artifacts such as selection effects have not been well considered in multivariate meta-analysis contexts.

Generalized meta-analysis. For several decades, researchers and practitioners of psychotherapy concluded that it was virtually ineffective (Smith & Glass, 1977). This view entered mainstream psychology after Eysenck (1965) examined reports of the differences in recovery rates of neurotic individuals who did and who did not receive psychotherapy. He systematically reviewed literature on the recovery rates of the two groups and found that there was no discernible difference, declaring that psychotherapy had no apparent benefit for neurotic patients. Unfortunately, he applied a vote counting method to the findings in the literature by which one simply counts and categorizes findings. This can be misleading because, typically, the counting relies upon statistical significance without accounting for study sample sizes or the size of the observed effects. Thus, underpowered studies might be taken as evidence of the absence of

an effect—a wrong conclusion on several levels—when, in fact, an effect does exist. In response to Eysenck's findings, Smith and Glass (1977) applied the meta-analysis methods described by Glass (1976) in his seminal meta-analysis article to 375 studies synthesizing 833 effect sizes. They showed that, contrary to Eysenck's findings, the effects of psychotherapy were actually quite beneficial with patients engaged in psychotherapy having, on average, an outcome .68 standard deviations greater than those without psychotherapy. Put another way, their results showed that 75% of those who engaged in psychotherapy had better outcomes than those who did not. Furthermore, they found that this effect held regardless of the therapeutic approach (e.g. Freudian vs Gestalt). Smith and Glass' conclusions were unabashedly direct:

> Scholars and clinicians are in the rather embarrassing position of knowing less than has been proven, because knowledge, atomized and sprayed across a vast landscape of journals, books ,and reports, has not been accessible. Extracting knowledge from accumulated studies is a complex and important methodological problem which deserves further attention. (p. 760).

Around the same time, Robert Rosenthal was publishing methods for statistically combining probabilities (i.e., p-values) and effect sizes (Rosenthal, 1978; Rosenthal & Rubin, 1978). Aggregating p-values was a unique focus of Rosenthal's work, which he later supplemented with mean effect sizes (Bangert-Drowns, 1986). His paper with Rubin on interpersonal expectancy effects was the result of a synthesis of 345 studies. They concluded that the existence of expectancy effects was basically certain and that the effects in some research domains were very large (d = 1.73). As a result of this synthesis, Rosenthal was able to give greater credence to the existence and impact of expectancy effects—a phenomenon that challenged the objectivity of even experimental research (Shadish & Lecy, 2015).

Following the initial developments by Glass and Rosenthal and Rubin, Larry Hedges along with Ingram Olkin began to rigorously develop the distribution theory that would become

the foundation of generalized meta-anlysis (Tipton et al., 2019). Their seminal 1985 work was a compilation of distribution theory for a variety of effect sizes and included parametric and non-parametric methods of synthesis, robust estimation methods, and hypothesis tests. They also presented a multivariate synthesis model for those situations in which correlated effects are harvested from the same sample. Generalized meta-analysis methodologists, exclusively, have expanded multivariate methods of meta-analysis including their development of statistical methods of pooling vectors of correlations.

The generalized meta-analysis literature is not totally absent references to artifacts and their corrections. Hedges and Olkin (1985, chapter 6.G.) provide a discussion and example of an analysis correcting correlations for measurement error and imperfect construct validity. Raudenbush (2009) presents a random-effects model of meta-analysis noting, in passing, that the effects may have been corrected for artifacts. Tanner-Smith, Tipton, and Polanin (2016) corrected correlations for artifacts in a methodological demonstration of robust variance estimation, which is a multivariate meta-analysis procedure. They did not, however, use simulated data in their demonstration so would be unaware of problems with the analysis. Finally, Schmidt, Le, and Oh (2009) provided a single, short discussion of artifact corrections in the handbook edited by Cooper et al. (2009) that otherwise presents meta-analysis from a generalized meta-analysis approach. So it is not that those working in the generalized meta-analytic tradition are unaware of psychometric meta-analysis. Instead, what is absent from this literature is rigorous research and guidance on integrating psychometric meta-analysis corrections with multivariate methods of meta-analysis.

Psychometric meta-analysis. Simultaneous with Glass and Rosenthal's research, Schmidt and Hunter were developing their method of validity generalization. In their seminal work, Schmidt and Hunter (1977, 2003) challenged the widely held view that validity coefficients (i.e. correlations) in personnel selection were so situation specific that it was impossible to make

any generalizations about them even when the tests and jobs being considered were very similar. This view was based on the fact that observed validity coefficients were highly variable even under seemingly similar conditions. Conventional wisdom held that the variation was due to differences that simply could not be articulated (Schmidt & Hunter, 1977). As a consequence, researchers' ability to arrive at a generalized model to describe personnel selection processes was severely limited, even within similar test-job combinations. For practitioners, this also meant that a validity study, consuming time and other resources, ought to be conducted every time a test was used in a new setting no matter how similar it was to prior administrations. Schmidt and Hunter posited—and showed in their 1977 work—that, in fact, much of the variability observed in validity coefficients was not the result of situation-specific true validities but of statistical artifacts affecting the observed correlations. As a result, they developed a meta-analysis framework that uniquely focuses on corrections for statistical artifacts in observed effects to improve meta-analytic estimates and reduce their artificial variability.

Psychometric meta-analysis is a completely univariate framework. Schmidt and Hunter (2015) specifically recommend that correlated effects within studies be either reduced to a single correlation by combining them into a composite if they are conceptually related or performing a separate analysis on each relationship if they are conceptually distinct. They discuss problems associated with correlated effects, but stop short of recommending multivariate meta-analysis techniques. I discuss their views in more detail later in the text. At this point, what is important to understand is that multivariate methods are never recommended in the psychometric meta-analysis framework. It is easy to see why psychometric meta-analysts would not consider artifact corrections in the multivariate context.

It should be clear from this brief sketch that artifact corrections and methods of multivariate meta-analysis developed in two bodies of literature. Integrating them has the potential to provide a single framework within which to study and conduct meta-analyses. Perhaps more importantly, my review of the literature has revealed that generalized and psychometric meta-analysis methods are

sometimes combined in an ad hoc fashion. Despite this, methodologists have given little rigorous consideration to their integration. As a result, for those times when the methods have been blended in practice there is little methodological research to suggest whether and how that should be done.

Roh et al. (2018) synthesized 208 correlations from 51 studies. They corrected the correlations for measurement error as described in psychometric meta-analysis. To estimate the mean effects, they transformed the correlations into a Fisher's z metric[1] first, a generalized meta-analysis recommendation that is recommended against in psychometric meta-analysis (Borenstein et al., 2009; Schmidt & Hunter, 2015). They followed up with a meta-regression analysis, regressing the observed effects on a series of predictors such as country where the study was conducted, another generalized meta-analysis procedure. Finally, as would be done in psychometric meta-analysis (Viswesvaran & Ones, 1995), they assembled the results of their series of univariate meta-analyses into a single correlation matrix rather than pooling them statistically although studies clearly reported more than one effect. They then used the matrix as input for an SEM model, a procedure found in both literatures.

In another example, Kriegbaum et al. (2018) studied the effects of intelligence and motivation on school achievement. They applied corrections for measurement error and range restriction to the correlations among all of the variables. They converted the correlations to the Fisher's z metric and ran random-effects models for each relationship. They do not detail how they pooled an input matrix for testing three path analyses, but it appears they assembled the results of the independently run random-effects models in a matrix, rather than using a statistical pooling method.

A number of questions arise. In a meta-analysis, a weighted average is computed to pool the effects of interest. Generalized meta-analysis methodologists recommend the use of the effect's inverse variance (Borenstein et al., 2009; Hedges & Olkin, 1985). When the authors used

[1]Given by $z = \frac{1}{2}\ln(\frac{1+r}{1-r})$

the Fisher's z transformation, they presumably used its corresponding variance $(n - 3)$ as a weight. However, correcting correlations for artifacts requires that we also correct their variances (Schmidt & Hunter, 2015). Therefore, a different estimator of this variance is required for corrected correlations and the only formulation applicable to Fisher's z transformed correlations that I uncovered was found to be imprecise, inflating Type I error rates by three to four times its nominal value (Raju & Brand, 2003). Because no integrative literature exists, it is not surprising that the authors did not appear to consider this problem. Furthermore, they did not apply statistical pooling methods to model dependent effects, which would have been complicated not only by the proper estimation of variances but of the covariances among the corrected correlations. Instead, they used the assembly approach by Viswesvaran and Ones (1995).

Studies such as these provide a motivation to clarify the consequences of applying artifact corrections within the generalized meta-analysis framework. In particular, these studies would have benefited from methods to integrate them into the multivariate, model-based meta-analysis framework because the studies in their samples contributed more than one effect size to the analysis. We need to better understand artifact corrections applied to multivariate meta-analyses. Here, I focus on selection effects. In the following sections, I describe how selection processes affect correlation coefficients and the problems it poses for making correct inferences. I also introduce methods for correcting for selection effects and discuss their relative advantages in the multivariate meta-analysis context.

Selection in Research

Selection processes are common phenomena in education and psychology (M. J. Allen & Yen, 1979; Schmidt & Hunter, 2015). In college admissions the SAT, ACT, TOEFL, MCAT, GMAT, and others are used to select candidates with a subset of scores on these exams (e.g. Klieger et al., 2014; Kuncel et al., 2001). In community colleges students typically take

placement tests to assess their readiness for college-level work. They act as a selection mechanism on the premise that students who are less prepared, as indicated on the placement test, are less likely to be successful. Conceptualized this way, placement tests act like tests used for admission to a university. In the community college setting admission to college-level work, rather than to the college itself. In personnel selection research, similar situations occur when employers use proficiency exams, interviews, assessment centers, or a combination of these to make hiring decisions (e.g. Joseph, Jin, Newman, & O'Boyle, 2015; Joseph & Newman, 2010).

Recruitment into research studies is also a selection process. Lawson, Hook, and Farah (2017) conducted a meta-analysis of the relationship between SES and executive function performance among children. In evaluating their sample of 25 studies, they recognized that the sampling procedures in the primary studies were likely to have an effect on the mean correlation resulting from the meta-analysis. They coded each study to indicate the target population in that study which included students with low SES, middle SES, high SES, diverse SES, and convenience sampling. They further coded the studies by whether or not the samples were heterogeneous on SES. Not surprisingly, they found that those studies with higher SES variability had a mean correlation with executive function performance of .22, while those that with low SES variability had a mean correlation of .08. Among all the moderators they tested including country where the study was done, gender, race, average age, and publication type only SES variability was a statistically significant. It is worth noting that the authors used the within-study average effect when studies reported multiple measures and, thus, did not account for the dependence among the correlations. Therefore, their hypotheses tests may not have been accurate. It is unclear why the authors chose not to correct for range restriction since their interest was in the relationship between the full range of SES values and executive functioning performance. In any case, had they left SES variability unexamined they might have been left with the impression that some substantive moderator affected the relationship. Removing this artificial variability in the effects highlights a key contribution of psychometric meta-analysis.

Self-selection can be harder to identify, but also induces selection effects. It was a concern in the meta-analysis of the relationship between psychological characteristics and academic interventions by Robbins, Oh, Le, and Button (2009), but they did not take steps to correct it due to insufficient information in the primary studies. Kriegbaum et al. (2018) were also concerned about the role of self-selection in their study of the predictive validity of motivation and intelligence on academic achievement and performed sensitivity analyses on different degrees of self-selection in the face of a lack of information on the true selection ratio.

Finally, selection on an outcome can occur through study attrition (Ones, Viswesvaran, & Schmidt, 2017) or might take place on several variables simultaneously. Powers (2004) examined the predictive validity of the GRE for admission to colleges of veterinary medicine. Since the GRE was used for admission, direct selection occurred on GRE scores. He also investigated the degree to which attrition from the programs might have further restricted the sample through the outcome. He obtained data from 12 of the 27 participating schools, all of which indicated that five or fewer students left the program. As a result, the effect of attrition in this example was negligible, but selection on both the predictor (GRE) and criterion (GPA) was observed.

Effects of selection on $\hat{\rho}_{XY}$. Whatever their form, selection processes bias estimates of population correlations when their effects result in a sample that is no longer representative of the target population (Schmidt & Hunter, 2015). Correcting biases such as these in the estimation of validity coefficients was what motivated Schmidt and Hunter to explore psychometric meta-analysis, which was originally known as validity generalization (Murphy, 2003). In studies of the predictive validity of tests used for college admission or hiring, selection processes bias estimates of the validity coefficient. A validity coefficient is simply the correlation between the test used for selection—the predictor—and a criterion or outcome. In the case of predictive validity, scores on the outcome are collected at some time later than the scores on the predictor (M. J. Allen & Yen, 1979).

Suppose we are interested in establishing the predictive validity of a test for the purposes of admission into a program of study. Because we want to select those applicants who will be successful in the program, we want to verify that the scores from the test are sufficiently correlated with an indicator of academic success such as GPA. Unfortunately, we only have GPAs for those students who were admitted to the program so we can only calculate the correlation between the test scores and GPA of those students who are admitted. We need to know that the test is valid for applicants, not just those who are admitted. Therefore, the only correlation we can calculate from the available data represents the wrong population—admitted students rather than applicants. As we will see, we can correct the correlation to better estimate the correlation in the population of admitted students.

The effects of selection and the appropriate method to correct it depend upon the form of selection that took place. Selection can be explicit or incidental. Keeping with our admissions example, if we review a pool of applicants' SAT scores and select which applicants to admit based on those scores, then we have explicitly selected the students on the SAT scores (Sackett & Yang, 2000). By selecting only those students with a minimum score on the SAT, we systematically exclude all students with SAT scores below the minimum score. Later, when we observe their GPA's we will only observe this range of SAT scores, which is smaller than the range in the applicant pool. This restriction on the range of SAT scores is direct range restriction (DRR) when it is induced by explicit selection.

Let X and Y be two random variables and ρ_{XY} their population correlation. Figure 2 illustrates the effect of explicit selection on $\hat{\rho}_{XY}$ when selection occurs on the basis of X. In this example the true population correlation, ρ_{XY}, is .70. Without selection (i.e., observing the entire random sample) the correlation of a random sample ($N = 500$) from the population is an accurate estimate, $\hat{\rho}_{XY} = .70$. After selecting scores greater than or equal to the 1st quartile of X, however, $\hat{\rho}_{XY}$ is reduced to .57, a 19% decrease. Selection of X at the 2nd and 3rd quartiles produces sample correlations of .48 and .38 respectively, a 31% and 45% reduction from the unrestricted $\hat{\rho}_{XY}$. Thus,

for constant ρ_{XY}, smaller selection ratios result in greater truncation and greater negative bias in $\hat{\rho}_{XY}$.

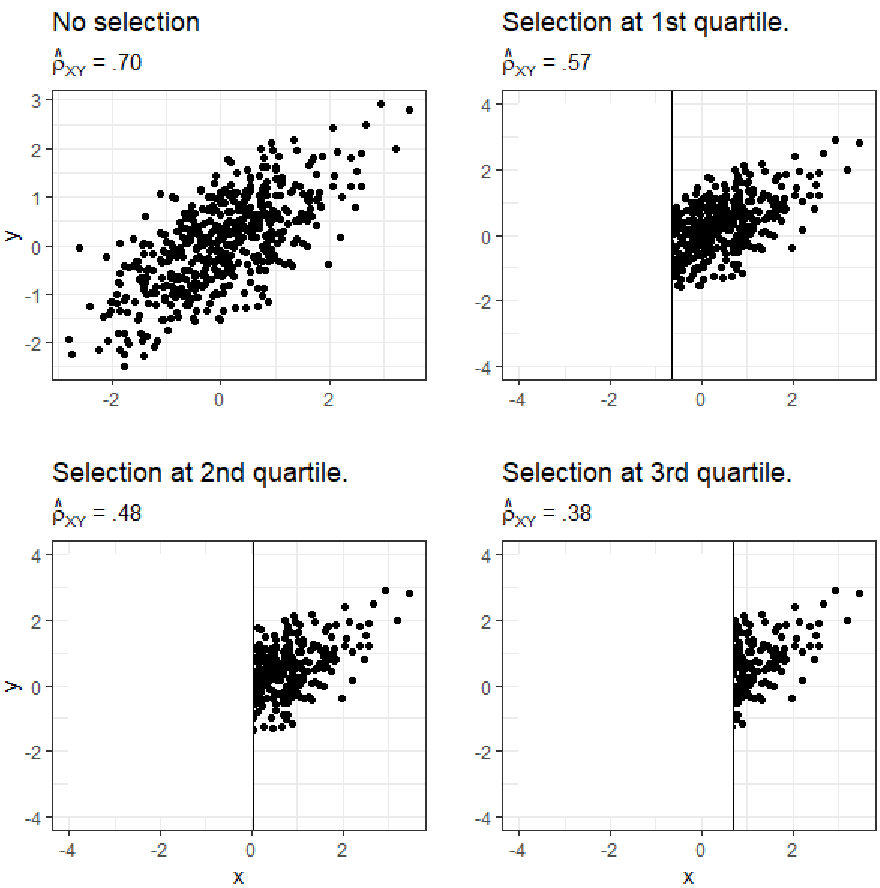

Figure 2. The effects of explicit selection on X with $\rho_{XY} = .70$.

Now, let us assume that we are also interested in how SES is related to the admissions selection process. When we explicitly select students based on their SAT scores, the observed values on SES in our sample of admitted students will depend on the SAT scores of those individuals we selected to the degree that SAT and SES are correlated. In this scenario, SES has undergone incidental selection and its values have undergone indirect range restriction (IRR) (Sackett & Yang, 2000). SES was not used to make a decision about the applicants, but

observations on SES are still affected by the explicit selection on the SAT scores because SES is correlated with SAT score.

Let X, Y, and Z, be three random variables. We are interested in estimating ρ_{XY}, but Z has undergone selection. To the degree that Z is correlated with X and Y, they undergo incidental selection, attenuating $\hat{\rho}_{XY}$ (M. J. Allen & Yen, 1979). Figure 3 demonstrates the effect of incidental selection through explicit selection on Z ($N = 500$) when $\rho_{XY} = .70$ and $\rho_{ZX} = \rho_{ZY} = .50$. The effects on $\hat{\rho}_{XY}$ are less extreme than explicitly selecting on X. After selecting only scores of Z greater than or equal to the first quartile, $\hat{\rho}_{XY}$ is reduced to .68, a 3% decrease. At both the 2nd and 3rd quartiles, the correlation is .61, a 13% decrease. Rather than completely truncate the values on X or Y, incidental selection through Z has more of a thinning effect on their distribution.

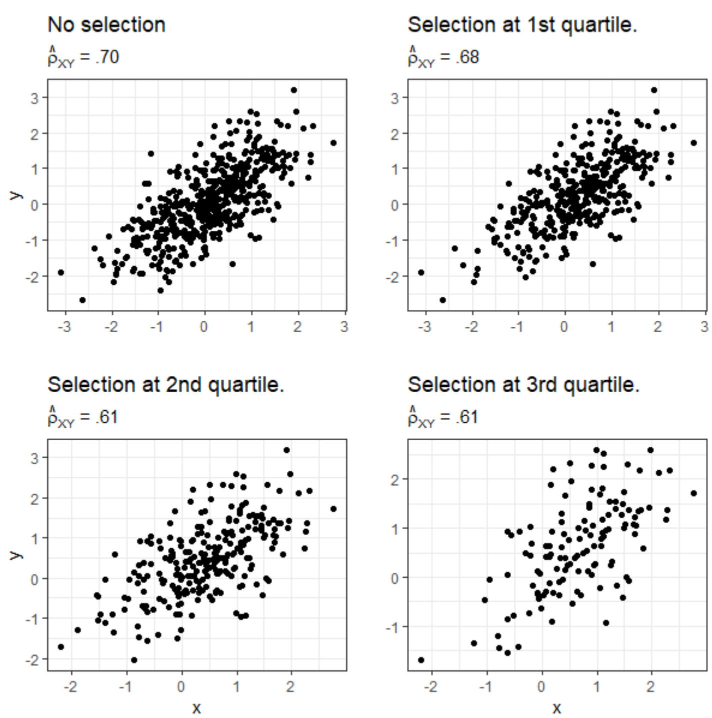

Figure 3. The effects of incidental selection on Z with $\rho_{XY} = .70$ and $\rho_{ZX} = \rho_{ZY} = .50$.

It is very clear that selection effects can dramatically impact correlation estimates.

Fortunately, we can correct the correlations observed in the restricted sample to better estimate the unrestricted target population value. In a psychometric meta-analysis framework each correlation estimated from a range restricted sample would be corrected individually using one of several available formulas. In a multivariate meta-analysis setting, however, studies contribute more than one effect size to the analysis. In this setting, selection effects have to be considered for all of the correlations within each study. Recall that in the SAT, SES, and GPA example explicit selection on the SAT score induced DRR on the correlation between SAT score and GPA, but also induced IRR on the correlation between SES and GPA. It is easy to show that selection on only SAT impacts all of the correlations. Figure 4 demonstrates the effects that explicit selection of the SAT scores has on all three correlations. The data was randomly generated with a mean of zero and standard deviation of one. The labels are purely pedagogical. I randomly generated 500 observations with arbitrary population correlations $\rho_{SAT,SES} = .50$, $\rho_{SAT,GPA} = .70$, and $\rho_{SES,GPA} = .40$. The first column of the figure contains plots of the full, unrestricted sample. The second column contains plots after explicit selection only on the SAT score. As can be seen, all of the relationships were attenuated by the selection on SAT.

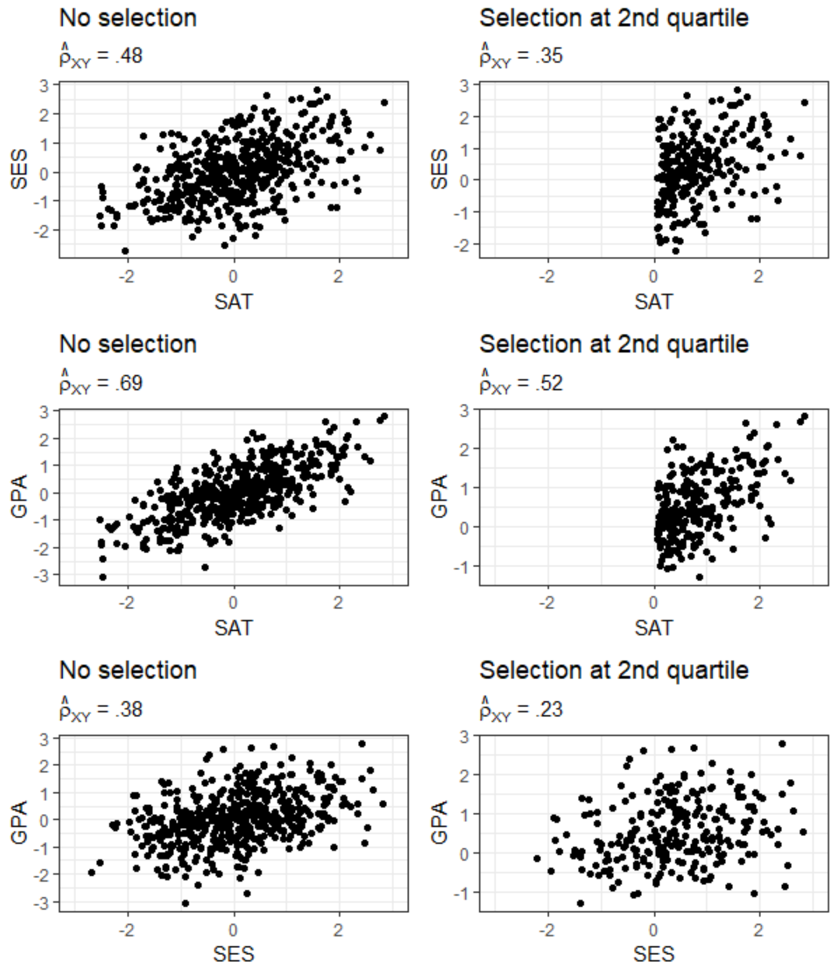

Figure 4. The effects of explicit selection on SAT score on all correlations from the same sample with $\rho_{SAT,SES} = .50$, $\rho_{SAT,GPA} = .70$, and $\rho_{SES,GPA} = .40$

A more formal examination of the formula for a correlation coefficient will provide a bit more insight into the exact impact of selection processes on observed correlations. Let X and Y be two random variables and ρ_{XY} their population correlation. Now let X^*, Y^*, and ρ^*_{XY} be the corresponding values in a subsample created from explicit selection on X. Limiting the observed values of X decreases the observed variance in X^* compared to X. Because $\sigma^{*2}_X \neq \sigma^2_X$ then $\rho^*_{XY} \neq \rho_{XY}$ since $\rho^*_{XY} = \frac{cov(XY)^*}{\sigma^*_X \sigma^*_Y} \neq \rho_{XY} = \frac{cov(XY)}{\sigma_X \sigma_Y}$. It is clear, then, that the magnitude and direction of the difference between the two correlations depend upon the changes in the covariance

relative to the change in the product of the standard deviations (Sackett & Yang, 2000). Typically, selection results in a smaller variance in X^* compared to X and a corresponding reduction in ρ^*_{XY} compared to ρ_{XY} as I have already demonstrated (Sackett & Yang, 2000). This change on X^* relative to X is DRR. However, range enhancement occurs if the extreme scores on X are selected excluding the middle scores, which causes the variance in X^* to become larger than in X corresponding to an increase in ρ^*_{XY} compared to ρ_{XY}. Because we are more likely in practice to observe restriction than enhancement (Schmidt & Hunter, 2015), I set aside explicit discussions of enhancement, but it should be noted that the corrections and considerations I discuss are just as applicable to that case (Sackett & Yang, 2000; Schmidt & Hunter, 2015). It follows that the results of this study are also applicable to those cases.

As selection ratios decrease the bias in the correlation estimates increases. In the case of IRR, as the correlation between Z and X or Y increases, so too does the effect of selection on Z given a constant selection ratio. I have shown that when multiple correlations are calculated on the same sample, selection on any one variable will attenuate all of the correlations. Under either explicit or incidental selection, $\rho^*_{XY} \neq \rho_{XY}$. The problem we face is that our goal is to estimate ρ_{XY}, but we can only compute $\hat{\rho}^*_{XY}$, which we know will be a biased estimate of ρ_{XY}. The purpose of the corrections for selection effects is to mitigate this bias and target the correct population. I now turn to those corrections and their theoretical limitations in the multivariate synthesis context.

Correcting the effects of selection. Correcting individual studies for selection effects is the first stage of a psychometric meta-analysis.Under Hunter and Schmidt's framework, the goal of artifact correction is to provide a parameter estimate as if all of the studies in the meta-analysis sample had been conducted perfectly—untainted by measurement error, range restriction, and other purely statistical artifacts. Because, they argue, the purpose of research is ultimately to discover theoretical relationships, the corrected effects are more appropriate for use in model testing (Schmidt & Hunter, 2015). We should acknowledge that correcting artifacts is not always

appropriate and there are those who question the utility of theoretically perfect correlations for practice (Crocker & Algina, 2008; LeBreton, Scherer, & James, 2014; Sackett & Yang, 2000). I do not take up this debate here. I do, however, offer an argument to justify the application of corrections for range restriction to correlations: that corrected correlations permit correct inferences based on a restricted sample to target the unrestricted population. These corrections improve the statistical conclusion and external validity of the correlation estimates (Shadish, Cook, & Campbell, 2002). That is reason enough to apply them and provides motivation for incorporating them into the MDM framework where they have been ignored.

Furthermore, failing to correct for range restriction in the meta-analysis context introduces variability among the observed correlations resulting from variability in the artifact. This additional variability can be mistaken for evidence of substantive moderators on the effects (Schmidt & Hunter, 2015). Because the investigation of moderators on the observed effects is a keystone of meta-analysis (Borenstein et al., 2009), the variance in effects caused by the presence of artifacts is detrimental to the goal of the analysis. I showed earlier that this was the case in the study by Lawson et al. (2017). To relate the problem to the admissions examples I have presented, consider that admissions criteria vary across institutions with some being more selective than others. Assume the simple case of admission to an undergraduate program being contingent only on an SAT score. Selection on the SAT will reduce the observed correlation between SAT score and undergraduate GPA, but the ratio of selection will not be constant at all institutions. I showed earlier that smaller selection ratios introduce more bias. More selective institutions will have smaller selection ratios that results in a greater reductions in the observed correlations. The variability resulting from differences in selection ratios and, hence, range restriction could be mistaken for the presence of a substantive moderator such as average SES of the applicant pool. The correlation between the SAT and achievement at more selective institutions will be more attenuated than that calculated at less selective institutions. This variance in the correlations is purely artificial. Left uncorrected, it would appear that there is greater variation in the correlations

than actually exists. Schmidt and Hunter (1977) demonstrated that this phenomenon resulted in decades of erroneous conclusions about validity coefficients in personnel research. Clearly, then, corrections for range restriction are appropriate so that we target the correct population parameters in our estimates and remove artificial variability from the sample of observed effects.

Schmidt and Hunter (2015) provide a full and thorough treatment of how to correct correlations for both explicit and incidental selection. They employ a multiplier to correct correlations for various artifacts. I do not present their methodology in full and, instead, focus on the traditional correction formulas. The multipliers of psychometric meta-analysis are transformations of these formulas and both methods produce the same corrections so there is nothing lost by this approach. Rather, it is simply convenient to refer to the more common formulas to more readily discuss the broader literature on the corrections. I also limit my discussion here to the substantive issues. I reserve technical presentations of the exact correction formulas for Appendices A and B.

Turning to the corrections themselves, Pearson (1903) first derived a correction for range restricted samples in response to his observations that natural selection processes modified the correlations between the characteristics of two organs. Thorndike (1949) later discussed such corrections in the context of personnel research. They are now commonly referred to as Thorndike's Cases I, II, that correct for DRR and III that corrects for IRR. Thorndike's cases are all univariate corrections—adjusting one correlation at a time. Building upon the work of Pearson and Aitken (1935), Lawley (1943) derived a generalized multivariate correction that adjusts for multiple selection variables and that can accommodate any of the Thorndike cases by simultaneously correcting a correlation matrix (Ree, Carretta, Earles, & Albert, 1994; Sackett & Yang, 2000).

Recall that when studies contribute more than one correlation to a meta-analysis, we are actually extracting a vector of correlations rather than a single correlation. These correlations can be arranged into a correlation matrix. Thus, we are extracting matrices and not single

correlations. The univariate procedures presented by Schmidt and Hunter (2015) based on Thorndike's formulas, require corrections to the matrix element-wise—one correlation at a time. Element-wise corrections are not necessarily equivalent to simultaneous corrections of the entire matrix (Ree et al., 1994). One of the goals of my research is to examine the degree of departure between element-wise and simultaneous corrections. Because psychometric meta-analysis was developed in a univariate context, it is not equipped to handle dependent effects arising from a single sample. This is reflected not only in the fact that there is no consideration of multivariate synthesis models, but in how psychometric meta-analysts apply artifact corrections—independently to correlations that, in truth, are related (Steiger, 1980). I have shown that selection on a single variable affects the entire set of correlations (i.e., the matrix) when the variables arise from the same sample. Element-wise corrections are theoretically problematic because they ignore this fact.

Ignoring the fact that correlations calculated on the same sample are related may invalidate the matrix. In particular, it might create triangular inequalities within the matrix or make it non-positive definite. Triangular inequalities occur when, given two correlations, the third falls outside the range of possible values. Stanley and Wang (1969) provided Equation 1 for evaluating this range.

$$r_{12}r_{13} \pm \sqrt{(1 - r_{12}^2)(1 - r_{13}^2)} \tag{1}$$

Given two correlations, r_{12} and r_{13}, the possible values of r_{23} are determined. Figure 5 shows the boundaries of the possible values of r_{23} over values of r_{13} given constant r_{12}. The possible values of r_{23} lie within each ellipse. When r_{12} is low, the range tends to be wide for any value of the second correlation. In fact, when $r_{12} = r_{13} = 0$, r_{23} can be any value between -1 and 1. When $r_{12} = 1$, $r_{13} = r_{23}$. This makes sense since if $r_{12} = 1$ then $X_1 = X_2$ and so their

correlations with X_3 must be the same. The likelihood that out of range values will occur, then, depends on the size of the larger of the two known correlations[2]. As either correlation approaches 1, the range of possible values of r_{23} narrows. Recalling that range restriction corrections will always increase the size of correlations, it is clear that the range of possible values will always shrink after correction. Applying the corrections element-wise would then apparently increase the risk of a resulting matrix containing an out-of-range value. The risk may be compounded if all of the correlations in a study are not corrected as was the case in Kuncel et al. (2001) who left all relationships that excluded the GRE uncorrected and, thus, attenuated. Studying these cases through simulation will help to quantify the risks of this taking place.

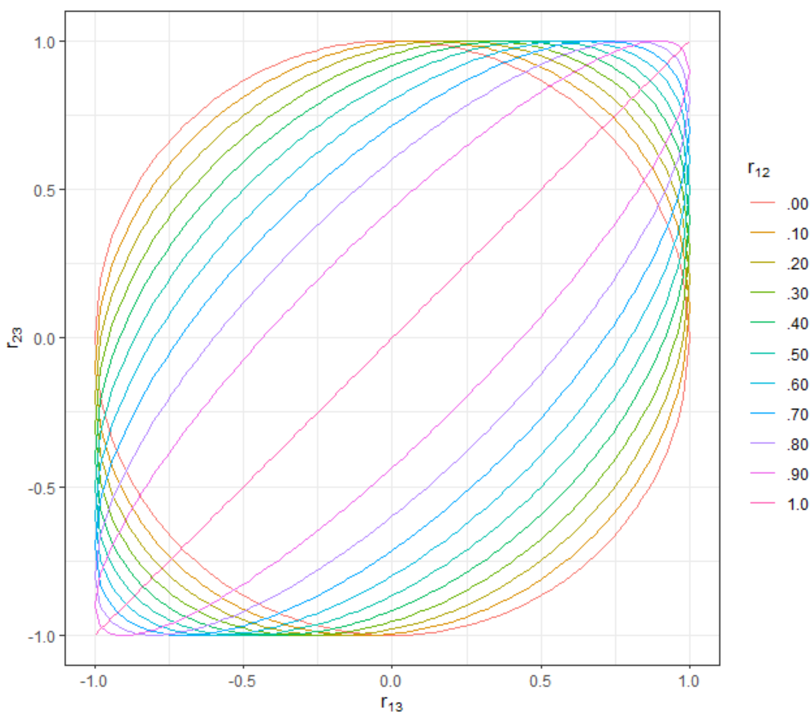

Figure 5. Boundaries of r_{23} over values of r_{13} given constant r_{12}.

A 3×3 correlation matrix can be analyzed for triangular inequalities using Equation 1 and the three-variable case is useful for illustrating the concept of the boundaries imposed within

[2]Reference to r_{12} is arbitrary. The results for constant r_{13} would be the same

the matrix. When the matrix is larger, however, a more general form to evaluate the validity of the matrix is necessary (Wothke, 1993). Hubert (1972) provided a generalization of Stanley and Wang's equation that evaluates each submatrix of a correlation matrix. The inequality to be tested for a set of standarized variables is given by Equation 2 where, for p variables, $\mathbf{G}$ is a $p \times 1$ column vector of covariances of one of the variables X_n with all others, and $\mathbf{A}$ is a $p-1 \times p-1$ covariance matrix of all variables excluding X_n. So for a set of four variables, we would first test X_1. $\mathbf{G}$ would be a column vector of covariances of X_1 with $X_2 ... X_p$ and $\mathbf{A}$ a covariance matrix of covariances of $X_2 ... X_p$ (excluding X_1). This repeats for each variable in turn until the entire matrix is evaluated.

$$\mathbf{G}'\mathbf{A}^{-1}\mathbf{G} \leq 1 \tag{2}$$

By definition, all correlation matrices must be positive definite. The potential for non-positive definite matrices was highlighted by Viswesvaran and Ones (1995) when they proposed their method of gathering correlations independently meta-analyzed on different samples of studies and assembling them into a matrix in an ad hoc fashion. Ad hoc assembly is analogous to univariate correction procedures because there is no regard for the interrelationships of the correlations from the same study, impacted by the same selection processes. As such, analogous concerns arise, but it is unclear what the risk is at the correction stage. Landis (2013) pointed out that these concerns have not been examined empirically or through a simulation and have simply been assumed to be minimal. My goal is to address this question in the first stage of this research.

While univariate corrections pose several potential problems, a multivariate range restriction correction that simultaneously corrects the entire matrix would mitigate them. Building on prior work by Pearson (1903) and Aitken (1935), Lawley (1943) derived a multivariate correction for range restriction. The Lawley correction permits the simultaneous correction of a correlation matrix—technically, a covariance matrix that can be converted to a

correlation matrix—overcoming the weaknesses of the univariate corrections. An advantage of the multivariate correction is that we do not know how selection occurred in order to apply it (Birnbaum, Paulson, & Andrews, 1950). We only need to know the variables on which selection occurred (Sackett & Yang, 2000). This avoids the problem of applying the wrong univariate formula to a correlation, which can have very serious consequences for the accuracy of the correction (Hunter, Schmidt, & Le, 2006; Sackett, Lievens, Berry, & Landers, 2007).

While this procedure avoids the apparent problems of the univariate corrections by correcting all elements simultaneously, it is more computationally burdensome requiring operations on matrices. With the growth of open source computing (e.g., R) it only takes the development of a simple function or package to overcome this limitation. Lawley's correction is available in *psychmeta* (Dahlke & Wiernik, 2018), a package for R. A more serious obstacle is that implementing it requires that one knows the variance-covariance matrix of all the variables involved. Until more researchers report full correlation (or, better yet, covariance) matrices along with their studies, this problem will persist. Other authors have cited the informational requirements of Lawely's correction in past methodological literature as a reason not to give it full consideration (Hunter et al., 2006; Schmidt, Oh, & Le, 2006). Lawley (1943) is not cited at all in Schmidt and Hunter (2015), which is the primary psychometric meta-analysis methodological text. Again, this is evidence of its wholly univariate focus.

While it is reasonable to limit methodological studies based on practical considerations, the lack of primary study data is also the most fundamental problem of the entire meta-analytic enterprise. More to the point, there are contexts in which Lawley's correction has been applied in meta-analytic studies as in the validation of the GRE (Klieger et al., 2014; Powers, 2004) or the ACT, high school grades, and SES (Westrick et al., 2015) to predict undergraduate achievement. Furthermore, because the IPD meta-analysis literature uses primary study data, the full covariance matrix would be available to the synthesist and corrections may be applicable in that context. In fact, Westrick et al. (2015) had raw data available and only used meta-analysis because they could

not identify a way to perform an HGLM analysis with the data corrected for range restriction. In such cases, the application of Lawley's correction may provide for better estimates. This is a procedure that has not been given full consideration in the psychometric meta-analysis literature but has reason to be fully evaluated along with other methods both because of the theoretical problems with univariate corrections and the fact that it is clearly applied in practice.

I have now raised three major concerns with the application of univariate range restriction corrections to correlation matrices. Compared to Lawley's multivariate correction, it is unclear whether they are sufficiently accurate, how frequently they produce triangular inequalities, and how frequently they produce non-positive definite corrected matrices. These comprise the first set of research foci for this study.

The application of range restriction corrections to correlation matrices has gone unexplored because psychometric meta-analysis is a wholly univariate methodology. In the sections that follow, I provide an overview of the multivariate meta-analysis model and discuss the challenges that follow from applying artifact corrections in this setting.

Model-based Meta-analysis

The motivation behind model-based meta-analysis is to synthesize a correlation matrix that can be used as an input matrix for a path analysis (Becker, 2009). The analysis takes place in two stages, first pooling correlation matrices from a series of studies and then using the resulting summary matrix as the basis for testing a path model. In this research, I focus on methods for pooling the matrices. Before I introduce methods of pooling, I introduce the model that underlies meta-analysis, reference to which will serve to highlight points of departure between generalized and psychometric meta-analysis methods.

All methods of meta-analysis treat effect sizes as outcomes. The random-effects meta-analysis model is based on the assumption that studies themselves are random samples from

a universe of possible studies. It follows that their effects are also randomly sampled from a distribution of effect sizes. One way to conceptualize this is to imagine sampling study designs from the universe of possible designs where *design* encompasses any combination of study characteristics (Raudenbush, 2009). Underlying each such design is a vector of true population effects and, so, these effects too are essentially sampled from a distribution of effect sizes. In the next step, random samples of observations for each study are taken. The effects observed within each study will themselves randomly deviate from the true study effects underlying it due to the sampling of observations within the study. By assuming a random-effects model, we can make inferences to the universe of possible studies.

An alternative to the random-effects model is the fixed-effects model. Under the fixed-effects model, we assume that a common vector of effects underlie all of the studies. Compared to the random-effects model, there is no random sampling from the universe of study—these studies *are* the universe. The sampling of observations within each study is the only source of error, or deviation, from this common vector of effects. The fixed-effects model permits inference only to those studies represented by the sample.

It is widely recognized that random-effects models are, *a priori*, typically more realistic models than fixed-effects models (Hedges & Vevea, 1998; Schmidt & Hunter, 2015). In fact, psychometric meta-analysis only ever considers the random-effects model, with Schmidt and Hunter (2015) pretty well rejecting the feasibility of a fixed-effects model. That random-effects models are more reasonable is a reflection of the larger goal of meta-analysis and the state of most bodies of research literature. Often, we want to make inferences to any study measuring the effects in which we are interested, not only to those alike the sampled studies. In addition, fixed-effects may be appropriate if the studies in a meta-analysis sample are close replications such that only the random sampling within the studies is expected to create variation in the observed effects. Given the fact that direct replication is not common and that meta-analytic studies often include studies conducted under diverse conditions, justification of a fixed-effects

model seems like it would be rare. Furthermore, as Hedges and Vevea (1998) point out, the problem in making a choice between the fixed-effects and random-effects model is determining how closely replicated the studies have to be to consider them close enough to be a replication.

Given the fact that the random-effects model is typically more appropriate, and the fact that it subsumes the fixed-effects model, I present it here. I focus on the multivariate notation of the model, which reduces to the univariate when each study in a sample contributes only a single effect size. Although my focus in this study is on Pearson's r, I adopt the more general notation found in the generalized meta-analysis literature to emphasize that this model applies to any effect size (e.g. Borenstein et al., 2009; Cooper et al., 2009).

Let $\boldsymbol{T}_i$ be a vector of observed effects in study i from a sample of k studies. The vector of observed effects in the studies varies about a vector of mean true study effects, $\boldsymbol{\theta}_i$. The true study effects, in turn, vary about a vector of population effects $\boldsymbol{\theta}$. Assume a p-variate normal distribution such that $\mathbf{T}_i \sim N_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{\theta_i} \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Psi})$ where $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Psi}$ are the $p \times p$ covariance matrices of the observed (within-study) and true study (between-study) effects respectively. We can represent this as

$$\mathbf{T}_i = \boldsymbol{\theta}_i + \mathbf{e}_i \tag{3a}$$

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \mathbf{u}_i \tag{3b}$$

$$\mathbf{T}_i = \boldsymbol{\theta} + \mathbf{u}_i + \mathbf{e}_i \tag{3c}$$

where $\mathbf{e}_i \sim N_p(0, \boldsymbol{\Sigma}_i)$ and $\mathbf{u}_i \sim N_p(0, \boldsymbol{\Psi})$ and all error terms are assumed to be independent.

The between-study variance in effects is given by $\boldsymbol{\Psi}$ and is captured in the model by $\mathbf{u}_i$ as the deviation of the true effects underlying study i, $\boldsymbol{\theta}_i$, from the true effects, $\boldsymbol{\theta}$. The within-study variance is given by $\boldsymbol{\Sigma}_i$ and is captured in the model by $\mathbf{e}_i$ as the deviation of the observed effects, $\mathbf{T}_i$, from the true study effects underlying study i, $\boldsymbol{\theta}_i$. If we set the off-diagonal elements of $\boldsymbol{\Sigma}_i$ and

$\boldsymbol{\Psi}$ to zero we have the equivalent of a series of univariate random-effects meta-analyses. When there is no hypothesized between-study variance and all elements of $\boldsymbol{\Psi}$ are zero, the random-effects model reduces to a fixed-effects model.

The goal of a random-effects model is to estimate the population effect, $\boldsymbol{\theta}$, and its variance $\boldsymbol{\Psi}$ (Wei & Higgins, 2012). While they both share a common underlying model, we can distinguish between the generalized meta-analysis and psychometric met-analysis approaches at a fundamental level by how the models are specified within the methodological literature. Consider first the estimate of $\boldsymbol{\theta}$. In psychometric meta-analyses, the model is fit with $\mathbf{T}_i^c$—a vector of corrected effect sizes, usually correlations. If, however, we strictly follow the generalized meta-analysis procedures as they are typically outlined, we would fit the uncorrected $\mathbf{T}_i$. The bias in $\hat{\boldsymbol{\theta}}$ is left as unexplained error variance. Subsequent moderator analyses can be conducted to test for the presence of artifacts by regressing the effects on one more predictors, which may include continuous or categorical indicators of the bias (Borenstein et al., 2009). In the end, however, the generalized meta-analysis framework provides no corrected effect estimate—the summary effect is estimated on the attenuated effects. Importantly, then, psychometric meta-analysis methods provide corrected estimates of $\boldsymbol{\theta}$ while generalized meta-analysis methods indicate that the presence of statistical artifacts explains the variation in observed effects. Although one could adjust effect estimates based on the coefficient of the moderator, moderator analyses typically have low power (Hedges & Pigott, 2001) so are prone to Type II errors, which would lead investigators to conclude there was no biasing effect. Schmidt and Hunter (2015) also discuss these problems with using moderator analyses instead of directly correcting the correlations.

Turning now to $\boldsymbol{\Psi}$, the psychometric meta-analysis literature only ever considers random-effects models which include estimates of $\boldsymbol{\Psi}$, but they are always univariate with off-diagonal elements set to zero. Off-diagonal elements of $\boldsymbol{\Sigma}_i$ are also always assumed to be zero. As a result, the total variance in effects is always underestimated in psychometric meta-analyses when correlated effects from the same study are synthesized. By contrast,

generalized meta-analysis methods include extensive treatments of cases where the off-diagonal elements of $\mathbf{\Sigma}_i$ and $\mathbf{\Psi}$ are non-zero and caution against, but do not outright discourage, fixed-effects models where all elements of $\mathbf{\Psi}$ are zero.

Table 1

Contrasts in the Generalized and Psychometric Meta-analysis Model

Vector/Matrix	Framework	
	Generalized	Psychometric
$\hat{\boldsymbol{\theta}}$	$\hat{\theta}$	$\hat{\theta}^c$
$\mathbf{\Sigma}_i$	$\begin{bmatrix} \sigma_{i1}^2 & \rho_{i12}\sigma_{i1}\sigma_{i2} & \dots & \rho_{i1p}\sigma_{i1}\sigma_{ip} \\ \rho_{i21}\sigma_{i2}\sigma_{i1} & \sigma_{i2}^2 & \dots & \rho_{i2p}\sigma_{i2}\sigma_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{ip1}\sigma_{ip}\sigma_{i1} & \rho_{i2p}\sigma_{ip}\sigma_{i2} & \dots & \sigma_{ip}^2 \end{bmatrix}$	$\begin{bmatrix} \sigma_{i1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{i2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{ip}^2 \end{bmatrix}$
$\mathbf{\Psi}$	$\begin{bmatrix} \tau_{11}^2 & \rho_{12}\tau_{11}\tau_{22} & \dots & \rho_{1p}\tau_{11}\tau_{pp} \\ \rho_{21}\tau_{22}\tau_{11} & \tau_{22}^2 & \dots & \rho_{2p}\tau_{22}\tau_{pp} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1}\tau_{pp}\tau_{11} & \rho_{2p}\tau_{pp}\tau_{22} & \dots & \tau_{pp}^2 \end{bmatrix}$	$\begin{bmatrix} \tau_{11}^2 & 0 & \dots & 0 \\ 0 & \tau_{22}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tau_{pp}^2 \end{bmatrix}$

As summarized in Table 1, psychometric meta-analysis methods theoretically provide better estimates of the mean effects if statistical artifacts bias observed effects, while the generalized met-analysis methods permit more flexibility in the specification of the effect covariances and explicit modeling of dependency among the observed effects. The most flexible method would be to adopt the psychometric meta-analysis approach to artifact correction, while also providing the flexibility of the generalized meta-analysis approach for specifying the covariance matrices.

Correlated effects in meta-analyses. Correlated effect sizes arise commonly in meta-analyses. Ahn, Ames, and Myers (2012) reviewed 56 meta-analyses in education published

in leading (AERA) journals between 2000 and 2010. Of those reviewed, 28 (62%) reported dependent effects in their samples. Treating dependent effects as if they were independent results in erroneous estimates of the total variance in the effects. When effects are independent, their covariances are zero. Thus, the total variance in the effects can be characterized by the sum of their variances. If, however, the effects are not independent, their covariances are non-zero. The additional non-zero values increase the total variance in the effects. Thus, ignoring these covariances results in an underestimation of total variance. When we underestimate these variances, we falsely believe that our estimates are more precise than they are. This, in turn, increases the likelihood of Type I errors (Raudenbush et al., 1988). As a result, it is theoretically important to model the dependency among effect sizes arising from the same sample, at the very least when we are interested in their values and significance testing—both of which are concerns in meta-analysis.

The key to modeling the dependence within each study to is to properly estimate $\mathbf{\Sigma}_i$. In particular, it is the off-diagonal elements that are assumed or known to be non-zero. Recall from Table 1 that in psychometric meta-analysis these elements are always set to zero so correlated effects are simply not modeled. Schmidt and Hunter (2015) acknowledge that correlated effects underestimate the true variance in the effects. In psychometric meta-analysis, hypothesis tests are not used, but the error variances are subtracted from the total variance in effects leaving an estimate of the true variance in the effects without sampling error. Underestimates of the error variances lead to overestimates of the true variance. Thus, it is not that they were not aware of the problem nor that their procedures are immune to its effects. Instead, they only worry about correlated effects if they are considered to be conceptual replications because when they are not, they simply conduct a separate meta-analysis.

Conceptual replication is best explained with an example. In some cases, effects reported in the same study measure the same relationship. In other cases, the effects measure different underlying relationships. Consider, for example, a meta-analysis in which we are interested in the

correlations between ACT and undergraduate GPA (UGPA), SAT and UGPA, and SES and UGPA. Suppose that we find one more studies that report all three relationships among these variables measured on the same sample. The relationships of UGPA with the ACT and SAT are conceptually similar in that they both represent the relationship between cognitive ability as measured by the exams and academic achievement as measured by UGPA. On the other hand, SES is not a measure of cognitive ability. Borrowing Schmidt and Hunter's terminology, ACT-UGPA and SAT-UGPA are *conceptual replications* of the same relationship—cognitive ability and student achievement. On the other hand, we cannot consider the SES-UGPA relationship to be a conceptual replication of such a measurement because SES does not measure cognitive ability. In studies where the ACT-UGPA and SAT-UGPA measurement are both reported, cognitive ability is measured twice on each subject within the study resulting in conceptual replication within the study.

Hunter and Schmidt specifically recommend that when two variables are not conceptually related they be synthesized separately even if they come from the same study (Schmidt & Hunter, 2015). In arguing against treating conceptual replications within the same study as independent they point out the problems with variance estimation that I presented earlier. They do not, however, address this problem at all for conceptually distinct effects from the same sample. Yet, the correlations between such effects still have a bearing on the total error variance. To the degree that a correlation covaries with another conceptually distinct correlation, its variance is partially shared with it. Ignoring it underestimates the total variance for that correlation.

When conceptual replications exist within a study, Schmidt and Hunter (2015) recommend creating a composite of the variables and computing the correlation between the composite and outcome. That correlation is then used in the meta-analysis and each study then provides only one effect making univariate models appropriate. The question arises, then why we might want to explicitly model the correlated effects. The most obvious situation is when there are correlated effects that are not conceptual replications. Even when conceptual replications within a study exist, there are times we might not want to create a composite.

Aside from proper variance estimates, there are other advantages to modeling the dependency among effects. It is very common in meta-analyses that the sampled studies report different sets of effects (Becker, 2009). In other words, it is rare for all of the studies to examine all of the same relationships. As a result, some studies will be missing information about some of the relationships. In a univariate meta-analysis, those studies must be excluded from the analysis. In a multivariate meta-analysis, however, as long as a study includes at least one effect of interest it can be included. When the studies are analyzed in a multivariate context, the information across all studies can be used to provide better estimates of all of the relationships (Gleser & Olkin, 2009; Wei & Higgins, 2012). Consider an analysis with five studies and six relationships between the ACT, high school GPA (HGPA), undergraduate GPA (UGPA), and SES. Table 2 contains fictitious data on the correlations among the four variables reported in five studies. Only study 5 examined SES so it is the only source of information about the relationship of SES with the other variables. A univariate meta-analysis cannot be performed on the SES relationships since there is only a single study with information on the relationship. If, however, we perform a multivariate analysis of the six relationships simultaneously, the correlations among the correlations can be used to improve the estimates of any studies with missing observations, what some refer to as *borrowing strength* (Riley, 2009; Wei & Higgins, 2012).

To understand conceptually how this is possible, consider that the correlation between any two variables can be used to predict one given the other. In the fictitious study data, only one study reports the correlation with SES. By modeling the correlation of that effect with the other effects in Study 5, we can use that information to improve the estimate of SES correlations overall. Gleser and Olkin (2009) also discuss this advantage and provide an illustrative example.

If we analyze the fictitious study data in Table 2 using GLS estimation, we get a mean correlation vector $\bar{\mathbf{r}} = (.48, .52, .49, .53, .57, .49)'$, the last three elements of which represent ACT-SES, HGPA-SES, and UGPA-SES respectively. Their values are slightly different than those reported in study 5. Compared to their values in study 5, their variances are also slightly reduced

indicating greater precision in the estimate.

Table 2

Fictitious Study Data

Study ID	Correlation					
	ACT-HGPA	ACT-UGPA	HPGA-UGPA	ACT-SES	HGPA-SES	UGPA-SES
1	.45	.53	.50	-	-	-
2	.51	.54	.49	-	-	-
3	.42	.49	.47	-	-	-
4	.55	.51	.54	-	-	-
5	.44	.50	.43	.52	.55	.47

Note. Source: Author's compilation, not based on real data. $n_1 = n_2 = n_3 = n_4 = n_5 = 500$.

In addition to borrowing strength from all of the correlations together, one might wish to test for differences between the elements of the summary matrix. Olkin and Finn (1990) provide an example of the need to test the best predictor of three alternative measures of food intake based on their correlation with a more precise, but too costly, measure. This has clear implications in the development of selection procedures. Should admission decisions be made on the basis of multiple measures that include, for example, study skills inventories they might become overly burdensome to administer. Selecting the right inventory with desirable statistical properties would be of interest. The model provided by Olkin and Finn permit testing whether or not the measures equally predict an outcome while accounting for the correlations between them. The results of the test can be combined with substantive expertise to inform the best choice. This scenario actually demonstrates that modeling all of the conceptually related effects rather than creating a composite from them is sometimes advantageous.

Because psychometric meta-analyses are fully univariate, we cannot look to that literature for advice on applying its corrections in the multivariate case. On the other hand, the generalized meta-analysis literature contains numerous treatments, but without examining effects corrected for artifacts.

Pooling vectors of correlations. In psychometric meta-analysis, Viswesvaran and Ones (1995) simply recommend that the meta-analytic results from series of univariate meta-analyses be assembled into a single matrix rather than statistically pooling them. In a paper cited 1,120 times according to Google Scholar, Joseph and Newman (2010) integrated several strands of research by first proposing a model predicting job performance from emotional intelligence, conscientiousness, cognitive ability, and emotional stability. To test their model, they harvested the correlations from 21 prior meta-analyses. They also performed their own meta-analyses on some of the relationships of interest. These were assembled ad hoc into a single matrix and used to test their path model. In another example, Meriac, Hoffman, Woehr, and Fleisher (2008) sought to quantify the incremental validity of seven dimensions of assessment centers to predict job performance. They computed weighted average correlations for each relationship of interest and arranged these into a matrix. Both of these studies individually corrected each relationship for measurement error and range restriction.

I refer to this method as a simple assembly method. Using this approach, we gather the results of other meta-analyses and insert their summary correlation effects into a single matrix instead of statistically estimating a matrix. Viswesvaran and Ones (1995) acknowledge that this could lead to improper (non-positive definite) matrices and that without examining the underlying studies contributing each summary effect, the final matrix could contain estimates from different populations. Without replicating the analyses, it is difficult to determine how frequently these problems occur. In any case, I look to the generalized meta-analysis literature for more statistically justifiable methods.

It will be helpful at this point to express the meta-analysis model from Equation 3 more succinctly in matrix form and substitute the general notation with that for correlations. Let k be the number of studies and p_i the number of observed effects in study i such that $n = \sum_{i=1}^{k} p$. The model is $\mathbf{r} = \mathbf{X}\boldsymbol{\rho} + \boldsymbol{\epsilon}$ where $\mathbf{r}$ is an $n \times 1$ vector of observed correlations, $\mathbf{X}$ is an $n \times p$ design matrix, $\boldsymbol{\rho}$ is a $p \times 1$ vector of population correlations to be estimated, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of

errors.

When we wish to model the dependence among the $\mathbf{r}_i$ the estimate of $\boldsymbol{\rho}$ needs to be weighted by their covariances, $\boldsymbol{\Sigma}_i$. Becker (1992) proposed the use of a generalized least squares (GLS) estimator. GLS is a generalized form of ordinary least squares (OLS) regression that incorporates $\boldsymbol{\Sigma}_i$ to weight estimates and adjust standard errors to account for the fact that the error terms are not independent due to the correlation among the effects (Kaufman, 2014). In a GLS model, we put all of the individual estimates of $\boldsymbol{\Sigma}_i$ for each study, i, into a block diagonal matrix, $\hat{\boldsymbol{\Sigma}}$. The vector of summary correlations and their covariances are then given by

$$\hat{\boldsymbol{\rho}} = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{r} \tag{4}$$

$$Var(\hat{\boldsymbol{\rho}}) = (\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1} \tag{5}$$

In a typical meta-analysis without artifact corrections, we can estimate each $\boldsymbol{\Sigma}_i$ using closed form estimators. They appear to have been first derived by Pearson and Filon (1898), later by Elston (1975), and again by Olkin and Siotani (1976)[3]. The covariance of two correlations calculated on the same sample is a function of all the correlations among the variables referenced by their indices. As an example, assume we have continuous data on students' high school GPA, college admissions test score, study skills inventory score, and college GPA. For simplicity in the notation, let these be indexed by 1, 2, 3, and 4 respectively. Among the $p = 4$ variables there are $p^* = \frac{p(p-1)}{2} = 6$ unique correlations. The sampling covariance between any two of the six

[3]Perhaps given Olkin's contributions to meta-analysis, his and Siotani's work is typically cited in the meta-analysis literature (e.g. Becker, 2009; Hafdahl, 2008).

correlations is given by

$$Cov(r_{12}, r_{34}) = \frac{\begin{bmatrix} 0.5\rho_{12}\rho_{23}(\rho_{13}^2 + \rho_{14}^2 + \rho_{23}^2 + \rho_{34}^2) + \rho_{13}\rho_{24} + \rho_{14}\rho_{23} \\ - (\rho_{12}\rho_{13}\rho_{14} + \rho_{21}\rho_{23}\rho_{24} + \rho_{31}\rho_{32}\rho_{34} + \rho_{41}\rho_{42}\rho_{43}) \end{bmatrix}}{n} \tag{6}$$

when there is no common index (i.e. no common variable) and

$$Cov(r_{12}, r_{13}) = \frac{[0.5(2\rho_{23} - \rho_{12}\rho_{13})(1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2) + \rho_{23}^2]}{n} \tag{7}$$

when an index is shared among them and, finally,

$$Cov(r_{12}, r_{12}) = Var(r_{12}) = \frac{(1 - \rho_{12}^2)^2}{n} \tag{8}$$

when the same correlation is examined we obtain its variance. So, given the correlation between high school GPA and admissions test score (r_{12}) and the correlation between study skills inventory and college GPA (r_{34}) their covariance is estimated by Equation (6). The covariance between the correlations of high school GPA with admissions test score (r_{12}) and study skills (r_{13}) is given by Equation (7) and the variance of the correlation between high school grades and admissions test score is given by Equation (8).

The challenge is to estimate the covariance of correlations corrected for selection effects. If we wanted to apply a GLS method, we would need to estimate each $\mathbf{\Sigma}_i$. The standard errors that form its diagonal elements have been derived, the square of which would provide their variances (Bobko & Rieck, 1980; Raju & Brand, 2003; Schmidt & Hunter, 2015). We do not, however, have estimates of their covariances that form the off-diagonal elements of $\mathbf{\Sigma}_i$. We can

use the multivariate form of the delta method to derive the asymptotic distributions of scalar-valued functions and, thus, their means, variances, and covariances (Bishop, Fienberg, & Holland, 1975). In fact, we could use the delta method to reproduce the exact results in Equations (6) to (8) (Pustejovsky, 2018). In the context of this research, I am interested in modeling the covariance of correlations corrected for selection effects so the scalar-valued functions to use in the delta method would be the correction formulas in Appendix A. Bobko and Rieck (1980) used the delta method to derive the standard error of correlations corrected for direct range restriction, but did not include a solution for their covariances. Raju and Brand (2003) also used the delta method to determine the variance of correlations corrected for both range restriction and unreliability. More recently, Wei and Higgins (2012) derived a series of estimators for the covariances between effect size pairs commonly found in medical research including log odds ratios, standardized mean differences, log risk ratios, and risk differences. If this work were done for corrected correlations, we could use the closed form estimators to compute the corrected correlation covariances in order to specify $\mathbf{\Sigma}_i$.

Deriving exact estimators may be useful for some applications, but the enduring problem in meta-analyses is the lack of information available in primary studies (Borenstein et al., 2009). To compute the covariance of correlations, we are likely to need all of the intercorrelations among all the variables. That information will not always be available from primary studies. Even if it were available, the estimators would certainly require software implementation to be of any practical use for researchers who are not inclined to perform or program their own calculations.

In addition to the practical limitations of exact estimators, the delta method would only be applicable to the univariate corrections. The univariate functions to correct correlations return scalar values, but Lawley's multivariate correction returns a matrix. I have argued that Lawley's multivariate correction would be more appropriate. If we derived estimators from the univariate formulas, but applied Lawley's correction in practice, we would have to assume that the covariances do not change meaningfully between correction methods. We would need to show

that the correction to the covariance matrix is the same in the univariate and multivariate case although the corrections to the correlations themselves are not the same. To do this, we need estimates of their covariances, the absence of which is the problem at hand.

A better solution is to avoid the need to directly estimate the covariances but still model them. Theoretically, robust variance estimation (RVE) (Hedges et al., 2010) provides a method for doing just that. RVE is a method of multivariate meta-analysis that accounts for the within-study correlation of effects arising from the same sample without the need to specify any information about their within-study covariances. Instead, $\hat{\mathbf{\Sigma}}$ is estimated empirically from the data. The method works by substituting a weight matrix, $\mathbf{W}_i$, in place of the covariance matrix, $\mathbf{\Sigma}_i$. Estimation of the population effects and their variances is then a three-step process. First, the weight matrix is used to estimate the population effects. The residual error terms are then computed and their cross-products are used as an estimate of $\mathbf{\Sigma}$ to compute the variance of the effects. The estimators are given by Equations 9 and 10 where the vector of errors is $\mathbf{e} = \mathbf{r} - \mathbf{X}\hat{\boldsymbol{\rho}}$.

$$\hat{\boldsymbol{\rho}} = \left(\sum_{i=1}^{k} \mathbf{X}'\mathbf{W}^{-1}\mathbf{X} \right)^{-1} \left(\sum_{i=1}^{k} \mathbf{X}'\mathbf{W}^{-1}\mathbf{r} \right) \tag{9}$$

$$Var(\hat{\boldsymbol{\rho}}) = \left(\sum_{i=1}^{k} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{k} \mathbf{X}_i'\mathbf{W}_i\mathbf{e}_i\mathbf{e}_i'\mathbf{W}_i\mathbf{X}_i \right) \left(\sum_{i=1}^{k} \mathbf{X}_i'\mathbf{W}_i\mathbf{X}_i \right)^{-1} \tag{10}$$

Moving from a GLS to an RVE framework, we move from needing to estimate $\mathbf{\Sigma}_i$ to needing to specify $\mathbf{W}_i$. The asymptotic results of RVE are robust to the choice of weights in $\mathbf{W}_i$. However, Hedges et al. (2010) recommend setting the total weight of each study equal to the inverse of the average variance of the effects within each study, $\frac{1}{\bar{v}_i}$, because inverse variance weights provide the most efficient estimates. Efficiency in a statistic is a desirable property that

refers to its ability to provide consistent (i.e. less variable) estimates in different samples. Thus, each $\mathbf{W}_i$ is a diagonal matrix with $\frac{1}{p_i \bar{v}_{\cdot i}}$ as its elements. Because the standard errors of corrected correlations have been derived, we can use their squares to complete the weight matrix making RVE a viable alternative to GLS. Another advantage of the RVE method is that it does not make any assumptions about the distribution of effects (Hedges et al., 2010).

RVE has been shown to perform well in several studies. In their original article, Hedges et al. (2010) conducted a simulation study to examine the confidence intervals produced by RVE estimates. Over a variety of conditions the confidence intervals contained the true population value between 93% and 96% of the time. López-López, den Noortgate, Tanner-Smith, Wilson, and Lipsey (2017) compared the performance of RVE with a three-level hierarchical model on data with study effects correlated either .30 or .70 and average effects per study of four, six, or eight. They found that RVE generally controlled Type I error rates more effectively than the three-level model. Park and Beretvas (2018) also examined the performance of a three-level model and RVE. Only the RVE method converged 100% of the time under all conditions and otherwise performed well. The results of these studies indicate good performance of RVE compared with multilevel models that can also be used to estimate the covariance among dependent effects. All of the effect sizes in these studies were standardized mean differences.

Tanner-Smith et al. (2016) demonstrated the application of RVE to correlation coefficients on real data, but did not conduct a simulation study. Following generalized meta-analysis procedures, they converted the raw correlations to the Fisher's z metric. They also applied artifact corrections to the correlations and appear to have used the standard variance estimate for the Fisher's z metric, $(n-3)^{-1}$. I have only located one study that considers the standard error of a Fisher's z-converted correlation when it has been corrected for artifacts and it was shown that Type I error rates were inflated three to four times their nominal rates (Raju & Brand, 2003). Without a good estimator of the variance, we cannot properly specify a weight matrix with inverse variance weights. However, if $(n-3)^{-1}$ provides even an approximation of the standard

error that would be obtained for a corrected correlation it may operate well enough to be of practical use (Hedges et al., 2010). I set this issue aside for future research. Here, I instead focus on the synthesis of the raw correlations since we have methods to estimate their standard errors after they have been corrected.

In summary, while we cannot use a GLS approach to pooling the corrected correlations, we can employ the RVE method. Instead of estimating Σ_i, we can specify a weight matrix using the inverse variances of the corrected correlations. The inverse variances are available by squaring the standard errors given by Raju and Brand (2003) for corrections of explicit selection or N. L. Allen and Dunbar (1990) in the case of incidental selection. Schmidt and Hunter (2015) also provide approximations of the standard errors that are less computationally burdensome.

Summary and Hypotheses

To this point I have shown that both selection effects and correlated effects occur in research domains including education, psychology, and personnel selection. I have argued that the application of univariate corrections for selection effects to vectors or matrices of correlations that share a common sample introduces unique problems. I have also identified that among the challenges in synthesizing correlations corrected for selection from the same sample is estimating the within-study correlation among those correlations, but that RVE methods may provide a solution to this obstacle. I now summarize these arguments in a series of research questions and corresponding hypotheses.

Evaluating the correction procedures. One of the concerns in applying psychometric meta-analysis procedures to a multivariate context is that they entail correcting one correlation at a time rather than correcting all of the elements of the correlation matrix simultaneously. Most importantly, the degree to which the accuracy of the corrections is affected by such an approach is unclear. Ree et al. (1994) claimed that making univariate corrections to each element of a matrix is

not the same as applying Lawley's multivariate correction and is likely to lead to undercorrection of the correlations, but provide no citations. In addition, the univariate procedures introduce the possibility that the resulting correlation matrix will be invalid. I found no evidence of studies that rigorously compare the two correction procedures. A single study by Held and Foley (1994) compared univariate and multivariate corrections for selection based on military tests. They found that the multivariate correction outperformed the univariate. They did not, however, study a range of population correlations or sample sizes nor did they examine the results for invalid matrices.

Perhaps the best outcome will be to find that the univariate corrections perform well compared to the multivariate in a variety of conditions and that the risk for invalidating the resulting matrix is minimal. Such a result would provide evidence that the simpler implementation of correction formulas can be followed, although available software makes the application of Lawley's procedure straightforward. More importantly, Lawley's procedure requires more information to implement than do the univariate procedures. The variances and covariances of all of the variables are necessary to implement Lawley's correction. All meta-analyses are challenged by incomplete reporting in primary studies. Thus, confirming that univariate procedures operate sufficiently well would be a welcome result. On the other hand, all things being equal, the exact selection processes need not be known to apply Lawley's correction as long as one knows which variables were subject to selection effects. Misspecification of the selection process and application of the wrong procedures can produce seriously misleading results (Hunter et al., 2006; Ree et al., 1994; Sackett & Yang, 2000). Applying Lawley's correction would mitigate this problem.

Research Question 1: Bias in the correction methods. As indicated by absolute median bias, how well do univariate, element-wise corrections for selection recover the unrestricted correlations compared to a multivariate, simultaneous correction when they are both applied to a matrix of correlations? The univariate correction procedures were not originally derived to be

applied to the elements of a correlation matrix. Lawley's correction, however, permits simultaneous correction of all the elements of the matrix.

Hypothesis 1.1. Prior research suggests that there is no relationship between the accuracy of the corrections and the true population correlation ρ in the univariate case (Duan & Dunlap, 1997; Mendoza, Hart, & Powell, 1991). As a result, for constant sample size and selection ratio the accuracy of the corrections should not meaningfully vary. I did not find any similar studies of Lawley's correction, which may owe to the fact that in such a context there are numerous population correlations involved. I did not find any studies of incidental selection.

Hypothesis 1.2. I found no studies that examined this, but I expect that as the correlations among the variables within a study increase, Lawley's will provide more accurate results than the univariate correction. This is based on the fact that as the correlations among the variables increases, so too will the covariances of their correlations. The univariate corrections ignore that covariance, while Lawley's accounts for it in the correction procedure. Put another way, if all of the correlations were totally independent univariate corrections would pose no risk of producing out-of-range values and should be equivalent to Lawley's correction.

Hypothesis 1.3. Holding all else constant, the corrections should be more accurate as sample size increases. Previous simulation studies have shown evidence that there is a positive relationship between samples size and the accuracy of the univariate correction formulas (Bobko, 1983; Duan & Dunlap, 1997). I found no studies evaluating this with Lawley's correction nor for incidental selection, but it seems reasonable this relationship will hold.

Hypothesis 1.4. Holding all else constant, as the selection ratio increases, the accuracy of the corrections will also increase. Prior research has shown that as the selection ratio increases, the accuracy of the univariate corrections improves (Duan & Dunlap, 1997; Held & Foley, 1994) as does the multivariate (Held & Foley, 1994). Although the corrected correlations are better

estimates than leaving them uncorrected, there is still a good deal of bias in small selection ratios. More specifically, even in large samples ($n = 200$) the corrected correlations still contain a good deal of bias when selection ratios are small, though less than their uncorrected counterparts (Duan & Dunlap, 1997). This phenomenon is well established and related to the fact that as the selection ratio becomes smaller, the regression of the criterion on the predictor begins to differ between the selected and unselected groups (M. J. Allen & Yen, 1979; Linn, 1968). As a result of this M. J. Allen and Yen (1979) recommend that the corrections not be used when less than 30% of the unrestricted sample is selected. As the selection ratios become smaller, the assumption of linearity underlying the corrections is more increasingly violated—smaller correlations imply less linearity. Lawley's correction makes the same assumption and should perform similarly.

Research question 2: Inadmissable values. How frequently do univariate, element-wise corrections for selection produce inadmissible vectors of correlations compared to a multivariate, simultaneous correction when applied to a matrix of correlations? I did not locate any prior research evaluating the frequency with which these risks occur for these corrections. Schram (1995) did examine these risks in applying corrections for reliability and found no practical difference in the methods although univariate methods did occasionally reproduce invalid results. In a slightly different context, Landis (2013) notes that concerns about invalid matrices in the construction of a single input matrix from independently estimated meta-analytic correlations have not been empirically tested. Applying univariate corrections element-wise to a correlation matrix is analogous to assembling a matrix element-wise.

Hypothesis 2.1. I expect to observe at least some values out of range when applying the univariate corrections and none at all when applying Lawley's correction. I expect to find corrected correlation values exceeding the bounds of Hubert's inequality (1972) formula more frequently in the univariate case as ρ becomes larger because, as I showed earlier, in the extreme tri-variate case where $r_{12} = 1$ then $X_1 = X_2$ and so $r_{13} = r_{23}$. Thus, when $r_{13} = 1$ is known, r_{23} can only be

one value, which is equal to r_{13}. It stands to reason, then, that as r_{12} or r_{13} become stronger, the possible values of r_{23} become more restricted and this is when we would expect to observe out-of-range values. The same logic should follow for more than three variables and given that more correlations are involved, the restraints may become even greater. Violations are more likely in the univariate correction cases because when we correct a single correlation at a time and ignore the others completely. There is nothing in the procedure to bound the values of the correlations. I do not expect to find any violations of Hubert's inequality after applying the multivariate correction since the entire correlation matrix is transformed simultaneously and should respect the bounds on values of the elements.

Research question 3: Non-positive definite matrices. How frequently do univariate, element-wise corrections for selection result in a non-positive definite correlation matrix compared to a multivariate, simultaneous correction when applied to a matrix of correlations? If inadmissible values are entered into a correlation matrix, it will become non-positive definite. Hubert's inequality (1972) is a variable-wise evaluation of the matrix, examining its submatrices. A simultaneous evaluation of the entire matrix is also necessary. This can be accomplished through the determinant.

Hypothesis 3.1. The frequency of non-positive definite matrices will be greater in the univariate corrections and non-existent in the multivariate-corrected matrices. We can use the matrix determinants to evaluate each matrix. Any that are less than or equal to zero are non-positive definite. In using the determinant of the matrices to evaluate them, it is not clear analytically how to predict the conditions under which the corrected matrices might be made non-positive definite. The prediction that it might happen more frequently with the univariate cases is simply due to the fact that, by definition, the correlation matrix starts out as positive definite. Any simultaneous transformation of the matrix, as through Lawley's correction, ought to preserve that property with greater frequency than a procedure whereby we manipulate the

elements independently. Lawley's procedure, by definition, returns a covariance matrix that can always be standardized to a correlations matrix and, so, should never return a non-positive definite correlation matrix. On the other hand, the univariate corrections need not, by definition, result in a positive definite matrix.

Modeling the correlated effects. We already know that correlations corrected for artifacts, including range restriction, are more variable than their uncorrected counterparts (Raju & Brand, 2003). Analytic solutions exist for estimating the standard errors of the corrected correlations. To model dependence, however, we also need an estimate of their within-study covariances. As yet, an estimate of this covariance has not been published. As a result, it is prudent to examine current meta-analysis procedures for a reasonable alternative to exact estimates of the covariance matrix. I will examine robust variance estimation (RVE) for this purpose.

RVE provides a method for accounting for the dependence among effects that does not require knowledge of the covariance among the effects (Hedges et al., 2010). Theoretically, RVE is agnostic to the actual effect size being synthesized and its true distribution and to dependence structure of the effects. As a result, RVE is very promising, but its applications have not been thoroughly studied. I will evaluate how well RVE performs when the effects being modeled have been corrected for range restriction.

Research question 4: RVE C.I. coverage rates. What are the coverage rates of the confidence intervals produced by robust variance estimation when pooling correlations corrected for selection effects? Prior studies of the robust variance estimation procedure have shown it to be effective (Hedges et al., 2010; Hong, Riley, & Chen, 2017; López-López et al., 2017). Both Hedges et al. (2010) and Hong et al. (2017) used the confidence interval coverage rates as evidence that the method produced properly estimated standard errors.

Hypothesis 4.1. Coverage rates of the confidence intervals produced by the RVE method will be close to their nominal value of 95% and these rates will improve as both the number of studies and within-study sample size increase.

CHAPTER 3: METHOD

To examine the application of corrections for selection effects to the multivariate meta-analysis context, I conducted two simulation studies. Simulations provided me with direct control over and knowledge of the true parameters in the study. This, in turn, permitted me to accurately evaluate the statistics derived from the methods applied to data with known parameters, which cannot be done using empirical data. All of the simulation and analysis for this study was performed using R 3.6.0—Planting of a Tree (R Core Team, 2018).

I begin by describing the first simulation, which I used to answer research questions 1–3. I then move to the second simulation, which I used to answer research question 4. For both simulations, I first describe how I generated the data to be evaluated and then the methods by which I made the evaluations. Throughout, I liberally subscript mathematical objects at the risk of being more explicit than is sometimes necessary. Because some aspects of my research concern the statistics and parameters of observations, while others concern the statistics and parameters of correlations I have tried to be maximally explicit. Complete programs for both simulations are available in Appendices C and D.

Evaluating the Univariate and Multivariate Corrections, RQ1–RQ3

The first three research questions concern how well the univariate and multivariate corrections for selection effects perform. To answer each research question, I evaluated their performance using a different criterion. Multiple criteria provided a more complete picture of overall performance. In the following sections I describe the process I used to generate unrestricted observations, to introduce selection effects, to apply the corrections, and the analyses used to respond to each research question.

Step 1: Data generation. The first step was to generate 1,000 sets of complete, unrestricted observations. In a practical context these observations represent a sample prior to any selection process taking place. Thus, in the context of admissions selection research the unrestricted samples represent an applicant pool. In the study of an intervention, these unrestricted samples would represent the target population prior to recruitment into the study.

Let $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})' \sim N_4(0, 1)$ be a vector of random variables in sample i. Then $\boldsymbol{\rho}_i = (\rho_{i12}, \rho_{i13}, \rho_{i14}, \rho_{i23}, \rho_{i24}, \rho_{i34})'$ is the vector of correlations among the variables in $\mathbf{X}_i$. For each $\boldsymbol{\rho}_i$ I sampled $\mathbf{X}_i$ 1,000 times with sample size N using the `mvrnorm` function from the *MASS* package (Venables & Ripley, 2002). I used seven variations of $\boldsymbol{\rho}_i$ and four variations of N for a total of 28 conditions under which I randomly generated the unrestricted samples and computed and stored both their observed correlation matrices and the standard deviations of each variable. The conditions included:

- For $\boldsymbol{\rho}_i = (\rho_{i12}, \rho_{i13}, \rho_{i14}, \rho_{i23}, \rho_{i24}, \rho_{i34})'$:
  - $\boldsymbol{\rho}_1 = (.00, .00, .50, .00, .50, .50)'$
  - $\boldsymbol{\rho}_2 = (.10, .10, .50, .10, .50, .50)'$
  - $\boldsymbol{\rho}_3 = (.30, .30, .50, .30, .50, .50)'$
  - $\boldsymbol{\rho}_4 = (.50, .50, .50, .50, .50, .50)'$
  - $\boldsymbol{\rho}_5 = (.70, .70, .50, .70, .50, .50)'$
  - $\boldsymbol{\rho}_5 = (.90, .90, .50, .90, .50, .50)'$
  - $\boldsymbol{\rho}_7 = (.42, .30, .24, .47, .19, .21)'$
- $N = 30, 50, 100, 300$

Population correlations, ρ**.** The population correlations represent the relationships among the variables being examined in a meta-analysis. For the first six $\boldsymbol{\rho}_i$ I varied the correlations among X_1, X_2, and X_3 from a value of .00 to .90 where $\rho_{12} = \rho_{13} = \rho_{23}$. The same intervals were used by Duan and Dunlap (1997) in their evaluation of the univariate correction for explicit selection

and their standard errors. Millsap (1989), also evaluating corrections for explicit selection, used the same range of correlations but in increments of $.10$. On the other hand, I kept all of the correlations with X_4 constant such that $\rho_{14} = \rho_{24} = \rho_{34} = .50$. This permitted me to examine how corrections of correlations with a constant population value performed as the other correlations changed. For $\boldsymbol{\rho}_1$ the common correlation, $.00$, represented the case where all of the variables in $\mathbf{X}$ were independent of each other except in their relationship with X_4. I took the last set of population correlations, $\boldsymbol{\rho}_7$, from a meta-analysis conducted by Klieger et al. (2014). They examined the predictive validity of verbal (V), quantitative (Q), and analytic writing (W) scores of the GRE with cumulative graduate school GPA at public universities in Florida. After correcting for artifacts including selection effects, they reported the following correlations among doctoral students in all programs: $.42_{Q,V}, .30_{Q,W}, .24_{Q,GPA}, .47_{V,W}, .19_{V,GPA}, .21_{W,GPA}$. The population correlations in $\boldsymbol{\rho}_1$–$\boldsymbol{\rho}_6$ cover the parameter space that includes $\boldsymbol{\rho}_7$, but I included it to anchor the results in a realistic set of correlations.

Sample size, N**.** The samples sizes in N (30, 50, 100, and 300) represent the size of the unrestricted samples before selection takes place. In practice, they would represent the size of an applicant pool or sample of the true target population prior to selection. Duan and Dunlap (1997) used sample sizes of 50, 100, and 200 in their investigation of univariate range restriction corrections for explicit selection and their standard errors. Millsap (1989) used sample sizes of 25, 60, and 100 in his study of the standard errors of correlations corrected for direct range restriction.

Step 2: Introducing selection. Up to this point in the procedure I generated 7 population correlations $\times$ 4 samples sizes $\times$ $1,000$ replications $= 28,000$ samples of $\mathbf{X}_i$, all of which were free of any selection bias. All of the error in these samples was due to random sampling error alone. I then introduced selection bias into each of the 28,000 samples. I selected observations based on the value of X_1. This simulates, for example, using a test score to admit students to a program of study. It also mimics situations in which self-selection operates such that all individuals with

a given range of scores on, say, an inventory of motivation choose not to participate in a study (e.g. Kriegbaum et al., 2018). Explicit selection on X_1 created direct range restriction on all of the relationships with X_1 and indirect range restriction on all of the relationships that excluded X_1 (M. J. Allen & Yen, 1979).

I examined the full range of selection ratios from $.10$ to $.90$ in increments of $.10$ (Duan & Dunlap, 1997; Greener & Osburn, 1979; Held & Foley, 1994; Millsap, 1989). Selection ratios of zero and one correspond to selecting no one and everyone respectively. Clearly, a ratio of zero makes no sense as a condition for the study. The original samples free of selection effects generated in the previous step represent a selection ratio of 1. To perform the selection based on the value of X_1, I calculated the decile of the unrestricted sample equal to 1 less the selection ratio and selected all of the values greater than or equal to that decile, saving each subset of observations. For each of the nine subsets corresponding to each selection ratio, I stored their correlation matrices and standard deviations.

Step 3: Applying the corrections. To this point I had generated 28,000 unrestricted samples $+ 28,000 \times 9$ restricted subsets $= 280,000$ sets of data. In the 252,000 restricted subsets, I corrected all of the relationships with X_1 for explicit selection using Thorndike's Case II formula and all relationships excluding X_1 for incidental selection using Thorndike's Case III formula and stored the resulting corrected correlations. I then applied Lawley's multivariate correction to each subset using the `correct_matrix_mvrr` function provided in the *psychmeta* package (Dahlke & Wiernik, 2018) and stored the resulting corrected correlations. This completed the data generation process.

Step 4: Analysis. I performed three separate analyses of the generated data, which I now describe. Each analysis corresponded to one of three research questions.

Research question 1: Bias in the corrections. To evaluate how well the univariate and multivariate corrections recovered the unrestricted correlations, I calculated the median bias in each condition. The median was more appropriate because the corrected correlations and, thus, the biases were negatively skewed. Greener and Osburn (1979) also used the absolute median bias in their study of corrections for explicit selection in empirical sets of data. In the unrestricted samples only sampling error was present. In the restricted samples, however, both sampling error and the error due to selection effects were present. Because I am correcting the correlations for the bias due to selection and not due to sampling error, it was more appropriate to determine the bias remaining after correction by comparing the corrected correlations to the observed correlations in the unrestricted samples, which were good approximations of the true population correlation on average.

Research questions 2 and 3: Invalid matrices. I assessed the degree of bias in each corrected correlation, but such a measure does not indicate whether or not the resulting matrix after correction is valid. Bias measures one element of the matrix, not the entire matrix itself. I used three metrics to measure the matrices themselves. To answer the question of whether inadmissible values were contained in any matrix (RQ2), I first examined the corrected correlations for any absolute values greater than 1, which would immediately invalidate any correlation matrix. Next, I evaluated each matrix according to the inequality derived by Hubert (1972) as shown in Equation 2 (RQ3). More explicitly, for X_1 I evaluated $\mathbf{G}'\mathbf{A}^{-1}\mathbf{G} \leq 1$ where $\mathbf{G} = [cov(X_1, X_2), cov(X_1, X_3), cov(X_1, X_4)]'$ and $\mathbf{A}$ was the covariance matrix of X_2, X_3, and X_4. For X_2, $\mathbf{G} = [cov(X_1, X_2), cov(X_2, X_3), cov(X_2, X_4)]'$ and $\mathbf{A}$ was the covariance matrix of X_1, X_3, and X_4 and so on for X_3 and X_4. Next, I evaluated the determinant of each matrix for values less than or equal to zero (RQ4). The determinant represents the most complete simultaneous evaluation of the entire matrix as even Hubert's inequality requires a variable-wise method of evaluation. Matrices with determinants less than or equal to zero are non-positive

definite and, by definition, are not correlation matrices (Wothke, 1993).

Evaluating the Robust Variance Estimation Procedure, RQ4

In a meta-analysis context, once we have sets of corrected correlations, we still need to pool them. In a multivariate meta-analysis context, this means accounting not only for their variances, but their covariances (Becker, 1992; Raudenbush et al., 1988). Recall from Chapter 2 that corrected correlations are more variable than their uncorrected counterparts so it stands to reason their covariances are also greater. While closed form estimators of their variances are available, we do not have a closed form estimator for their covariances. The robust variance estimation (RVE) method of synthesis requires no knowledge of the covariance structure of the correlations being synthesized. Thus, it holds promise as a method for properly synthesizing the estimates and providing proper standard errors (Hedges et al., 2010). To evaluate the method, I generated sets of meta-analytic observations and evaluated the coverage rates of the confidence intervals resulting from RVE. I begin by describing the data generation process then move to introducing selection effects, applying the corrections, specifying the RVE model, and evaluating the results.

Step 1: Generating the data. To generate the data for this part of the simulation, I chose to simulate a random-effects model since they are thought to be the most realistic meta-analysis models in practice (Hedges & Vevea, 1998; Schmidt & Hunter, 2015). In these models, the observed study-level effects vary around true study-level effects that themselves vary around a common population effect. Let $\mathbf{X}_i$ denote a vector of four random variables observed in study i such that $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})' \sim N_4(0, 1)$. The number of studies is k. When the number of variables, p, is four then the number of unique correlations among them is given by $p^* = \frac{p(p-1)}{2} = 6$. Now, let $\mathbf{r}_i$ be the vector of observed study correlations among the observations in $\mathbf{X}_i$ for study i such that $\mathbf{r}_i = (r_{iX_1X_2}, r_{iX_1X_3}, r_{iX_1X_4}, r_{iX_2X_3}, r_{iX_2X_4}, r_{iX_3X_4})' \sim N_6(\boldsymbol{\rho}_i, \boldsymbol{\Sigma}_i)$

where $\boldsymbol{\rho}_i$ is a vector of true study correlations for study i around which $\mathbf{r}_i$ is centered with covariance matrix $\boldsymbol{\Sigma}_i$. Finally, $\boldsymbol{\rho}_i = (\rho_{iX_1X_2}, \rho_{iX_1X_3}, \rho_{iX_1X_4}, \rho_{iX_2X_3}, \rho_{iX_2X_4}, \rho_{iX_3X_4})' \sim N_6(\mathbf{P}, \boldsymbol{\Psi})$ where $\mathbf{P}$ is a vector of population correlations around which $\boldsymbol{\rho}_i$ is centered with covariance $\boldsymbol{\Psi}$. This denotes a random effects meta-analysis model:

$$\mathbf{r}_i = \mathbf{P} + \mathbf{u}_i + \mathbf{e}_i \tag{11}$$

where $\mathbf{u}_i \sim N_6(0, \boldsymbol{\Sigma}_i)$ and $\mathbf{e}_i \sim N_6(0, \boldsymbol{\Psi})$ are vectors of between-study and within-study error terms respectively.

I defined the population correlations as $\mathbf{P} = (.42, .30, .24, .47, .19, .21)'$, which are the correlations among the quantitative, verbal, and writing portions of the GRE, and graduate school GPA studies by Klieger et al. (2014). I chose a population vector of six correlations, which corresponds to four variables in each study. Three variables would have represented the simplest multivariate case, but would not permit a thorough evaluation (Hafdahl, 2008). Recall from Equation 6 that the covariance of two correlations that do not share an index is a function of all six intercorrelations among the four variables. A minimum of four variables is necessary to simulate this situation.

To generate study-level correlations, I first had to specify a between-study covariance matrix, $\boldsymbol{\Psi}$. Specifying $\boldsymbol{\Psi}$ must be done with care because large variances and covariances will produce correlations that exceed an absolute value of one (Hafdahl, 2008). I set all off-diagonal elements of $\boldsymbol{\Psi}$ to zero. Thus, there was no between-studies covariance of the effects. To determine the variances Hafdahl (2008) and Hedges et al. (2010) reduced τ^2, the between-study variance of the effects, by a factor to produce a desired ratio of between- and within-study variance. The variance components in $\boldsymbol{\Psi}$ represent the degree of heterogeneity in true study effects. Because the random effects model is considered more realistic in practice, it is important

to examine varying degrees of heterogeneity. The case where all elements in $\mathbf{\Psi}$ is zero represents no heterogeneity and a fixed-effects model. Some conventions for small, medium, and large degrees of heterogeneity were proposed by Hedges and Pigott (2001, 2004). These correspond to ratios of τ^2 to v of $\frac{1}{3}$, $\frac{2}{3}$ and 1 respectively where v is the expected variance within studies. These values were also used by Hedges et al. (2010) when they proposed and tested the RVE method. Thus, to test these degrees of heterogeneity, I set the diagonal elements of $\mathbf{\Psi}$ to $v * h$ where v is the expected variance given by $\frac{(1-\rho^2)^2}{N}$ and h is the ratio of heterogeneity equal to $\frac{1}{3}$, $\frac{2}{3}$ or 1.

With the population correlations and their covariances defined, I generated k vectors of true study-level correlations. I then used these study-level correlations to generate unrestricted observations of sample size N following the same procedure outlined in the data generation step in the first simulation. In this simulation, I replicated each set of conditions 500 times. For this simulation, then, the data generating conditions were:

- $\mathbf{P} = (.42, .30, .24, .47, .19, .21)'$
- $k = 10, 30, 100$
- $N = 50, 100, 300$
- $\frac{\tau^2}{v} = \frac{1}{3}, \frac{2}{3}, 1$

Sample sizes, k ***and*** N***.*** The number of studies, k, represents the size of the meta-analysis sample. The within-study sample size, N, is the size of the sample within each of the k studies in the meta-analysis sample. I originally planned to use $N = 30$, but based on the results of the first simulation excluded that condition here due to the number of invalid matrices produced by both correction methods. The values of both k and N I specified for this study are similar to the values in other methodological studies of this nature (e.g. S. F. Cheung et al., 2018; Hafdahl, 2008; Hedges et al., 2010). In their original presentation of the RVE method, Hedges et al. (2010) used both k and N of 10, 20, and 40.

Step 2: Introducing selection. At this point, I will have generated 3 meta-analysis sample sizes $\times$ 3 within-study sample sizes $\times$ 3 τ^2 ratios $\times$ 500 replications $=$ $13,500$ unrestricted samples. I introduced selection effects into each of these as described in the prior section resulting in $135,000$ samples total.

Step 3: Applying the corrections. Based on the results of the first simulation, I applied Lawley's multivariate correction to all of the restricted samples and stored the corrected correlations.

Step 4: Calculating the effect size variances. Hedges et al. (2010) recommend using asymptotic inverse variances to weight the effects in a robust variance meta-analysis. For any unrestricted correlation, I calculated its variance as in Equation 8 (Olkin & Siotani, 1976). To estimate the variance of the corrected correlations, I used the approximation in Equation 12 provided by Schmidt and Hunter (2015).

$$\left(\frac{r_c}{r}\right)^2 \frac{(1-r^2)^2}{N} \tag{12}$$

where r_c is the corrected correlation, r is the observed correlation before correction, and the second term is just the variance of an unattenuated correlation as given in Equation 8.

Step 5: Applying the RVE method. With the corrected correlations from each of the 500 replications, I applied the RVE method to estimate mean correlations and their standard errors. RVE is implemented in the R package *robumeta* (Fisher, Tipton, & Zhipeng, 2017). The complete RVE model implemented in *robumeta* is described in Hedges et al. (2010) and a tutorial with examples is available in Fisher and Tipton (2015). The package includes corrections to both the estimator and degrees of freedom for models that include fewer than 40 studies by default (Fisher & Tipton, 2015; Tipton, 2015), which I left in place regardless of the number of studies as

recommended by Fisher and Tipton (2015). Furthermore, I specified a correlated effects model. An alternative hierarchical effects model is available through the package, but is appropriate when effects are thought to be correlated between-studies due to, for example, their originating from the same authors or labs (Fisher & Tipton, 2015). Recall that I set all between-study covariances to zero.

RVE requires that one specify a linear model of the effects. I specified a model with the vector of study correlations as the outcome, no intercept, and a dummy variable indicating each of the six correlations. Without this specification, *robumeta* will treat the effects as measuring a single outcome and return a single estimate instead of an estimate of each effect. Park and Beretvas (2018) used the same approach to compare the performance of three-level hierarchical meta-analysis models to RVE with multiple outcomes per study as did Hong et al. (2017) in their evaluation of RVE as proposed by Hedges et al. (2010) and their proposal of an alternative RVE model based on prior work by Riley, Thompson, and Abrams (2007).

Step 6: Analysis. It is important to understand that although accounting for the within-study correlations is important, the goal of the random effects model is to estimate $\mathbf{P}$ and $\boldsymbol{\Psi}$—the true population effects and their variances (Wei & Higgins, 2012). Whether or not one models the covariances has little-to-no effect on the point estimates of the effects (Schmidt & Hunter, 2015; Tracz, Elmore, & Pohlmann, 1992). Instead, the standard errors are of interest. To evaluate the results, I examined the coverage rates of the confidence intervals produced by the RVE procedure. This is similar to the evaluation by Hedges et al. (2010) in their original presentation of RVE and that of the evaluation by Hong et al. (2017).

CHAPTER 4: RESULTS

In this chapter, I discuss the results of the simulation studies. The information is organized by research question. Throughout, I refer to the different specifications of $\boldsymbol{\rho}$ by the shared correlation among r_{12}, r_{13}, and r_{23}. For example, in the first condition I specified $\boldsymbol{\rho} = (.00, .00, .50, .00, .50, .50)'$ and so I refer to this condition as having a population correlation of .00. Recall from the research design that in all conditions $r_{14} = r_{24} = r_{34} = .50$. Finally, it will be helpful to keep in mind that the correlations from Klieger et al. (2014) were included to investigate a set of correlations found in empirical research. All of the correlations from the study fall within the parameter space covered by the other conditions I investigated. I include the results from the study in tables and figures for completeness, but they do not suggest any conclusions that differ from the other conditions, and so I do not discuss them separately.

Data Generation

I generated the data as described in the preceding chapter. At a selection ratio of $.10$ and $N = 30$, the selection on X_1 itself created data with three observations and their correlation matrices were sometimes non-positive definite. I retained this condition and those matrices in the study because it was the selection process that made them invalid, not the method by which the original unrestricted observations were generated. Furthermore, I evaluated these matrices after they were corrected, so they did not represent the final matrix of interest. If they remained invalid after correction, then they were counted in the results for research questions two and three. For completeness and in the interest of transparency, I report the number of invalid matrices in Table 3.

Table 3

Number and Percentage of Invalid Correlation Matrices Produced by Selection

ρ	N	%
.00	12	1.2%
.10	35	3.5%
.30	24	2.4%
.50	24	2.4%
.70	20	2.0%
.90	57	5.7%
Klieger	14	1.4%

Note. $N = 30$, selection ratio = $.10$.

Research Question 1: Bias in the Corrections

As indicated by absolute median bias, how well do univariate, element-wise corrections for selection effects recover the unrestricted correlations compared to a multivariate, simultaneous correction when they are both applied to a matrix of correlations? I begin by making some general statements about the observed correlations, and then focus on the corrections, discussing those correlations affected by explicit and incidental selection separately. The reader will find it helpful to refer to Figures 6–9, which show the median absolute bias in the observed correlations before correction, in univariate-corrected correlations, and in multivariate-corrected correlations. Tabled results for this section are available in Appendix E. Also note that I refer to any bias less than an absolute value of $.05$ as negligible. This is somewhat arbitrary with no standards in the literature on how great a departure in a correlation coefficient is meaningful. Readers should consider their own discipline and the consequences of error for their own studies.

Observed correlations. As expected, (M. J. Allen & Yen, 1979; Schmidt & Hunter, 2015), the bias in the observed correlations increased as the selection ratio decreased for constant sample size and population correlation. The degree of bias in observed correlations did not change with the sample size for constant selection ratio and population correlation. As expected, when $r_{12} = r_{13} =$

$r_{23} = .00$ the bias in observed correlations was negligible. This is most easily seen in Figures 6–9 by looking at the first column in each figure. This behavior was expected because when two variables are not correlated, then selecting on one should not affect the observations on the others and should not result in any bias. In some cases, however, when the population correlation was .00 the bias exceeded .05 when the selection ratio was less than .30 and $N < 300$. In these conditions, the bias was more likely to be due to sampling error given that there were only three observations at these extreme selection ratios. This can be seen by looking in Table E1 where the population correlation is .00.

As population correlations increased, the bias in the observed correlations also increased. This is best seen by looking at Figures 6–9 and reading across each row in the figure at a constant selection ratio. The bias in r_{14}, with population value of .50 in all conditions, remained consistent as the values of the other population correlations changed. On the other hand, the bias in r_{24} and r_{34}, both also equal to .50 in all conditions, was positive when the other correlations were .00 or .10 indicating some range enhancement. Recall from Chapter 2, that corrections for range enhancement and range restriction are equivalent. As the other correlations increased, the bias in these two became increasingly negative. Note that although $r_{14} = r_{24} = r_{34} = .50$, r_{14} was subject to explicit selection while r_{24} and r_{34} were subject to incidental selection. Changes in the other population correlations had little-to-no effect on the bias in observed r_{14} under explicit selection but did influence the bias in observed r_{24} and r_{34} under incidental selection. This helps to demonstrate one way in which selection on a single variable has differential effects on the correlations in the matrix and provides additional motivation for the study and support for more closely investigating the multivariate approach to making corrections when correlations are calculated on the same sample.

Corrections for explicit selection. All of the correlations with X_1 were subjected to explicit selection and were corrected using Thorndike's Case II formula. Under all conditions,

both the univariate and multivariate correction methods performed nearly identically. This is best seen in Figures 6–9 where it is evident that the results of both methods overlap. As seen in Table E1, absolute bias greater than 0.05 occurred consistently in population correlations of .30, .50, and .70 with low selection ratios of .10 and .20 and at unrestricted $N = 30$ or $N = 50$. The largest absolute bias was under the population correlation .30, $N = 30$, and selection ratio of .10 where the corrections were positively biased by 0.263 on average. Neither method offers a clear advantage when correcting for explicit selection on a single variable. Since there is no apparent advantage between the correction methods, the remaining discussion of corrections for explicit selection applies to both the univariate and multivariate correction methods.

Hypothesis 1.1: Population correlations. Prior research suggests that there is no relationship between the accuracy of the corrections for explicit selection and the true population correlation ρ in the univariate case (Duan & Dunlap, 1997; Mendoza et al., 1991). The authors of previous studies only examined a two-variable case, and the same result was not always borne out in the multivariate context. The two correlations to examine for this criterion are r_{12} and r_{13} because r_{14} was held at a constant value of .50. Holding sample size and selection ratio constant, changes in the bias across population values were only noticeable in selection ratios of .10 and .20. To see this, examine the first and second rows of Figures 6–9, which are at constant sample size for each figure. At selection ratios of .10, and sometimes .20, and $N = 30$ the bias increases through a population correlation of .30, falls at .50, increases again at .70, and falls again at .90. At these selection ratios and $N = 50$, the bias tends to increase through a population correlation of .50 and improves through .90. Thus, at unrestricted samples of $N \geq 100$ the population correlation did not affect the bias in the corrections.

Hypothesis 1.2: Population intercorrelations. I hypothesized that as the correlations among the variables within a study increase, the multivariate correction would provide more accurate results than the univariate correction. To evaluate this, I examined r_{14}, which was held

constant at $.50$ while the other correlations changed. As can be seen in Figures 6–9, the patterns of bias were the same in both the multivariate and univariate corrections as the population correlation changed. Thus, the multivariate correction did not perform better as I hypothesized it would. My original hypothesis was based on the fact that as the correlations among the variables increases, so too will their covariances. The univariate corrections ignore that covariance, while the multivariate correction accounts for it in the correction procedure. Under explicit selection, those relationships seem to be ignorable under the conditions simulated here. At $N < 100$, the corrections of this correlation tended to worsen from a population correlation of $.00$ to $.50$ then improve suggesting some relationship between the other correlations in the matrix.

Hypothesis 1.3: Sample size. I hypothesized that the corrections should be more accurate as sample size increases, holding all else constant. Previous simulation studies have shown evidence that there is a positive relationship between sample size and the accuracy of the univariate correction formulas for explicit selection (Bobko, 1983; Duan & Dunlap, 1997). This was true across all simulation conditions. As seen in Table E1, absolute bias never exceeded $.05$ in unrestricted samples of $N = 300$. Only when the population correlation was $.00$ and selection ratio $.10$ or $.20$ did absolute bias exceed $.05$ when $N = 100$. The relationship between the bias and sample size is more evident in lower selection ratios since both methods had negligible bias in selection ratios greater than $.20$ under all conditions.

Hypothesis 1.4: Selection ratio. I hypothesized that as the selection ratio increased, the accuracy of the corrections would also increase, holding all else constant. Prior research has shown that as the selection ratio increases, the accuracy of the univariate corrections for explicit selection improves (Duan & Dunlap, 1997). This was true of the corrected correlations in this study as well, though as I already noted, there was negligible bias in all conditions in selection ratios as low as $.30$. This is most easily seen in Figures 6–9 by looking at a single cell and observing the change in bias as the selection ratio increases. This finding aligns with Duan and Dunlap (1997) who

concluded that under small selection ratios, the corrections were not reasonable estimates of their population values.

The corrected correlations were always better estimates than their uncorrected counterparts at selection ratios as low as $.30$, and in this simulation, the absolute bias in those conditions was less than $.05$. It is worth noting, however, that at times the observed correlations were more accurate than their corrected values when selection ratios were less than $.30$, a finding that replicates Greener and Osburn (1979) who found this to occur at selection ratios $\leq .25$. Thus, it may be best to avoid corrections when selection ratios are less than $.30$ with the understanding that at such extreme selection we are studying a different population than the one represented by the unrestricted sample. This supports the recommendation by M. J. Allen and Yen (1979) to avoid using corrections when selection ratios are less than $.30$. Findings from this section are made plain in Figures 6–9, with Table E1 providing the actual biases.

Corrections for incidental selection. All of the correlations that excluded X_1 were subjected to incidental selection and were corrected using Thorndike's Case III formula. As with the corrections for explicit selection, both methods performed equivalently. Their performance is shown in Figures 6–9 and Table E1. As seen in the table, absolute bias greater than 0.05 occurred consistently at population correlations of $.00$, $.10$, and $.30$ with low selection ratios of $.10$ and $.20$ and at unrestricted $N = 30$ or $N = 50$. The largest absolute bias was under the population correlation $.30$, $N = 30$, and selection ratio of $.10$ where the corrections were positively biased by 0.251 on average. Neither method offers a clear advantage when correcting for incidental selection on a single selection variable. Since there is no apparent advantage between the correction methods, the remaining discussion of corrections for explicit selection applies to both the univariate and multivariate correction methods.

Hypothesis 1.1: Population correlations. Prior research suggests that there is no relationship between the accuracy of the corrections for explicit selection and the true population

correlation ρ in the univariate case (Duan & Dunlap, 1997; Mendoza et al., 1991), but I found no equivalent research for cases of incidental selection. The performance in r_{23} is the best indicator for this criterion because r_{24} and r_{34} were both held constant at $.50$. Holding selection ratios and sample size constant, the bias in both corrections was stable across population correlations except for $.00$ and $.10$ where greater bias was observed at selection ratios $< .50$ and $N = 30$.

Recall that in the corrections for explicit selection, when the unrestricted samples sizes were $N = 30$ or 50 the accuracy of the corrections changed with the population correlation at constant N and constant selection ratio. This only occurred at $N = 30$ in the corrections for incidental selection. This is not surprising since incidental selection attenuates the correlations less than explicit selection (see Figure 4). As a result, smaller unrestricted samples are more accurately corrected for incidental selection at lower selection ratios as compared to explicit selection. This difference disappears at an unrestricted sample size of ≥ 100. This is most easily seen in Figures 6–9.

Hypothesis 1.2: Population intercorrelations. This criterion is best examined through r_{24} and r_{34}, which were held constant at $.50$ while the other correlations were changed. I hypothesized that as the correlations among the variables within a study increased, the multivariate correction would provide more accurate results than the univariate correction. As with the corrections for explicit selection, this was not the case when correcting for incidental selection. At $N < 100$, the corrections tended to improve as the population correlations of the other variables increased. This differs from the explicit corrections that had increasing bias as the population correlation increased from $.00$ to $.50$, then improved, at the same N. Figures 6–9 make the results of this section easy to see.

Hypothesis 1.3: Sample size. I hypothesized that the corrections should be more accurate as sample size increases, holding all else constant. Previous simulation studies have shown evidence that there is a positive relationship between sample size and the accuracy of the

univariate correction formulas for explicit selection (Bobko, 1983; Duan & Dunlap, 1997). This was true for both corrections in every condition. As shown in Table E1, bias never exceeded $.05$ in unrestricted samples as large as 300 even at selection ratios of $.10$.

Hypothesis 1.4: Selection ratio. I hypothesized that as the selection ratio increased, the accuracy of the corrections would also increase, holding all else constant. Prior research has shown that as the selection ratio increases, the accuracy of the univariate corrections for explicit selection improves (Duan & Dunlap, 1997), but I found no equivalent research on incidental selection. The results from this study, however, suggest that this is also true for the corrections for incidental selection in both the univariate and multivariate cases. Figures 6–9 show this clearly.

Efficiency in the methods. With both the univariate and multivariate methods producing the same results in terms of the bias in corrections, I also examined their variability. Two equivalently biased estimators can be further distinguished by their efficiency (Hedges, 1982). Efficiency refers to the variability around an estimate. Less variable estimators are more desirable because they provide more consistent estimates from sample to sample. My examination revealed that neither method was more efficient than the other under the conditions studied here. The results are available in Appendix F. As shown in Figures F1-F4, efficiency improved as sample size increased, or population correlation increased, or selection ratio increased. For example, The variance in r_{12} at a selection ratio of .10 and population correlation of .10 was about 0.75 when $N = 30$, but declined to about 0.15 when $N = 300$ (Figures F1 and F4). The variance of the corrections of r_{12} at a selection ratio of .10 and $N = 50$ (Figure F2) was approximately 0.55 when the population correlation was .10 and declined to about 0.23 in a population correlation of .90. The changes as selection ratios increase are evident within each cell of each figure reading from right to left.

The only exception to these patterns was under incidental selection on correlations r_{24} and r_{34}, which were both held constant at $.50$ under all conditions. As the population value of the

other correlations increased, the variability in the corrections on these two correlations increased for constant sample size and selection ratio. The same pattern was not observed in r_{14}, which was also held at $.50$ but underwent explicit selection. This is best seen by examining each row in the figures comparing the observations holding selection ratio constant. At a selection ratio of $.10$ and population correlation of $.10$ the variance of the corrections in r_{24} is about 0.55 and increases to 0.70 with a population correlation of $.90$.

Summary. Altogether, the univariate and multivariate corrections performed equally well under the conditions I studied here. Their performance improved as sample size, and selection ratio increased with bias less than $.05$ in selection ratios greater than $.20$ and unrestricted $N > 30$. Notably, when bias did exceed $.05$, it was always positive and, therefore, not conservative as one might prefer. In general, corrections for incidental correction became less biased more quickly than corrections for explicit selection as the unrestricted sample size and selection ratio increased. This can be attributed to the fact that the effects of incidental selection are less extreme than those of explicit selection. Finally, in unrestricted samples as large as 300, absolute median bias never exceeded $.05$.

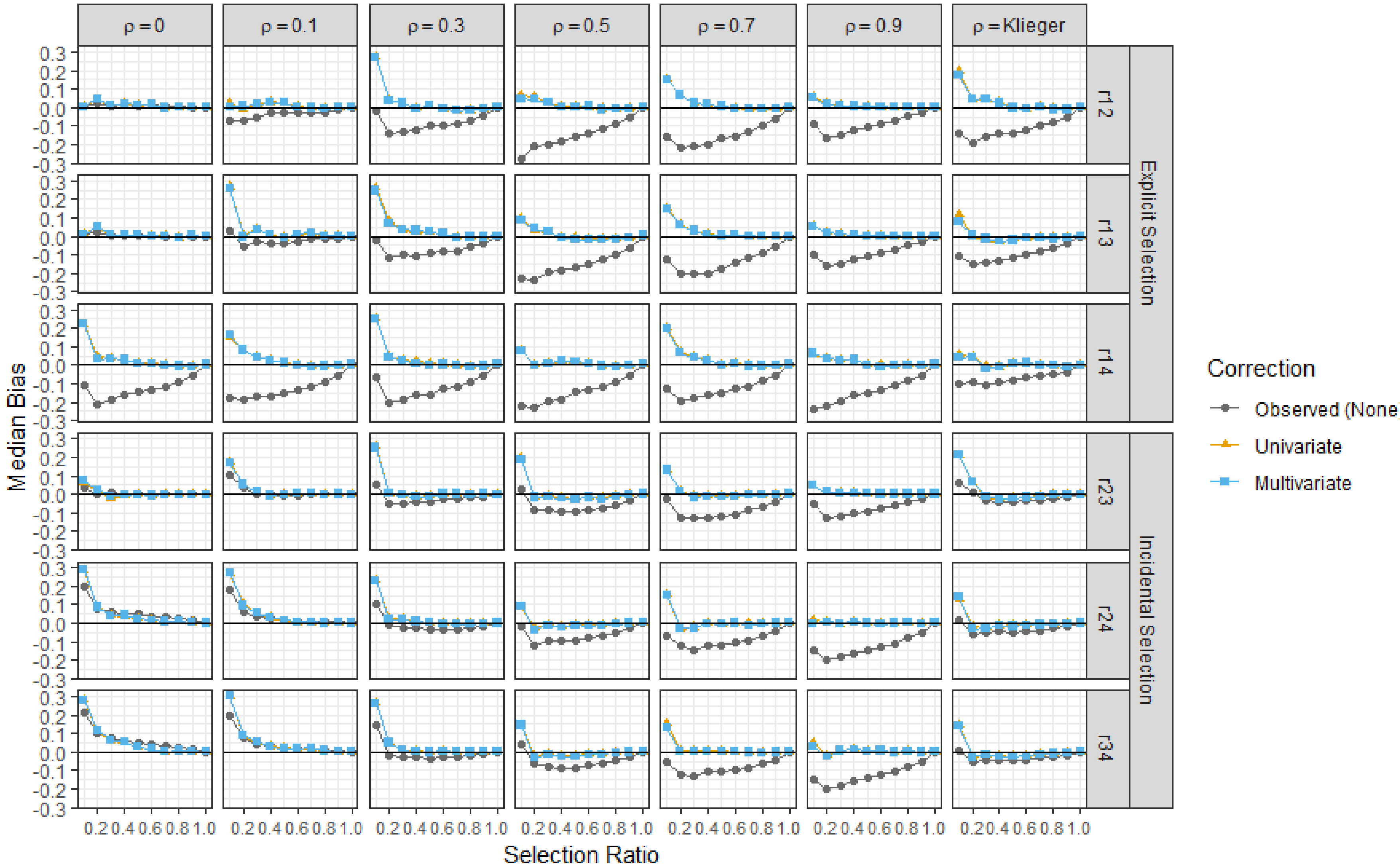

Figure 6. Bias in observed, univariate-corrected, and multivariate-corrected correlations by population correlation, ρ, and selection type, $N = 30$.

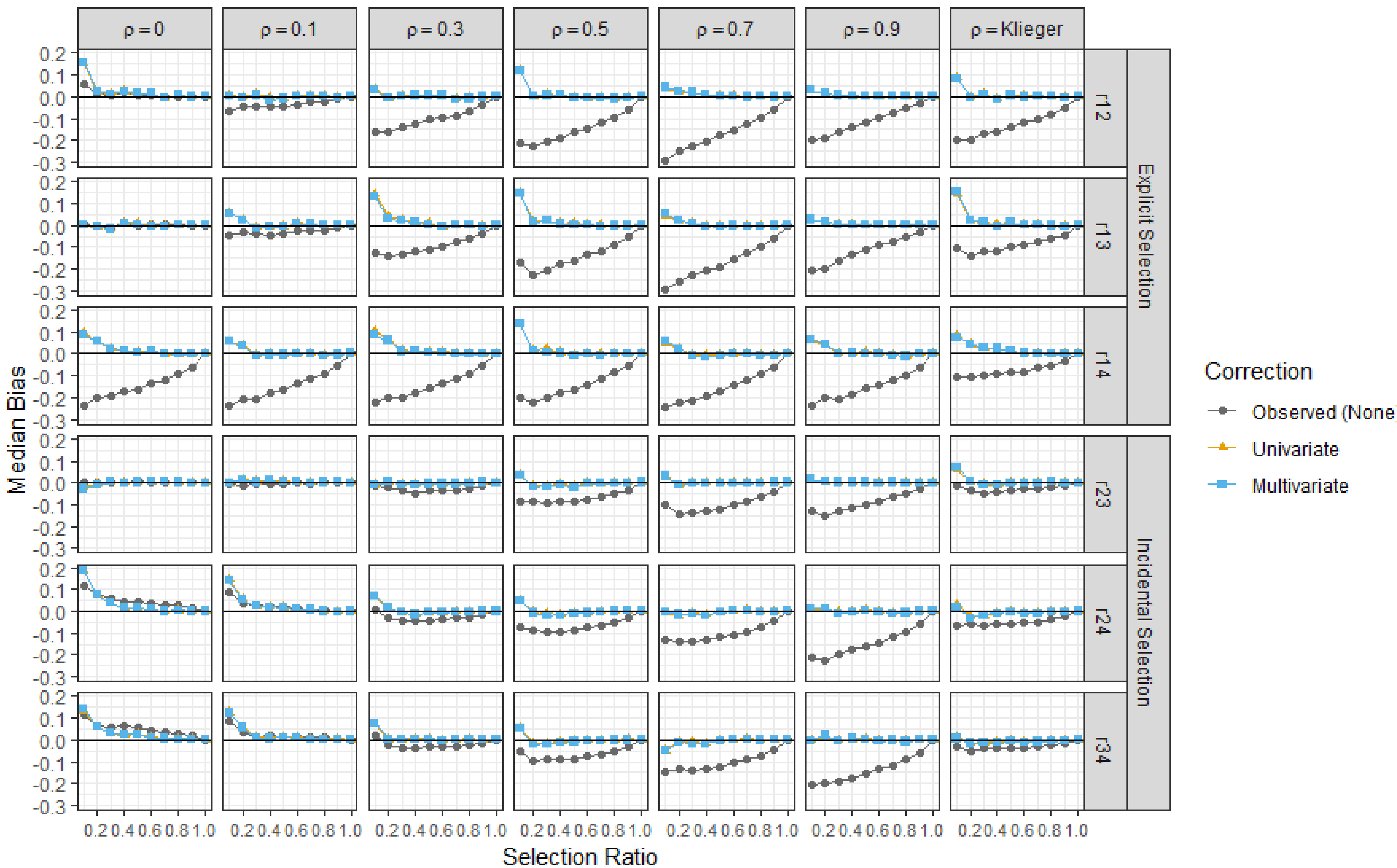

Figure 7. Bias in observed, univariate-corrected, and multivariate-corrected correlations by population correlation, ρ, and selection type, $N = 50$.

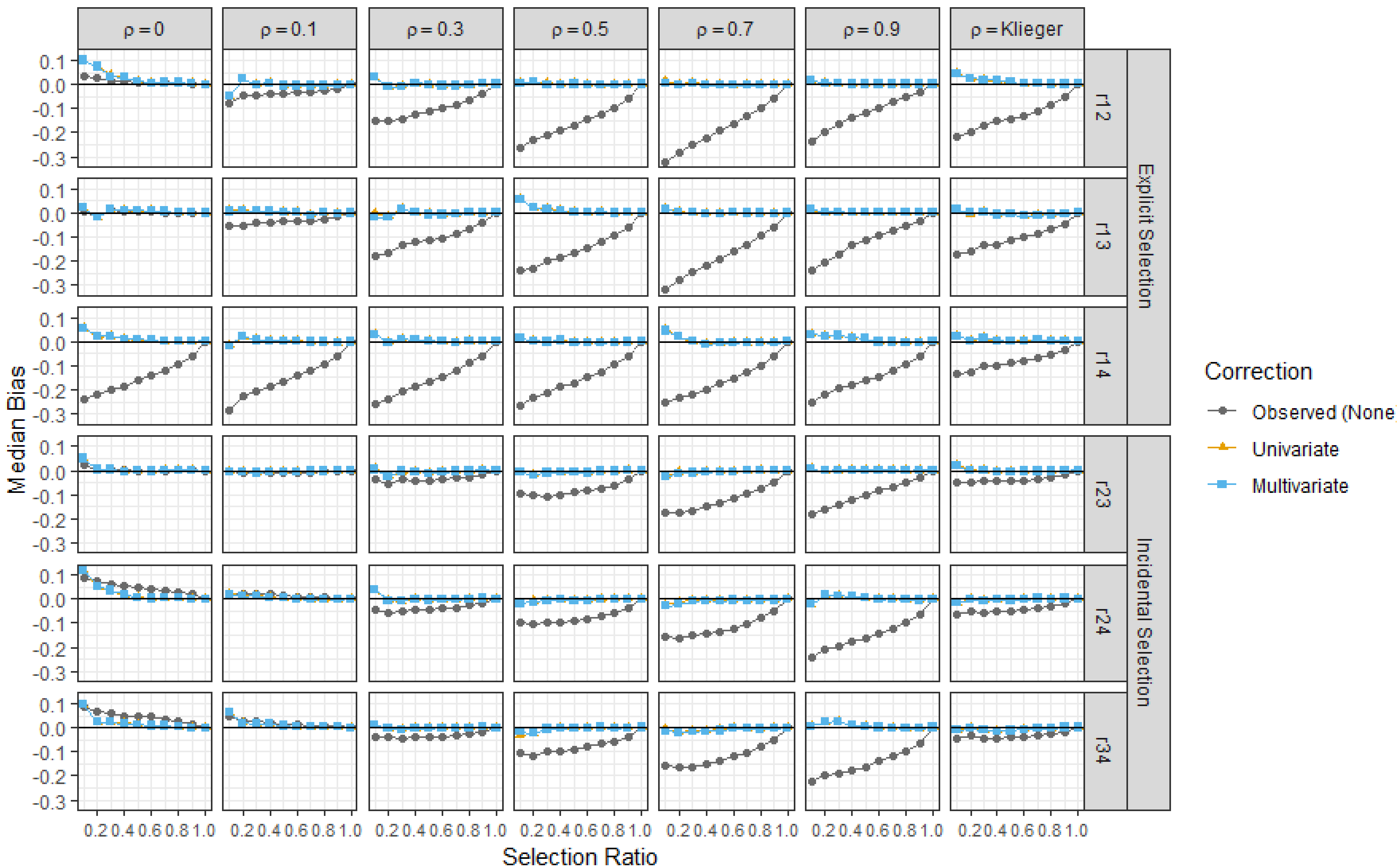

Figure 8. Bias in observed, univariate-corrected, and multivariate-corrected correlations by population correlation, ρ, and selection type, $N = 100$.

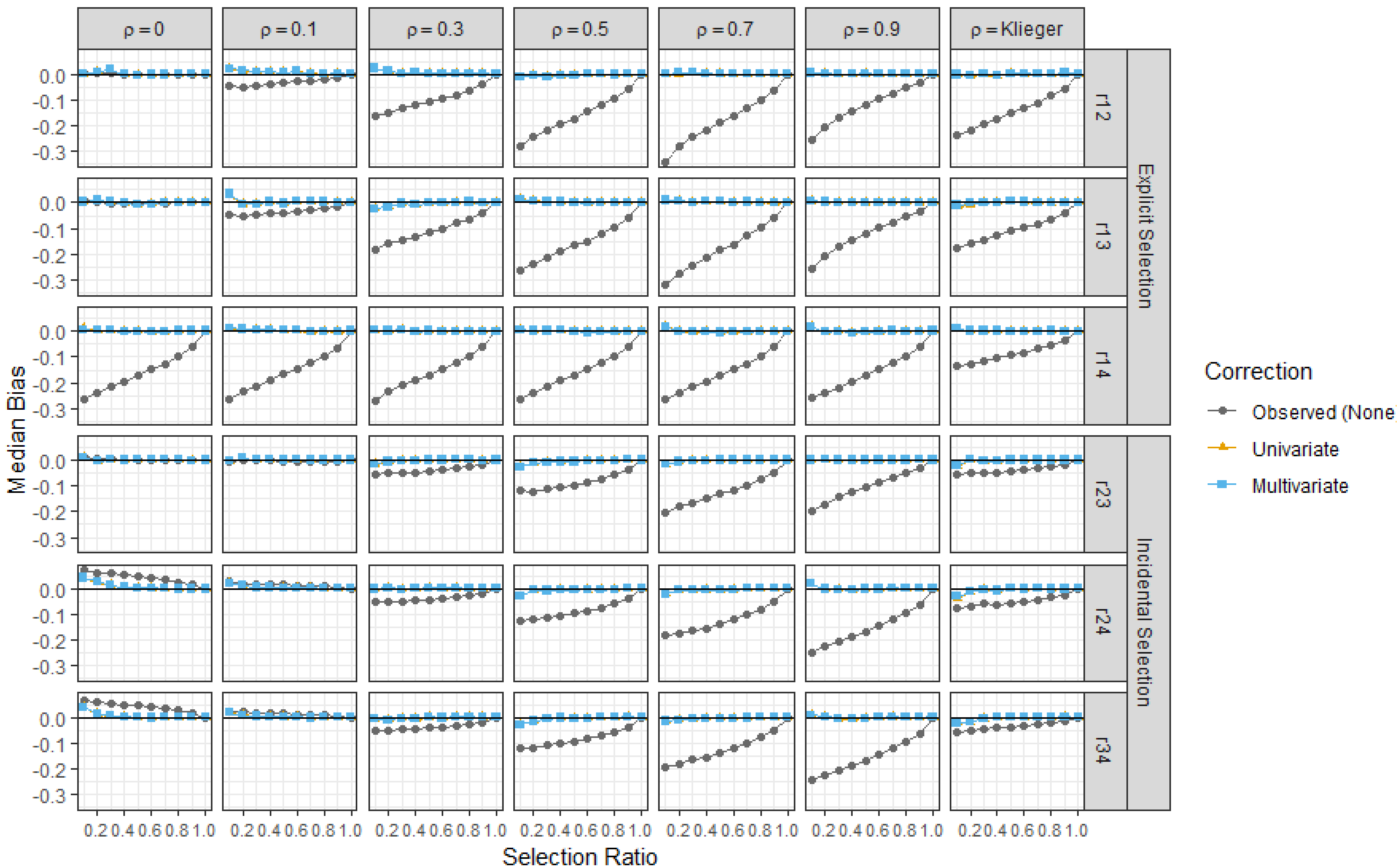

Figure 9. Bias in observed, univariate-corrected, and multivariate-corrected correlations by population correlation, ρ, and selection type, $N = 300$.

Research Question 2: Inadmissible Correlations

How frequently do univariate, element-wise corrections for selection produce inadmissible vectors of correlations compared to a multivariate, simultaneous correction when applied to a matrix of correlations? No correlations corrected by either method ever exceeded an absolute value of one. Table 4 and Figure 10 show the frequency with which each correction method violated Hubert's inequality under each simulation condition. I hypothesized that only the univariate corrections would produce inadmissible values. Both methods performed similarly, however. For all $\boldsymbol{\rho}$, both correction methods resulted in matrices with inadmissible values only when the selection ratio was $.10$ and $N = 30$. The poor results for both corrections methods at this extreme level of selection are consistent across population correlations. Furthermore, although some restricted correlation matrices were invalid after selection and before correction, the much larger number of invalid matrices after correction suggests that the corrections themselves do not reliability recover valid matrices after such extreme selection from a small unrestricted N.

Table 4

Number of Matrices Containing Inadmissible Values by Population Correlation.

$\boldsymbol{\rho}$	Univariate	Multivariate
.00	186	186
.10	195	203
.30	190	188
.50	156	165
.70	171	200
.90	251	211
Klieger	183	181

Note. $N = 30$, selection ratio = $.10$.

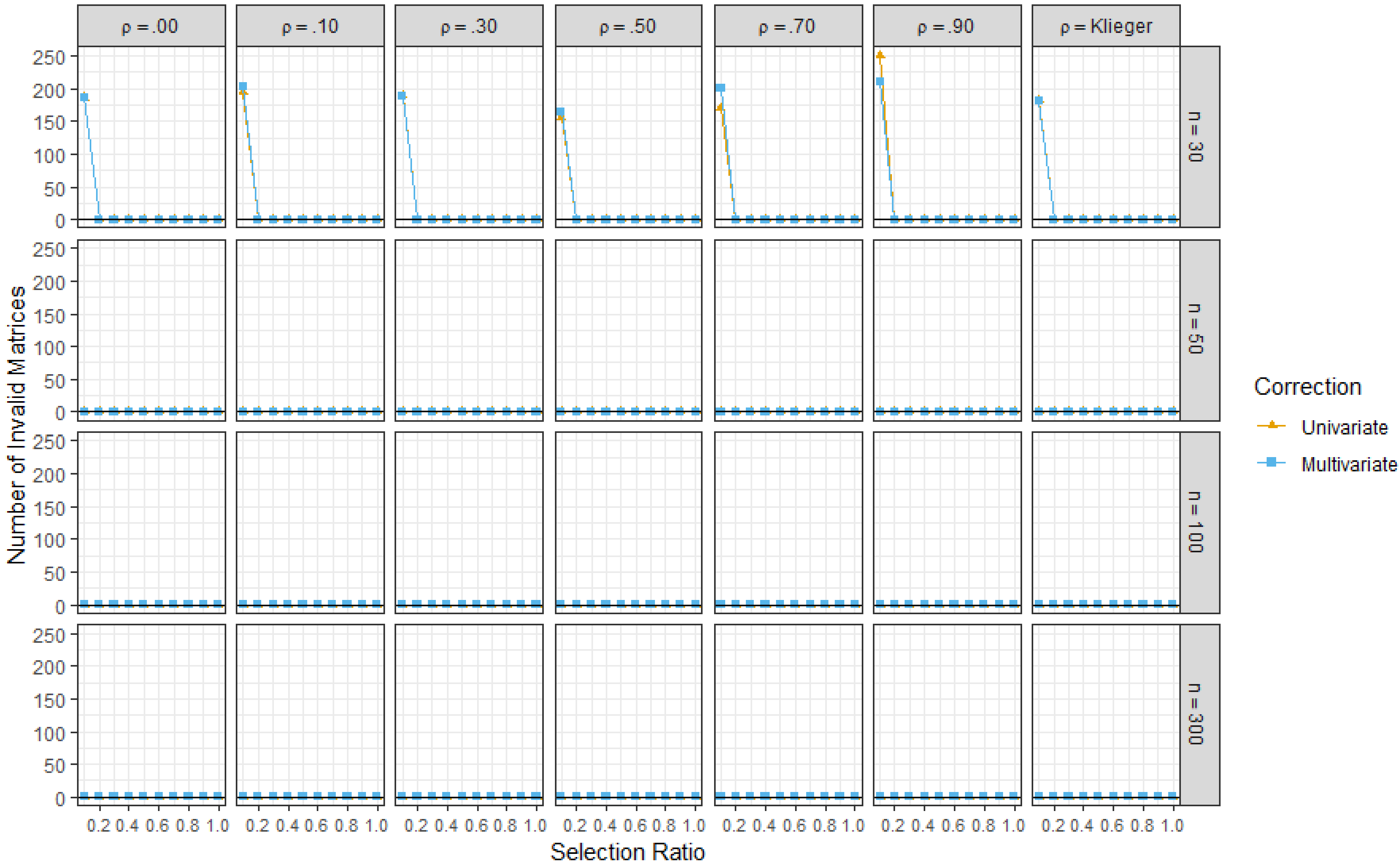

Figure 10. Count of matrices violating Hubert's inequality by population correlation, ρ; sample size, N; and correction method.

Research Question 3: Non-positive Definite Matrices

How frequently do univariate, element-wise corrections for selection result in a non-positive definite correlation matrix compared to a multivariate, simultaneous correction when applied to a matrix of correlations? As with the inadmissible values, the corrections only produced non-positive definite matrices when $N = 30$ and the selection ratio was $.10$. A notable difference, however, is that more than half of the matrices produced by each correction were non-positive definite. The univariate corrections produced non-positive definite matrices more frequently than did the multivariate correction, typically around 10% more often. So, although the values within the matrix did not violate Hubert's inequality, there were still problems across the matrix. Hubert's inequality is assessed by examining all of the submatrices of a correlation matrix. The determinant, however, is an assessment of the entire matrix. This may explain the difference in the results. This criterion, then, suggests that neither correction method is reliable under the extreme $.10$ selection ratio and small unrestricted sample $N = 30$.

Table 5

Number of Non-Positive Definite Matrices by Condition.

ρ	Univariate	Multivariate
.00	723	606
.10	715	575
.30	704	588
.50	700	597
.70	729	599
.90	694	611
Klieger	703	590

Note. $N = 30$, selection ratio = $.10$.

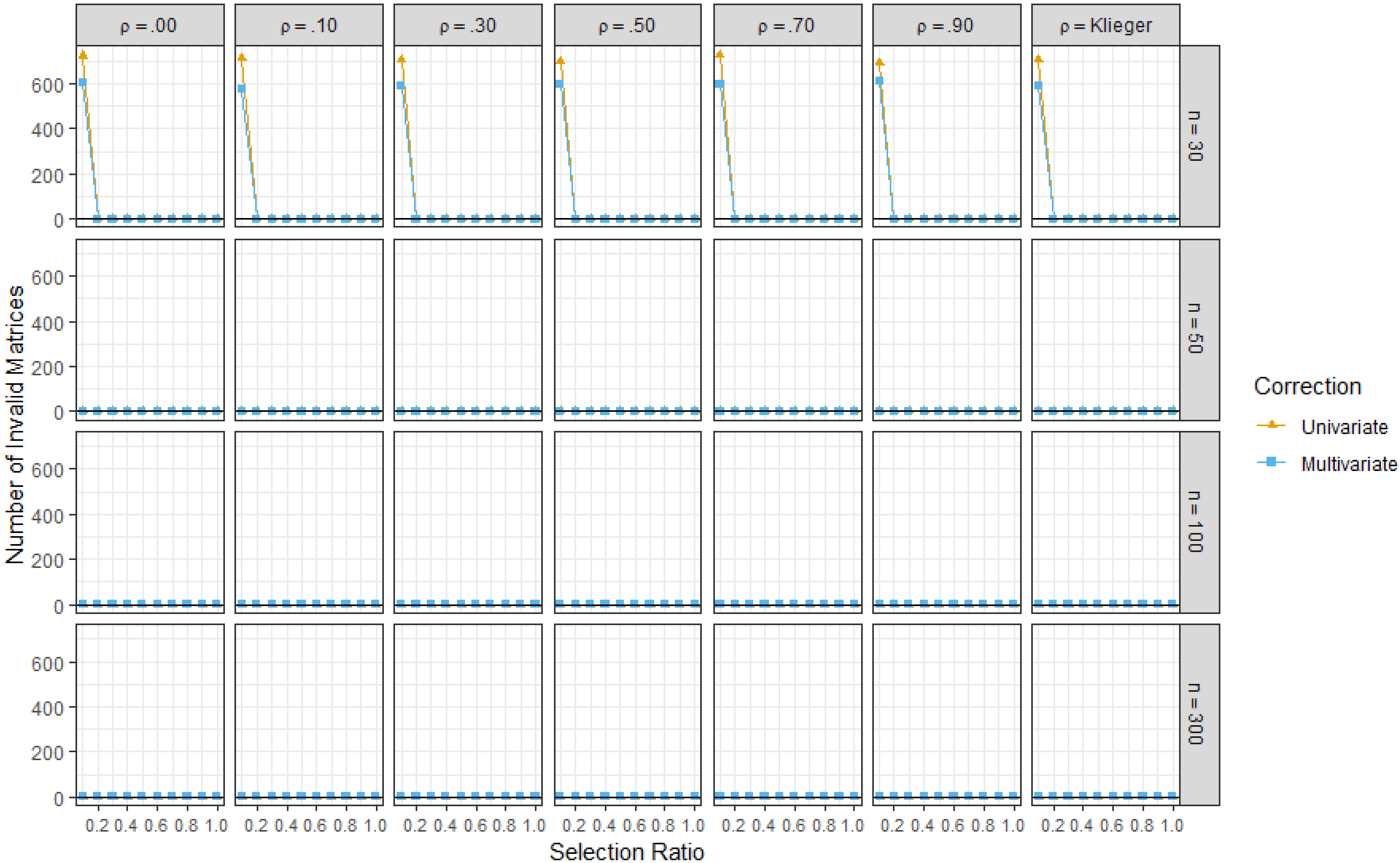

Figure 11. Count of corrected matrices with a determinant less than zero by population correlation, ρ; sample size, N; and correction method.

Research Question 4: Robust Variance Estimation

What are the coverage rates of the confidence intervals produced by robust variance estimation when pooling correlations corrected for selection effects? The motivation behind this research question was the fact that in a multivariate setting, standard error estimates are too narrow if we ignore the correlation between the corrected correlations (Becker, 1995; Raudenbush et al., 1988; Shadish, 1996). We have no closed form estimator of the covariance between two correlations corrected for selection, but Hedges et al. (2010) proposed robust variance estimator that will theoretically correctly model the covariance structure without the meta-analyst having to specify it. Instead, estimates are made using the empirical distribution of the data entered into the model.

Following similar evaluations, I examined the performance of RVE by the coverage rates of the confidence intervals it produces (Hedges et al., 2010; Hong et al., 2017). Coverage rates close to the expected value of 95% serve as one indicator that the standard errors of the estimates are appropriate. If the procedure works well, it obviates any immediate need for deriving the full asymptotic distribution of corrected correlations. In a situation like the one in this study, there are covariances between two correlations affected by explicit selection, covariances between two variables affected by incidental selection, and between two variables—one affected by explicit and one affected by incidental selection. The covariances of each of these pairings would take a different form and be quite complex (e.g. Wei & Higgins, 2012). Figures 12–15 show the coverage rates under the study conditions. Due to the number of invalid matrices produced by the correction methods at a selection ratio of $.10$, I do not present those results here.

I hypothesized that coverage rates would be close to their nominal value. Overall, the results were mixed but positive for situations that seem most likely to occur in practice. Generally, as the unrestricted sample size and selection ratio increased, the performance of the confidence intervals did as well. The estimates produced by the RVE method tended to have more skewed

distributions as the selection ratio decreased. This is the likely cause of their declining accuracy as the recommended confidence intervals are based on the t distribution (Hedges et al., 2010). Histograms of the estimates are available in Appendix G. There was no clear relationship between τ^2 and the coverage rate in any scenario. When the number of studies, k, was 10, the coverage rates were nearly always between 92% and 97% in all conditions. Under this condition, however, there were no consistent patterns in the rates as the selection ratio or sample size changed. The coverage rates fluctuated up and down over those factors.

As the number of studies increased, the coverage rates of the confidence intervals at low selection rates dropped dramatically and did so more steeply with decreasing sample size. This is most easily seen in Figure 12 looking across each row within each τ^2 condition. As the number of studies increased from 10 to 100, the coverage rates showed stronger relationships with sample size and selection ratio. Holding everything else constant, coverage rates improved as sample size or selection ratio improved. The number of studies, however, seems to have the greatest influence on the coverage rates.

Figure 16 is a heat map that shows the confidence interval coverage in all conditions with $\tau^2 = \frac{2}{3}v$. Recall that the value of τ^2 had no apparent effect on the results, so the heat map is representative of all conditions in the simulation. Importantly, the coverage rates in all conditions in the unrestricted sample were close to nominal, with only one instance of a coverage rate of 92% and one instance of 93% each never exceeding 97%. This suggests that the procedure works well for uncorrected correlations representing a new, though expected, finding for the literature since I have found no methodological studies of RVE using correlations. Coverage rates were nominal when $N = 300$ in selection ratios as low as $.40$. Given the fact that N represents the unrestricted sample size, this is a positive finding as most applicant pools, or unrestricted samples, seem likely to be larger than 300.

The coverage rates dropped as low as 92% in selection ratios $.90$–$.70$ and at a selection ratio of $.60$, $N = 50$, and $k = 100$ dropped to 88% in a single correlation, r_{12}, but with all others to this

point being above 90%. When the selection ratio was $.60$ there were 16 out of 162 conditions that produced coverage rates between 88% and 92.9%. All but one occurred with $k = 100$ and seven occurred on correlation r_{12}. At a selection ratio of $.50$, 12 conditions produced coverage rates less than 90%.

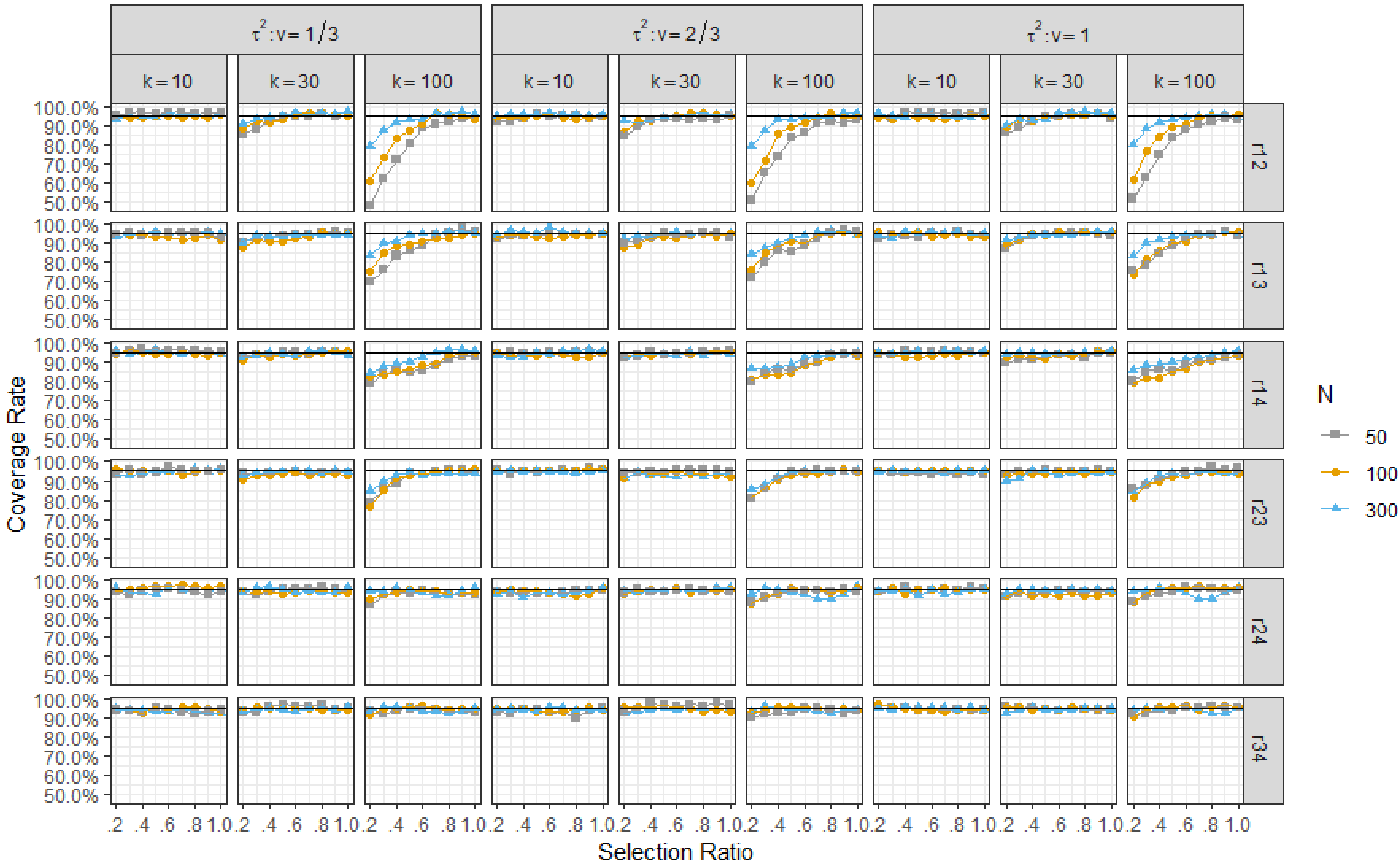

Figure 12. Confidence interval coverage by τ^2:v ratio; number of studies, k; and sample size, N.

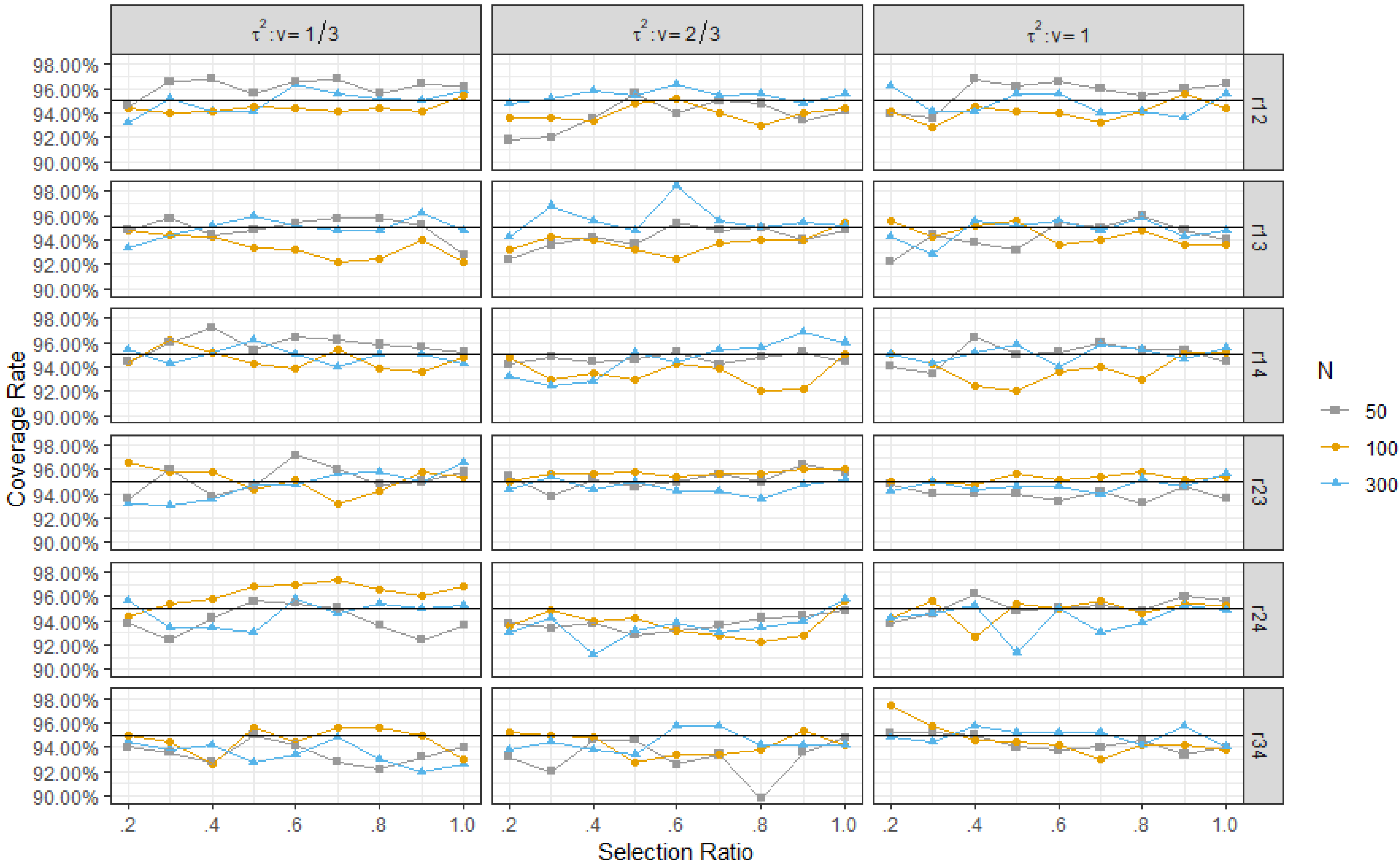

Figure 13. Confidence interval coverage by τ^2:v ratio and sample sample size, N; k = 10.

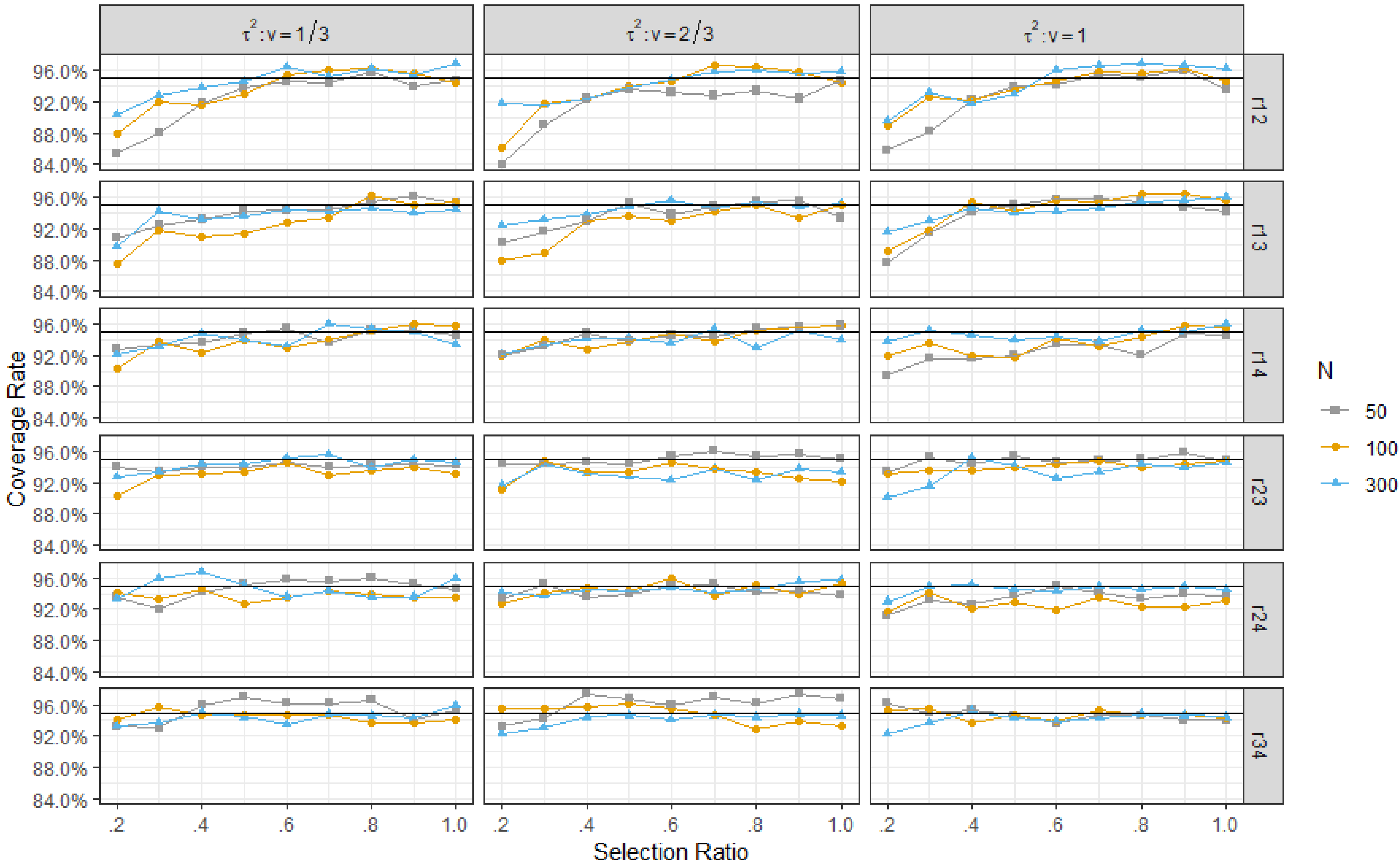

Figure 14. Confidence interval coverage by τ^2:v ratio and sample sample size, N; k = 30.

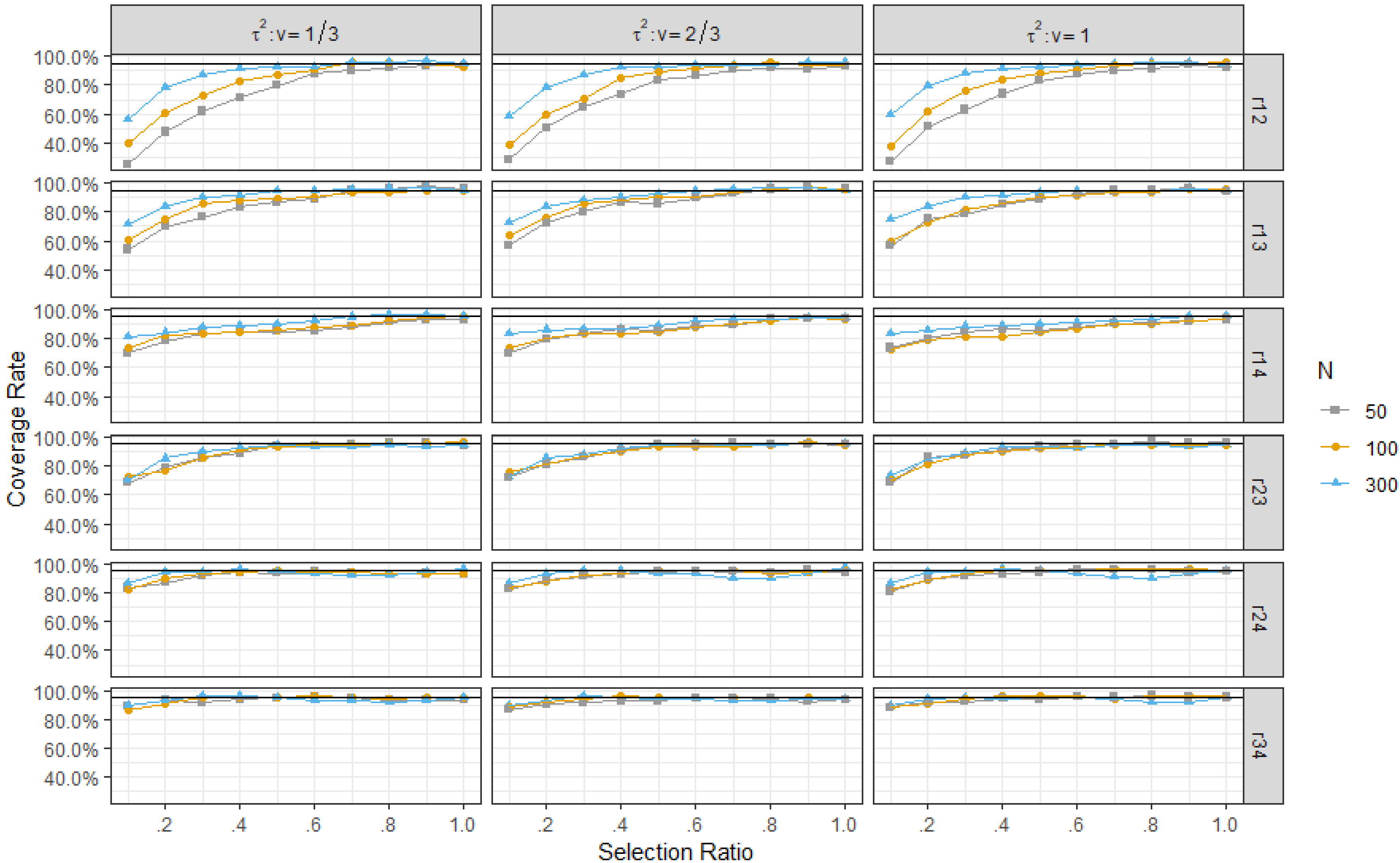

Figure 15. Confidence interval coverage by τ^2:v ratio and sample sample size, N; k = 100.

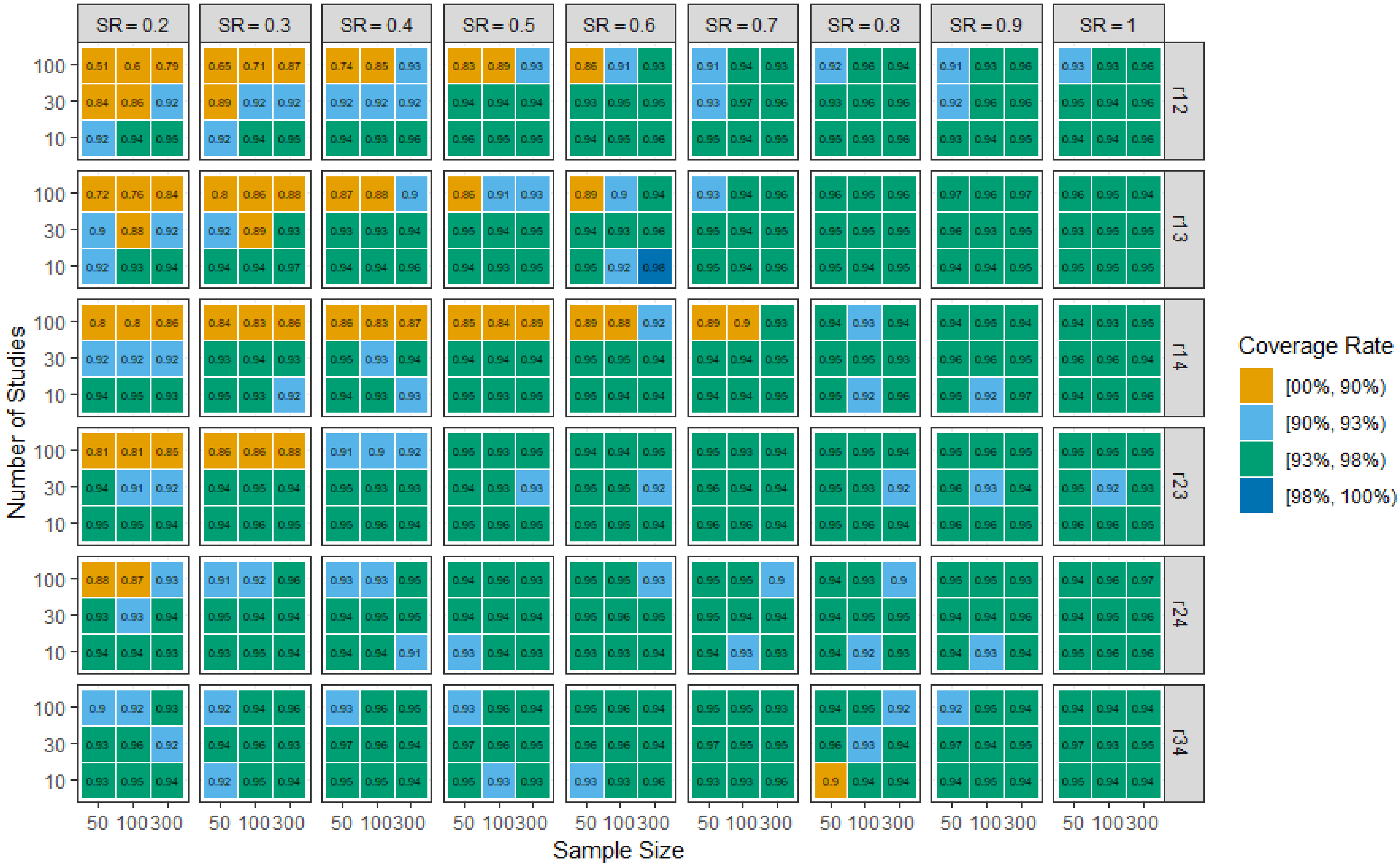

Figure 16. Confidence interval coverage for each correlation by sample size, N; number of studies, k; and selection ratio, SR; $\tau^2 = \frac{2}{3}v$. Actual rate inside each cell.

CHAPTER 5: DISCUSSION

I started this research with a broad motivation—to explore how we might apply the artifact corrections from psychometric meta-analysis to a multivariate meta-analysis. I chose to focus on the particular artifact of selection bias, which has applications in any research domain and has particular applications to personnel selection where Schmidt and Hunter developed psychometric meta-analysis and, in education, research on admissions processes, community college placement, and teacher selection. Methods for integrating artifact corrections with multivariate meta-analysis are not in the psychometric meta-analysis literature, which is wholly univariate (Schmidt & Hunter, 2015). Nor can they be found in the generalized meta-analysis literature where methods of multivariate meta-analysis have received a great deal of attention but without serious consideration of correlations corrected for artifacts (Becker, 1992; Becker & Wu, 2007; M. W.-L. Cheung, 2008; Gleser & Olkin, 2009; Hafdahl, 2008; Hedges et al., 2010; Raudenbush et al., 1988). I found myself engaged with two bodies of literature, both of which have made significant contributions to meta-analysis but lacked integration. Artifact corrections improve effect size estimates by removing the noise inherent in our measurements (Schmidt & Hunter, 2015). Multivariate meta-analysis led to model-based meta-analysis, providing a tool to test theory by drawing on the information contained in an entire body of literature (Becker, 1992; Raudenbush et al., 1988). These two contributions, however, have yet to be fully integrated in any rigorous sense.

I argued that it is not appropriate to assume that we can apply the univariate methods that dominate psychometric meta-analysis to the multivariate meta-analysis context (Hafdahl, 2008) and examined the two stages at which such assumptions might break down: (1) correcting the correlations and (2) pooling the corrected correlations. Applying the univariate corrections in psychometric meta-analysis to correlation matrices is theoretically inappropriate because doing so requires that one manipulates the elements of the correlation matrix independently although they

are related (Steiger, 1980). These relationships limit the possible values that can exist in a valid correlation matrix (Hubert, 1972; Olkin, 1981; Stanley & Wang, 1969), and the univariate corrections do not necessarily respect those limits. At the pooling stage, the corrected correlations need to be combined, but in a multivariate context, they are not independent of one another. Instead, the correlations that arise from the same sample are themselves correlated (Steiger, 1980). This correlation among the correlations should be modeled to accurately estimate the corrected correlations' variance and standard error (Raudenbush et al., 1988; Shadish, 1996). No one has yet derived this correlation structure—the variance-covariance matrix—of correlations corrected for selection effects through an exact estimator, but their variances have been (N. L. Allen & Dunbar, 1990; Bobko & Rieck, 1980; Raju & Brand, 2003). The RVE method provided an attractive alternative.

Implications for Meta-analysis Literature

The results of this study show that under the conditions I studied both the univariate and multivariate corrections provide equivalent results in terms of bias in their estimates. This is a welcome finding for the meta-analyses that have already been completed that apply univariate corrections to a multivariate context (e.g. Credé & Kuncel, 2008; Kuncel et al., 2001; Westrick et al., 2015). My results differ from those of Held and Foley (1994) who found the multivariate corrections to be more accurate. Theirs is the only study I located similar to my work here, but they did not systematically study the full parameter space of correlations and they used empirical data. In small unrestricted samples ($N = 30$) under an extreme selection ratio of $.10$, neither method performed well in recovering a positive definite matrix of corrected correlations, but the univariate corrections produced 10% more invalid matrices than did the multivariate. This is the only criterion on which they differ meaningfully.

I showed that while it performed well under many conditions, RVE did not always perform

optimally unless unrestricted sample sizes were at least 300. Recalling that the unrestricted sample size represents what would be a target population of inference or an applicant pool, groups as small as 300 seem unlikely in practice, but caution is warranted in such cases. The application of RVE to correlations corrected for selection bias opens up the possibility that correlations corrected for other artifacts can be synthesized in this same manner. I also showed for the first time that the method performs well on correlations not subject to selection effects (i.e., in the unrestricted samples).

Delimitations and Future Research

As the first study to examine pooling correlation matrices corrected for selection bias, I chose to delimit this research to establish some baseline understanding of the problem and provide a basis for future studies. Thus, there is a great deal of work to be done to fully explore this problem. Different simulation conditions, different models, and additional artifacts must all be considered.

Selection mechanisms can be very complex. I chose to simulate a single selection variable to examine the effects on an entire correlation matrix. In practice, this would be similar to using a single test to select students from an applicant pool or any single criterion to select a group of study participants. Deliberate selection decisions may be more complex, however. For example, applicant interviews or writing samples can be used in conjunction with cognitive tests to select students for admission into a program (e.g. Westrick et al., 2015; Zwick & Himelfarb, 2011). Some institutions are already using the College Board's Environmental Context Dashboard in conjunction with SAT scores to make selection decisions (Jaschik, 2019). I did not examine multiple selection criteria operating on the same sample in order to first address the simplest case. Lawley's multivariate correction is specifically called for in these instances (Sackett & Yang, 2000), but I have not found studies that examine the conditions under which the univariate

corrections might perform reasonably well. Given the fact that the univariate corrections require less information than the multivariate correction, such research would be worthwhile.

In practice, the within-study sample sizes and selection ratios represented in a meta-analysis sample will not be the same for all studies (Hafdahl, 2008). Additional simulations that vary these conditions over the studies may yield more information about how these conditions interact to influence the results of the corrections and effect pooling. Also, not every study in a meta-analysis will include all of the effects of interest either because they do not report the information or the effect was not measured at all (Becker, 2009). The variability over studies in the availability of information creates a missing data problem. I explained in Chapter 2 that one of the advantages of a multivariate meta-analysis is the ability to borrow information across studies when some studies did not report some effects of interest. Hong et al. (2017) claimed that the RVE method by Hedges et al. (2010) does not permit such borrowing, but they do not cite research or expand upon the point nor have I found any other authors making that claim. That no such sharing of information across studies occurs is not immediately evident, however. The estimation method operates in a fashion similar to GLS estimation, which does permit such sharing (Gleser & Olkin, 2009). Hedges et al. (2010) simulated conditions where the number of effects per study varied, but they did not study multiple outcomes in their simulation as I did here. Studies of the Hedges et al. (2010) and Hong et al. (2017) models under these conditions ought to be completed.

S. F. Cheung et al. (2018) proposed a method for synthesizing dependent correlations attenuated due to measurement error or dichotomization based on what they refer to as a sample-wise procedure for accommodating dependence (S. F. Cheung & Chan, 2004, 2008). The coverage rates of the confidence intervals produced by their procedure were close to nominal under all conditions with no apparent relationship to other study conditions. The independence of coverage rates from other study conditions is a desirable trait that I did not observe with the RVE method. Examining their method to incorporate selection effects and to evaluate the two methods

side-by-side would be worthwhile.

Under the conditions where the observed coverage rates of the confidence intervals were poor, more research should be done to examine whether better estimates of the standard errors would improve performance or whether distributions other than the t should be used to construct the confidence intervals. It is also possible that converting the correlations to Fisher's z and then performing the RVE correction would suffice. The Fisher's z transformation normalizes the distribution of the correlations (Hedges & Olkin, 1985) and so may improve confidence intervals constructed from the t distributions.

As noted in Chapter 3, there are hierarchical models of meta-analysis accommodated by RVE (Fisher & Tipton, 2015). These models are appropriate when, for example, several studies in a meta-analysis sample come from the same lab or authors (Fisher & Tipton, 2015). I chose not to investigate that model, instead focusing on the simpler case. Extracting multiple correlations from the same study implies the correlation model described by Fisher and Tipton (2015), but it is conceivable that a hierarchical model could also be applied in practice, providing another opportunity for additional research.

Finally, Schmidt and Hunter (2015) identified 11 artifacts that bias correlation coefficients. Selection effects are only one among them. More research is necessary to evaluate the other corrections in the multivariate setting and, indeed, to investigate correlations corrected for multiple artifacts. The number of conditions to study is quite large, and it will take a good deal of research to untangle how correlations corrected for several artifacts might be appropriately synthesized in a multivariate context. Schmidt and Hunter (2015, p. 92) state that measurement error and selection effects are among the most well-studied artifacts in psychometric literature providing some indication that they present the most common problems in applied research.

Recommendations for Practice

With deliminations in mind, it is helpful to condense the results of this work into a series of practical guidelines. It is essential to keep in mind that my recommendations for practice apply to correcting more than one correlation drawn from the same sample when selection took place on a single variable as those were the conditions of the study. I first discuss what considerations to make at the correction stage, then move to the findings of the RVE simulation.

Correcting for selection effects. The first stage of a meta-analysis of correlations is to correct for the biasing effects of artifacts such as selection effects. Typically, we would look to the methods outlined in Schmidt and Hunter (2015). However, if we are extracting multiple correlations from the same sample, we need to consider how to approach the corrections. The primary driving factor in practice will be the type of selection process at work and otherwise is most likely to be the unrestricted sample size. So we need this information to make an informed decision.

In general, no correction method works well when the selection ratio is $.10$ and the unrestricted $N = 30$ not only due to the bias but because of their inability to produce positive-definite matrices. Observed correlations should be used in these cases with the understanding that they are still biased estimates. The unrestricted sample refers to the reference population. So, for example, *unrestricted sample* can also be read *applicant pool*. I have shown that if an applicant pool is at least 300 strong, the corrections will always be better estimates than the observed correlations (see Figure 9). It seems likely that in most research domains, the unrestricted sample, or even the population if that data is available (e.g., national data on SAT scores), will exceed 300 members.

On the other hand, selection ratios as extreme as 10% might also be rare. For perspective, according to the National Center for Education Statistics (2017), in 2016-17, only 0.5% of 4-year

colleges and universities in the United States admitted fewer than 10% of their applicants. Instead, 79.7% admitted at least 50.0% of their applicants and an additional 17.7% admitted between 25.0% and 49.9%.

Explicit selection. Neither method of correction offers an advantage when explicit selection affects the correlations. My research suggests that corrections not be used at selection ratios less than $.30$ when the unrestricted sample is less than 100. In an unrestricted sample of at least 100, the corrections were always better estimates of their population values except when the population correlation was $.00$ and the selection ratio was $.10$ or, sometimes, $.20$. If the population correlation is $.00$, there should be no need to make any corrections.

Correcting for incidental selection. If any variable has undergone selection in the matrix, then any correlations that exclude that variable have undergone indirect selection. The corrections for indirect range restriction were also equivalent between the two methods. Bias in the correction only needs to be a concern with selection ratios less than $.30$ with unrestricted samples less than 100, where the observed correlations were less biased than their corrected values. In addition, as the population values of correlations subject to explicit selection increased, the accuracy of corrections for incidental selection also increased for constant selection ratio. We do not have access to the actual population parameters of the correlations, but if explicit corrections are made and the population correlation is likely to be around $.30$ or higher, with unrestricted $N \geq 50$ then accurate corrections can be made for selection ratios as low as $.20$.

Evaluating the resulting matrix. When we manipulate a matrix, we should check that the result is valid. In this study, invalid matrices only occurred when the selection ratio was $.10$ and $N = 30$, a time when I recommend that observed correlations be used. Still, I offer some advice on the problem. If we try to use an invalid matrix as input for a path analysis or SEM study, it will produce an error because the matrix cannot be inverted. If, however, we are not using the

matrix in a subsequent analysis the error may not be apparent, and the resulting matrix should be examined for problems because, bottom line, it is not a correlation matrix at all if it is non-positive definite! The easiest method is to check the determinant of the matrix to determine that it is greater than zero. From the results of this study, this criterion is more conservative than evaluating Hubert's inequality. It is also easier to calculate. The `det` function in the base version of R (R Core Team, 2018) where the most commonly used meta-analysis packages have been implemented can be easily used to check the determinant of a matrix. R is free. If we find that the matrix is non-positive definite and, therefore, invalid, then we can turn to Wothke (1993) who provides a discussion of potential remedies. This need should not arise, however, if the results of my research here are heeded.

Additional considerations. Since the two methods of correction are equivalent under the conditions studied here, one might ask whether there are situations when using one or the other is advantageous. The multivariate correction requires the covariances in the unrestricted sample or population among all of the variables used for selection and the observed covariances of all the variables that make up the correlation matrix. Ideally, then, each primary study in the meta-analysis sample would report a covariance matrix. This is rare, however (Hunter et al., 2006). When a covariance matrix is not reported, we can calculate one if we have the standard deviations of the variables that contribute to the correlations we are synthesizing. These standard deviations are also necessary to apply univariate corrections to all of the correlations of interest. To get a covariance matrix, $\mathbf{\Sigma}$, we would arrange the standard deviations into a diagonal matrix, $\mathbf{D}$, then pre- and post-multiply[1] it by the correlation matrix, $\mathbf{R}$, from the study: $\mathbf{\Sigma} = \mathbf{DRD}$. This can be done in any program that can handle matrix algebra, including R.

We would be justified in asking why we should ever go through that kind of trouble. The primary advantage of the multivariate correction in terms of its implementation is that we need not

[1]For readers not familiar with matrix multiplication, the order of the terms must be preserved. For example, $\mathbf{DRD} \neq \mathbf{DDR}$.

know the exact nature of the selection mechanisms at work (Lawley, 1943). In other words, we do not run the risk of applying the wrong correction formula as long as we can identify the variables on which selection occurs. Research has shown that applying the incorrect univariate formula can introduce serious bias in the estimates (Linn, 1968; Sackett & Yang, 2000; Schmidt & Hunter, 2015). In addition, when there are multiple selection variables (e.g., admitting based on GRE and interviews) that are used in a simultaneous selection, the multivariate correction can be used. If selection occurs in stages, we can apply the multivariate correction to each successive group. For example, if GRE scores are used to narrow a pool of applicants first, then the smaller groups if further narrowed using interviews, the correction can be applied to each group in succession (Sackett & Yang, 2000).

If we are sure of the selection mechanisms at work, then the univariate corrections are more easily applied. They do not require the covariances among all variables, which provides an advantage over the multivariate correction. In particular, using Thorndike's Case III formula to correct for incidental selection as shown in Equation A3 only requires information about the standard deviation of the selection variable in both the restricted and unrestricted samples. On the other hand, using Equation A4 requires no knowledge of the standard deviation of the selection variable in the restricted and unrestricted samples. Instead, it uses the standard deviations of the two correlated variables. The multivariate correction requires knowledge of the standard deviations of all three variables—the selection variable, and the two correlated variables, in both the restricted and unrestricted samples.

Pooling the correlations. With valid, corrected correlation matrices from each study in the meta-analysis sample, we can move to pooling them. The RVE method described by Hedges et al. (2010) produced mixed results but may be the best alternative for modeling the covariances of corrected correlations. I can offer some general guidelines. Recall that among the delimitations of this research is that I simulated constant selection ratios and sample sizes across studies within the

meta-analysis. It is unclear how varying those conditions might change the results of this research so it is best to evaluate the typical selection ratio and average within-study sample size present in the meta-analysis sample and make a judgement about how to interpret the results of the RVE analysis based on those values until more research is completed.

Figure 16 can be used to help guide decisions about how to interpret confidence intervals and hypothesis tests concerning correlations corrected for selection. When selection ratios are $.70$ or higher, the confidence intervals have near-nominal coverage in all conditions. In a selection ratio of $.60$ and larger numbers of studies, the coverage rates began to approach 90%. At that point, the performance of the confidence intervals tends to degrade more rapidly with coverage rates more frequently falling from 70%–89% with performance being poorer as the number of studies increased. Notably, in selection ratios as low as $.40$ and within-study sample size of $N = 300$, the performance was close to nominal. At a selection ratio of $.60$, within-study sample sizes as low as $N = 100$ performed well in all conditions. Thus, large within-study sample sizes will be required to offset the poor performance at lower selection ratios. Recall for perspective that 79.7% of 4-year institutions in the United States admit at least 50% of their applicants (National Center for Education Statistics, 2017).

At lower selection ratios, assume that the standard errors are underestimated and that Type I errors are more likely to occur in any significance testing. It is not clear from this research whether or not the Type I error rates are better than simply conducting a univariate analysis. More research will be necessary. From a purely pragmatic standpoint, there is no research to suggest that there is harm in implementing it. It is straightforward using the *robumeta* (Fisher et al., 2017) package in R, which also includes an adjustment for small samples. Fisher and Tipton (2015) provide a complete description and tutorial. The lack of distributional assumptions and theoretical robustness to the choice of weights makes RVE a very straightforward method that is likely to be at least as good as any other method available at this point. As I discussed in the prior section, S. F. Cheung et al. (2018) provide a different approach for correlations corrected for other artifacts that does

not incorporate selection effects as yet. Alternatively, it will be worth exploring the degree to which a series of univariate meta-analyses differs from a multivariate meta-analysis using the RVE framework, but that research has not been done.

Applications

Methodological research may, at times, seem removed from empirical research, so it is helpful for the methodologist to connect the results to research practice explicitly. My research here concerns the application of corrections for selection bias to multivariate meta-analysis. The potential for selection bias is present in any research domain. Recruitment for an intervention can introduce selection bias through participant self-selection (e.g. Kriegbaum et al., 2018; Robbins et al., 2004). When participants drop out of a study, that attrition can also introduce selection bias that ought to be considered (e.g. Powers, 2004). Synthesists should evaluate the studies they sample in their meta-analyses for potential selection effects and make efforts to correct for the bias they introduce.

A domain in educational research where selection effects are inevitable is in the study of the predictive validity of admissions criteria in colleges and universities and in placement tests used at the community college level. Postsecondary institutions have an interest in selecting those students who are most likely to be successful. The traditional centrality of tests of cognitive ability has come under fire because of their tendency to admit higher proportions of Asian and White students (Credé & Kuncel, 2008; Kuncel et al., 2001; Sackett et al., 2009; Sackett et al., 2001; Schmitt et al., 2009). In the face of legal challenges to affirmative action, more colleges and universities are adopting race-neutral admissions policies, but the proportion of Black and Latino students has generally declined (Glasener, Martell, & Posselt, 2019). Thus, the search for effective, multiple measures by which to admit students is ongoing.

As discussed in Chapter 1, while this study was being planned and carried out, the College

Board announced that it will begin to assign an adversity score to accompany SAT scores (Hartocollis, 2019). No research has yet been published on this new score, but how might my research be applied? To simplify the example, assume for the moment that only College Board data will be used to make admissions decisions. First, the predictive validity of the combined measures will be studied, perhaps at individual institutions. Institutions will have the option to use the SAT scores alone or together with the Environmental Context Dashboard (ECD) index to make a selection. Assume that undergraduate GPA is the criterion by which success will be measured. If institutions use the SAT scores alone to select students but still wish to examine the ECD scores, then SAT scores are explicitly selected while ECD scores are incidentally selected. Based on my research, either correction method can be used and the RVE method of synthesis will perform well in moderate-to-large selection ratios and unrestricted samples of at least 300.

As the number of studies of predictive validity grows and as the institutions studied become more heterogeneous, a larger scale meta-analysis can be used to synthesize the data across institutions. This will shed light on whether the new ECD information along with other admission criteria is a good predictor of academic success. This, in turn, will provide evidence to evaluate both the promises and concerns about the ECD. For example, some see the additional information as an opportunity to enhance diversity among their student body (Hartocollis, 2019; Jaschik, 2019), while others have criticized the College Board for potentially putting students who come from more affluent backgrounds at a disadvantage (Hartocollis, 2019; Jaschik, 2019; Wexler, 2019). Wexler (2019) argues that such a score still does not make up for a lack of preparation and the need to improve K-12 schools and, so, the score may not provide much of an advantage after all. This stance, though, requires that the scores be used in simultaneous rather than compensatory selection. Recall that with compensatory selection, a higher ECD score would make up for a lower SAT score, which appears to be how the pilot institutions have used it so far. Others have criticized the College Board for attempting to sum up the life experiences of its test takers with a single score (Paris, 2019). Ultimately, studies of predictive validity of this and other

admissions criteria and a meta-analysis of their results will help to clarify the relationship between the ECD and academic achievement.

In a related strand of research, the predictive validity of the various methods of placement is a domain in which my research can be applied. Waschull (2018) discusses some findings from the Florida College System's (FCS) implementation of Senate Bill 1720 in 2013, which required all of Florida's state colleges to submit remedial education plans to FCS. Among the requirements is that multiple measures be used to place students. As she notes, very little research has been conducted on the impact of this legislative change. A meta-analysis could be completed that synthesizes the correlations between predictors of academic success that are shared components of each college's developmental education plan and an outcome such as college GPA. In this case, one need not wait for individual studies of each college to be reported. Westrick et al. (2015), discussed earlier, used a similar strategy.

Psychometric meta-analysis and the artifact corrections that form its nexus were developed in the context of personnel selection research. In education, there are strands of research on teacher preparation and selection. Kimbrel (2019) pointed to a lack of attention in the literature to the selection process for hiring high-quality teachers despite a great deal of research on the qualities of effective teachers and their impact on student achievement. Indeed, my search in ERIC for articles indexed with a subject thesaurus term of *teacher selection* returned only 261 peer-reviewed works since 2009. The lessons from personnel research and the results of this research could be applied to that domain. In related research, where there is attrition among teachers, the effects of teacher preparation programs expressed as a correlation between the program and a measure of success such as student learning might be improperly estimated as those who do not remain in the profession or at a school do not have outcomes to measure. Their departure from teaching might lead to range restriction or, possibly, range enhancement on observed variables. Thus, if teachers leave due to, say, environmental conditions such as salary their attrition may lead to the belief that a teacher preparation program has a lower relationship with student success than is truly the case.

Selection effects are inevitable in these domains.

In research domains like those I have described, where researchers have an interest in studying the predictive validity of more than one measurement on a population and the available sample is missing data due to selection effects my research is relevant for making corrections to the observed correlations. At the point that a meta-analysis is being considered, my research on pooling the corrected correlations should be considered. As research in these domains accumulates, meta-analytic techniques that accommodate correlations corrected for selection will be necessary. Such methods can help to bring clarity and consensus to research on admissions, placement testing, and teacher preparation and selection, and thus advance our ability to contribute to equitable access and student achievement. More broadly, research like this and the research that follows from it can serve to build a cohesive approach to meta-analysis that integrates the lessons from both psychometric and generalized meta-analysis.

APPENDIX A: UNIVARIATE RANGE RESTRICTION CORRECTIONS

Explicit Selection

Under DRR, there is explicit selection on either x or y, attenuating their correlation, r_{xy}. How one address it will depend upon what is known about the variance in the unrestricted sample. When one knows the unrestricted variance on the selection variable, x, this is Thorndike's Case II. At times, one might not know the unrestricted variance on the selected variable, but does have knowledge of the unrestricted variance of the other variable—y, in this example. This situation describes Thorndike's Case I and is unlikely to occur in practice (Thorndike, 1949). In education, Thorndike's Case I might occur if one selected a sample of students for an intervention based on their GPAs. For some reason, the GPAs of the unselected students become unavailable so that one no longer has access to the unrestricted variance. If the intervention were designed to improve reading scores on a state assessment and one could obtain the unrestricted variance of those scores for all students in the correct population, it could be used to correct the correlation between the GPA and reading score for the study. Again, this situation seems unlikely in practice.

Let S_x be the standard deviation of the unrestricted sample, s_x the standard deviation of the restricted sample, and r_{xy} the observed correlation. Recall that designations of x and y are arbitrary and should not be taken to imply x is a predictor and y an outcome. For Thorndike's Case II, the corrected correlation, r_c is

$$r_c = \frac{(\frac{S_x}{s_x})r_{xy}}{\sqrt{1 + r_{xy}^2(\frac{S_x^2}{s_x^2} - 1)}} \tag{A1}$$

and for Thorndike's Case I,

$$r_c = \sqrt{1 - \frac{s_y^2}{S_y^2}(1 - r_{xy}^2)} \tag{A2}$$

The formula one applies, then, is dependent upon what is known about the unrestricted population. It is evident that r_c depends upon the ratio of the standard deviations and the observed correlation.

Incidental Selection

Under IRR, there is incidental selection on either x or y through a third variable, z. As with DRR, there are different formulas available depending upon what information about the variables is known. If we have measured z and know its unrestricted standard deviation, we can use Equation (A3). If, however, we do not have information about z we can use Equation (A4). We might not know z if we suspect, for example, that student motivation creates self-selection into an intervention but have no measure of the motivation. If we have information about the unrestricted standard deviation on both x and y, we can still correct the correlation between them. We use the $+$ if ρ_{zx} and ρ_{zy} have the same sign and $-$ if they do not. We do not need the exact correlation values, only reasonable information about their signs. Thus, if z is motivation score, x is high school GPA, and y is undergraduate GPA we would use the positive form since it is reasonable that motivation is positively correlated with both x and y. Information about the sign of the correlations can be based on a literature search or normative data from the population if it is available (Sackett & Yang, 2000).

$$r_c = \frac{r_{xy} + r_{zx} r_{zy} (\frac{S_z^2}{s_z^2} - 1)}{\sqrt{1 + r_{zx}^2 (\frac{S_z^2}{s_z^2} - 1)}\sqrt{1 + r_{zy}^2 (\frac{S_z^2}{s_z^2} - 1)}} \tag{A3}$$

$$r_c = r_{xy} (\frac{s_x}{S_x})(\frac{s_y}{S_y}) \pm \sqrt{[1 - (\frac{s_x}{S_x})^2][1 - (\frac{s_y}{S_y})^2]} \tag{A4}$$

APPENDIX B: MULTIVARIATE RANGE RESTRICTION CORRECTION

Lawley's multivariate correction for range restriction requires that we know the unrestriced variances for all selection variables. This, in turn, implies that we must be able to properly identify which variables are used in selection. Beyond this, however, we need no other knowledge of the selection processes (e.g. indirect vs. direct). Borrowing elements of both Birnbaum et al. (1950) and Ree et al. (1994), I present the methods for applying Lawley's correction. To be consistent with the rest of this work, however, I adopt a different set of subscripts. Let the total number of variables be designated p, the number of selection variables q, and the number of incidental variables not subject to selection $p - q$.

Given a p-dimensional random variable, $U = (X_1, X_2, X_3, ..., X_q, Y_{q+1}, Y_{q+2}, Y_{q+3}...Y_p)$, q-dimensional random variables $X = (X_1, X_2, X_3, ...X_q)$, and $p - q$-dimensional random variable $Y = (Y_1, Y_2, Y_3, ..., Y_{p-q})$ the variance-covariance matrix of U, the unrestricted sample, is

$$\mathbf{V}_U = \begin{bmatrix} \mathbf{V}_{qq} & \mathbf{V}_{q,p-q} \\ \mathbf{V}_{p-q,q} & \mathbf{V}_{p-q,p-q} \end{bmatrix} \tag{B1}$$

Let a superscript asterisk denote the same values, but in the restricted sample after selection so that $\mathbf{V}_U^*$ is the variance-covariance of the restricted sample and all of its elements are the same as in $\mathbf{V}_U$, but designated with the superscript asterisk.

The problem we face is that we know all of the elements of $\mathbf{V}_U^*$, but we wish to make inferences for the unrestricted population. Thus, we need $\mathbf{V}_U$. We did not observe the $p - q$ incidental variables in the unrestricted group. Therefore only $\mathbf{V}_{qq}$, the variance-covariance matrix of the selection variables, is known for the unrestricted group. The challenge, then is to estimate the other elements of $\mathbf{V}_U$ using the information we have, namely all of $\mathbf{V}_U^*$ and $\mathbf{V}_{qq}$. Lawley

derived such a solution given the known information, which is

$$\hat{\mathbf{V}}_U = \left[\begin{array}{c|c} \mathbf{V}_{qq} & (\mathbf{V}_{qq})(\mathbf{V}^*_{q})^{-1}(\mathbf{V}^*_{q,p-q}) \\ \hline (\mathbf{V}^*_{q,p-q})(\mathbf{V}^*_{qq})^{-1}(\mathbf{V}_{qq}) & \mathbf{V}^*_{p-q,p-q} - \mathbf{V}^*_{p-q,q}(\mathbf{V}^{-1}_{qq} - \mathbf{V}^{*-1}_{qq}\mathbf{V}_{qq}\mathbf{V}^{*-1}_{qq})\mathbf{V}^*_{q,p-q} \end{array}\right] \tag{B2}$$

Once this solution is obtained, a correlation matrix can be found by $\hat{\mathbf{R}} = \mathbf{D}^{-\frac{1}{2}}\hat{\mathbf{V}}_U\mathbf{D}^{-\frac{1}{2}}$ where $\mathbf{D}$ is a $p \times p$ diagonal matrix with all $\mathbf{D}_{ii} = \hat{\mathbf{V}}_{ii}$ and all $\mathbf{D}_{ij} = 0$ (Sackett & Yang, 2000). Finally, two assumptions underlie this procedure. The regression of incidental variables on selection variables is assumed to be linear and the variance-covariance matrix of the incidental variables is assumed to be independent of the selection variables (Lawley, 1943).

APPENDIX C: DATA GENERATION PROGRAM RESEARCH QUESTIONS 1–3

```r
#==============================================================================#
#                EVALUATING UNIVARIATE VS MULTIVARIATE CORRECTIONS              #
#==============================================================================#

# Program is coded to directly select on x1 producing DRR on all relationships
# with x1 and IRR on all others

# Load libraries ----------------------------------------------------------

library(MASS)
library(psychmeta)
library(data.table)
library(tidyverse)

# Define simulation conditions ---------------------------------------------

# v     = number of variables
# reps  = number of studies (i.e. replications)
# n     = within-study samples size
# Sigma = population variance-covariance matrix
# p     = truncation points 1-(Selection Ratio);small selection ratios with
#         small n may result in constants and undefined correlations

set.seed(093081)

v <- 4
reps <- 1000

n <- list(30, 50, 100, 300)
p <- list(0, .1, .2, .3, .4, .5, .6, .7, .8, .9)
Sigma <- list('.00' = reshape_vec2mat(cov = c(.00, .00, .50, .00, .50, .50)),
              '.10' = reshape_vec2mat(cov = c(.10, .10, .50, .10, .50, .50)),
              '.30' = reshape_vec2mat(cov = c(.30, .30, .50, .30, .50, .50)),
              '.50' = reshape_vec2mat(cov = c(.50, .50, .50, .50, .50, .50)),
              '.70' = reshape_vec2mat(cov = c(.70, .70, .50, .70, .50, .50)),
              '.90' = reshape_vec2mat(cov = c(.90, .90, .50, .90, .50, .50)),
              'Klieger' = reshape_vec2mat(cov = c(.42, .30, .24, .47, .19, .21))
          )

# Define a function to extract the lower triangle of matrices
ltr <- function(x) {x[lower.tri(x)]}

# Data generation ----------------------------------------------------------

x  <- list() # List of generated observations
sd <- list() # List of standard deviatations of observations
r  <- list() # List of correlation matrices of observations
d1 <- list() # Lists of temporary results of each stage
d2 <- list()
d3 <- list()
s1 <- list()
s2 <- list()
s3 <- list()
r1 <- list()
r2 <- list()
r3 <- list()

for (i in seq_along(Sigma)){
  for (j in seq_along(n)){
    for (k in 1:reps){
      d1[k] <- list(mvrnorm(n = n[[j]], mu = rep(0,v), Sigma = Sigma[[i]]))

      for (l in 1:length(p)) {
```

```r
        d2[[l]] <- (matrix(subset(d1[[k]],
                                  d1[[k]][, 1] >= quantile(d1[[k]][, 1],
                                                           probs = p[[l]])),
                           ncol = 4))

        s1[[l]] <- matrix(apply(d2[[l]], 2, sd), 1, 4)

        r1[[l]] <- cor(d2[[l]])

      }
      names(r1) <- (1-unlist(p))
      names(s1) <- (1-unlist(p))
      names(d2) <- (1-unlist(p))
      r2[[k]]   <- r1
      s2[[k]]   <- s1
      d1[[k]]   <- d2
    }
    names(r2) <- 1:reps
    names(s2) <- 1:reps
    names(d1) <- 1:reps
    r3[[j]]   <- r2
    s3[[j]]   <- s2
    d3[[j]]   <- d1
  }
  names(r3) <- unlist(n)
  names(s3) <- unlist(n)
  names(d3) <- unlist(n)
  x[[i]]  <- d3
  sd[[i]] <- s3
  r[[i]]  <- r3
}
names(x)  <- names(Sigma)
names(sd) <- names(Sigma)
names(r)  <- names(Sigma)

# Remove intermediate objects
rm(d1,d2,d3,r1,r2,r3,s1,s2,s3)


# Begin applying corrections ----------------------------------------------

# Define functions for univariate corrections
# drr = Thorndike's Case II
# irr = Thorndike's Case III

drr <- function(Sx, sx, rxy) {
  u <- (Sx / sx)
  (u * rxy) / sqrt(1 + rxy ^ 2 * (u ^ 2 - 1))
}

irr <- function(Sz, sz, rxy, rzx, rzy) {
  u <- (Sz / sz)
  (rxy + rzx * rzy * (u ^ 2 - 1)) /
    (sqrt(1 + rzx ^ 2 * (u ^ 2 - 1)) * sqrt(1 + rzy ^ 2 * (u ^ 2 - 1)))
}

ltri          <- list() # Lower triangles of r
rc1           <- list() # Correlations corrected for DRR and IRR
rc2           <- list()
rc3           <- list()
r.correct.uni <- list() # Vectors of corrected correlations

for (i in seq_along(Sigma)) {
```

```
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        # Extract the lower triangle of each matrix in r
        ltri[[l]]  <-
          matrix(ltr(r[[i]][[j]][[k]][[l]]), 1, 6)

        # Correct full vector of lower triangles for both DRR and IRR
        rc1[[l]] <-
          matrix(c(
            apply(
              matrix(ltri[[l]][, 1:3], 1, 3),
              2,
              drr,
              Sx = sd[[i]][[j]][[k]][[1]][, 1],
              sx = sd[[i]][[j]][[k]][[l]][, 1]
            ),
            irr(
              Sz  = sd[[i]][[j]][[k]][[1]][, 1],
              sz  = sd[[i]][[j]][[k]][[l]][, 1],
              rzx = ltri[[l]][, 1],
              rzy = ltri[[l]][, 2],
              rxy = ltri[[l]][, 4]
            ),
            irr(
              Sz  = sd[[i]][[j]][[k]][[1]][, 1],
              sz  = sd[[i]][[j]][[k]][[l]][, 1],
              rzx = ltri[[l]][, 1],
              rzy = ltri[[l]][, 3],
              rxy = ltri[[l]][, 5]
            ),
            irr(
              Sz  = sd[[i]][[j]][[k]][[1]][, 1],
              sz  = sd[[i]][[j]][[k]][[l]][, 1],
              rzx = ltri[[l]][, 2],
              rzy = ltri[[l]][, 3],
              rxy = ltri[[l]][, 6]
            )
          ), 1, 6)

        # Reshape the vectors back to matrices
        rc1[[l]] <- reshape_vec2mat(cov = rc1[[l]],)
      }
      names(rc1) <- c(1 - unlist(p))
      rc2[[k]] <- rc1
    }
    names(rc2) <- 1:reps
    rc3[[j]] <- rc2
  }
  names(rc3) <- unlist(n)
  r.correct.uni[[i]] <- rc3
}

names(r.correct.uni) <- names(Sigma)

t1 <- list() #Temporary list of lower right triangles
t2 <- list()
t3 <- list()
t4 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
```

```
      for (l in seq_along(p)) {
        t1[[l]] <- cbind(names(Sigma)[i],
                         n[[j]],
                         1-p[[l]],
                         k,
                         t(ltr(r.correct.uni[[i]][[j]][[k]][[l]])))
      }
      t2[[k]] <- do.call(rbind, t1)
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
}

correct.tri.uni <- as.data.frame(do.call(rbind, t4))
correct.tri.uni[, 2:10] <- apply(correct.tri.uni[, 2:10], 2, as.numeric)

names(correct.tri.uni) <-
  c('Sigma', 'n', 'sr', 'rep', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

# Lawley's correction is applied using the psychmeta package

r.correct.mvt  <- list() # Final list of corrected matrices

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        rc1[[l]] <- correct_matrix_mvrr(
          Sigma_i = cov(x[[i]][[j]][[k]][[l]]),
          Sigma_xx_a = matrix(1, 1, 1),
          x_col = 1,
          standardize = TRUE
        )
      }
      names(rc1) <- c(1 - unlist(p))
      rc2[[k]] <- rc1
    }
    names(rc2) <- 1:reps
    rc3[[j]] <- rc2
  }
  names(rc3) <- unlist(n)
  r.correct.mvt[[i]] <- rc3
}

names(r.correct.mvt) <- names(Sigma)

t1 <- list() #Temporary lists of lower right triangles
t2 <- list()
t3 <- list()
t4 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        t1[[l]] <- cbind(names(Sigma)[i],
                       n[[j]],
                       1-p[[l]],
                       k,
                       t(ltr(r.correct.mvt[[i]][[j]][[k]][[l]])))
      }
      t2[[k]] <- do.call(rbind, t1)
```

```r
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
}

correct.tri.mvt <- as.data.frame((do.call(rbind, t4)))
correct.tri.mvt[, 2:10] <- apply(correct.tri.mvt[, 2:10], 2, as.numeric)

names(correct.tri.mvt) <-
  c('Sigma', 'n', 'sr', 'rep', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

summary.correct.mean.uni <- (aggregate(. ~ Sigma + n + sr,
                                       correct.tri.uni,
                                       mean))

summary.correct.mean.mvt <- (aggregate(. ~ Sigma + n + sr,
                                       correct.tri.mvt,
                                       mean))

summary.correct.var.uni <- (aggregate(. ~ Sigma + n + sr,
                                      correct.tri.uni,
                                      var))

summary.correct.var.mvt <- (aggregate(. ~ Sigma + n + sr,
                                      correct.tri.mvt,
                                      var))

# Remove intermediate objects
rm(ltri, rc1, rc2, rc3, t1, t2, t3, t4)

# Calculate the bias in the corrections -----------------------------------

b1       <- list() # Temporary lists of matrix differences
b2       <- list()
b3       <- list()
bias.uni <- list() # List of matrix differences in the univariate corrections
bias.mvt <- list() # List of matrix differences in the multivariate corrections

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        b1[[l]] <-
          (r.correct.uni[[i]][[j]][[k]][[l]] - r[[i]][[j]][[k]][[1]])
      }
      names(b1) <- c(1 - unlist(p))
      b2[[k]] <- b1
    }
    names(b2) <- 1:reps
    b3[[j]] <- b2
  }
  names(b3) <- unlist(n)
  bias.uni[[i]] <- b3
}

names(bias.uni) <- names(Sigma)

t1 <- list()
t2 <- list()
t3 <- list()
t4 <- list()
```

```
for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        t1[[l]] <- cbind(names(Sigma)[i],
                         n[[j]],
                         1-p[[l]],
                         k,
                         t(ltr(bias.uni[[i]][[j]][[k]][[l]])))
      }
      t2[[k]] <- do.call(rbind, t1)
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
}

bias.tri.uni <- as.data.frame(do.call(rbind, t4))
bias.tri.uni[, 2:10] <- apply(bias.tri.uni[, 2:10], 2, as.numeric)

names(bias.tri.uni) <-
  c('Sigma', 'n', 'sr', 'rep', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        b1[[l]] <-
          (r.correct.mvt[[i]][[j]][[k]][[l]] - r[[i]][[j]][[k]][[1]])
      }
      names(b1) <- c(1 - unlist(p))
      b2[[k]] <- b1
    }
    names(b2) <- 1:reps
    b3[[j]] <- b2
  }
  names(b3) <- unlist(n)
  bias.mvt[[i]] <- b3
}

names(bias.mvt) <- names(Sigma)

t1 <- list()
t2 <- list()
t3 <- list()
t4 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        t1[[l]] <- cbind(names(Sigma)[i],
                         n[[j]],
                         1-p[[l]],
                         k,
                         t(ltr(bias.mvt[[i]][[j]][[k]][[l]])))
      }
      t2[[k]] <- do.call(rbind, t1)
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
}
```

```r
bias.tri.mvt <- as.data.frame(do.call(rbind, t4))
bias.tri.mvt[, 2:10] <- apply(bias.tri.mvt[, 2:10], 2, as.numeric)

names(bias.tri.mvt) <-
  c('Sigma', 'n', 'sr', 'rep', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

summary.bias.median.uni <- (aggregate(. ~ Sigma + n + sr, bias.tri.uni, median))
summary.bias.median.mvt <- (aggregate(. ~ Sigma + n + sr, bias.tri.mvt, median))

# Remove intermediate objects
rm(b1, b2, b3, t1, t2, t3, t4)

# Examine correlations for values greater than an abs value of 1 ----------

count.gt1.uni <- list(apply(correct.tri.uni[, 5:10], 2,
                            function(x) {
                              length(which(x < -1 | x > 1))
                            }))

count.gt1.mvt <- list(apply(correct.tri.uni[, 5:10], 2,
                            function(x) {
                              length(which(x < -1 | x > 1))
                            }))

# Examine correlations for violation of Hubert's inequality ---------------

# Define a function that evaluates Hubert's (1972) inequality

hubert <- function(A,G){
  # A is a correlation matrix containg only the correlations whose bounds need
  # to be determined
  # G is a column vector all of all fixed correlations

  value <- t(G) %*% ginv(A) %*% G
  return(value)
}

t1 <- list() # Temporary lists for the following loop
t2 <- list()
t3 <- list()
t4 <- list()
t5 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        for(m in 1:v){
        t1[[m]] <- cbind('Sigma' = names(Sigma)[i],
               'n' = n[[j]],
               'rep' = k,
               'sr'  = 1-p[[l]],
               'v'   = m,
               hubert(A = r.correct.uni[[i]][[j]][[k]][[l]][-m, -m],
                      G = r.correct.uni[[i]][[j]][[k]][[l]][-m, m]))
        }
        t2[[l]] <- do.call(rbind, t1)
      }
      t3[[k]] <- do.call(rbind, t2)
    }
    t4[[j]] <- do.call(rbind, t3)
  }
```

```
  t5[[i]] <- do.call(rbind, t4)
}

# Add results to a data frame

bound.uni <- as.data.frame(do.call(rbind, t5))
bound.uni <- spread(bound.uni, v, V6)
bound.uni[, 2:8] <- apply(bound.uni[, 2:8], 2, as.numeric)
bound.uni$vio <- c(apply((apply(bound.uni[, 5:8],
                                2,
                                function(x){x > 1})),1,sum))

t1 <- list() # Temporary lists for the following loop
t2 <- list()
t3 <- list()
t4 <- list()
t5 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        for(m in 1:v){
          t1[[m]] <- cbind('Sigma' = names(Sigma)[i],
                           'n' = n[[j]],
                           'rep' = k,
                           'sr'  = 1-p[[l]],
                           'v'   = m,
                           hubert(A = r.correct.mvt[[i]][[j]][[k]][[l]][-m, -m],
                                  G = r.correct.mvt[[i]][[j]][[k]][[l]][-m, m]))
        }
        t2[[l]] <- do.call(rbind, t1)
      }
      t3[[k]] <- do.call(rbind, t2)
    }
    t4[[j]] <- do.call(rbind, t3)
  }
  t5[[i]] <- do.call(rbind, t4)
}

# Add results to a data frame

bound.mvt <- as.data.frame(do.call(rbind, t5))
bound.mvt <- spread(bound.mvt, v, V6)
bound.mvt[, 2:8] <- apply(bound.mvt[, 2:8], 2, as.numeric)
bound.mvt$vio <- c(apply((apply(bound.mvt[, 5:8],
                                2,
                                function(x){x > 1})),1,sum))

summary.bound <- cbind.data.frame(aggregate(vio ~ Sigma + n + sr,
                           bound.uni,
                           function(x){length(which(x > 0))}),
                  'mvt' = aggregate(vio ~ Sigma + n + sr,
                             bound.mvt,
                             function(x){length(which(x > 0))})[, 4])

colnames(summary.bound)[4] <- 'uni'

# Remove intermediate objects
rm(t1, t2, t3, t4, t5)

# Evaluate all corrected matrices for determinants <= 0 -------------------
```

```r
t1 <- list() # Temporary lists for the following loop
t2 <- list()
t3 <- list()
t4 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        t1[[l]] <- cbind('Sigma' = names(Sigma)[i],
               'n' = n[[j]],
               'sr'  = 1-p[[l]],
               'rep' = k,
               'uni' = det(r.correct.uni[[i]][[j]][[k]][[l]]),
               'mvt' = det(r.correct.mvt[[i]][[j]][[k]][[l]]))
      }
      t2[[k]] <- do.call(rbind, t1)
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
}

# Add results to a data frame

det <- as.data.frame(do.call(rbind, t4))
det[, 2:6] <- apply(det[2:6], 2, as.numeric)

summary.det <- aggregate(. ~ Sigma + n + sr, det, function(x) {
  length(which(x <= 0))
})

# Remove intermediate objects
rm(t1, t2, t3, t4)

# Supplemental Analysis

# Evaluate underlying matrices after selection and before correction

t1 <- list()
t2 <- list()
t3 <- list()
t4 <- list()
t5 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        for(m in 1:v){
          t1[[m]] <- cbind('Sigma' = names(Sigma)[i],
                           'n' = n[[j]],
                           'rep' = k,
                           'sr'  = 1-p[[l]],
                           'v'   = m,
                           hubert(A = r[[i]][[j]][[k]][[l]][-m, -m],
                                  G = r[[i]][[j]][[k]][[l]][-m, m]))
        }
        t2[[l]] <- do.call(rbind, t1)
      }
      t3[[k]] <- do.call(rbind, t2)
    }
    t4[[j]] <- do.call(rbind, t3)
```

```r
  }
  t5[[i]] <- do.call(rbind, t4)
}

bound.r <- as.data.frame(do.call(rbind, t5))
bound.r <- spread(bound.r, v, V6)
bound.r[, 2:8] <- apply(bound.r[, 2:8], 2, as.numeric)
bound.r$vio <- c(apply((apply(bound.r[, 5:8],
                                2,
                                function(x){x > 1})),1,sum))

summary.bound.r <- aggregate(. ~ Sigma + n + sr,
                             bound.r,
                             function(x) {
                               length(which(x > 0))
                             })

# Bias in empirical r

b1     <- list() # Temporary lists of matrix differences
b2     <- list()
b3     <- list()
bias.r <-list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        b1[[l]] <-
          (r[[i]][[j]][[k]][[l]] - r[[i]][[j]][[k]][[1]])
      }
      names(b1) <- c(1 - unlist(p))
      b2[[k]] <- b1
    }
    names(b2) <- 1:reps
    b3[[j]] <- b2
  }
  names(b3) <- unlist(n)
  bias.r[[i]] <- b3
}

names(bias.r) <- names(Sigma)

t1 <- list()
t2 <- list()
t3 <- list()
t4 <- list()

for (i in seq_along(Sigma)) {
  for (j in seq_along(n)) {
    for (k in 1:reps) {
      for (l in seq_along(p)) {
        t1[[l]] <- cbind(names(Sigma)[i],
                         n[[j]],
                         1-p[[l]],
                         k,
                         t(ltr(bias.r[[i]][[j]][[k]][[l]])))
      }
      t2[[k]] <- do.call(rbind, t1)
    }
    t3[[j]] <- do.call(rbind, t2)
  }
  t4[[i]] <- do.call(rbind, t3)
```

```r
}

bias.tri.r <- as.data.frame(do.call(rbind, t4))
bias.tri.r[, 2:10] <- apply(bias.tri.r[, 2:10], 2, as.numeric)

names(bias.tri.r) <-
  c('Sigma', 'n', 'sr', 'rep', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

summary.bias.median.r <- (aggregate(. ~ Sigma + n + sr, bias.tri.r, median))
summary.bias.mean.r <- (aggregate(. ~ Sigma + n + sr, bias.tri.r, mean))
summary.bias.var.r <- (aggregate(. ~ Sigma + n + sr, bias.tri.r, var))

summary.bias.mean.uni <- (aggregate(. ~ Sigma + n + sr, bias.tri.uni, mean))
summary.bias.mean.mvt <- (aggregate(. ~ Sigma + n + sr, bias.tri.mvt, mean))

# Remove intermediate objects
rm(t1, t2, t3, t4, t5, b1, b2, b3)
```

APPENDIX D: DATA GENERATION PROGRAM RESEARCH QUESTION 4

```
#===============================================================================#
#                    EVALUATING ROBUST VARIANCE ESTIMATION                      #
#===============================================================================#

# Load libraries -----------------------------------------------------------

library(MASS)
library(robumeta)
library(psychmeta)
library(plyr)
library(data.table)
library(tidyverse)

# This does not loop through all conditions. Create these first then comment
# them out. Each subsequent run will be stored in these objects.

#summary.ci      <- NULL
#summary.var     <- NULL
#summary.results <- NULL

# Define functions ---------------------------------------------------------

# Variance of uncorrected correlation
rvar <- function(r, n) {
  (1 - r ^ 2) ^ 2 / n
}

# Variance of corrected correlation
crvar <- function(r, u) {
  1 / (((u ^ 2 - 1) * r ^ 2) + 1)
}

# Extract the lower triangle of matrices
ltr <- function(x) {x[lower.tri(x)]}

# Define simulation conditions ---------------------------------------------

# v       = number of variables
# k       = number of studies
# n       = within-study samples size
# h       = the ratio of between to within-study variance (t^2/v).
#           Set to zero for a fixed-effects model.
# p12-p34 = popoulation correlations
# P       = the vector of population correlations, for convenience
# psi     = population variance-covariance matrix
# p       = truncation points 1-(Selection Ratio)
# reps    = number of repetitions

set.seed(093081)

v    <- 4
k    <- 100
n    <- 300
h    <- 1
reps <- 500

p12 <- .42
p13 <- .30
p14 <- .24
p23 <- .47
p24 <- .19
p34 <- .21
P   <- c(p12, p13, p14, p23, p24, p34)
```

```r
p <- list(0, .1, .2, .3, .4, .5, .6, .7, .8, .9)

# This is the population covariance matrix
psi <- matrix(
  c(
    rvar(p12, n) * h, 0, 0, 0, 0, 0,
    0, rvar(p13, n) * h, 0, 0, 0, 0,
    0, 0, rvar(p14, n) * h, 0, 0, 0,
    0, 0, 0, rvar(p23, n) * h, 0, 0,
    0, 0, 0, 0, rvar(p24, n) * h, 0,
    0, 0, 0, 0, 0, rvar(p34, n) * h
  ),
  6,
  6,
  dimnames = list(
    c("p12", "p13", "p14", "p23", "p24", "p34"),
    c("p12", "p13", "p14", "p23", "p24", "p34")
  )
)

# Data Generation ---------------------------------------------------------

out1 <- list()
out2 <- list()
out  <- NULL

for(z in 1:reps){

  s <- list()

  # Generate study-level correlation matrices based on psi
  i <- 1
  while (i < (k + 1)) {
    c <- reshape_vec2mat(mvrnorm(n = 1, mu = P, Sigma = psi))
    if (det(c) > 0) {
      s[[i]] <- c
      i <- i + 1
    } else {
      i <- max(1, i - 1)
    }
  }
  names(s) <- rep(1:k)

x    <- list() # List of observations and subsets
x.t  <- list() # Temp list for gathering subsets of x
sd.t <- list() # Temp list for gathering sd of x
sd   <- list() # List of sd of x
r.t  <- list() # Temp list gathering correlations of x
r    <- list() # List of observed correlation matrices of x

for (i in 1:k) {
  # Randomly generate multivariate normal observations
  x[i]   <- list(mvrnorm(
    n = n,
    mu = rep(0, v),
    Sigma = s[[i]]
  ))

  # Subset observations by quantiles specified in p
  for (j in 1:length(p)) {
    x.t[[j]] <- (matrix(subset(x[[i]],
                               x[[i]][, 1] >= quantile(x[[i]][, 1],
```

```r
                                                                probs = p[[j]])),
                              ncol = 4))
  }
  x[[i]] <- x.t

  # Compute sd for all vars in x
  for (j in 1:length(p)) {
    sd.t[[j]] <- matrix(apply(x[[i]][[j]], 2, sd), 1, 4)
  }
  names(sd.t) <- c(1 - unlist(p))
  sd[[i]] <- sd.t

  # Compute correlation matrices for all vars in x
  for (j in 1:length(p)) {
    r.t[[j]] <- cor(x[[i]][[j]])
    if (det(r.t[[j]]) <= 0) {
      print("NPD MATRIX")
      break
    }
  }
  names(r.t) <- c(1 - unlist(p))
  r[[i]] <- r.t
}

names(sd) <- rep(1:k)
names(r)  <- rep(1:k)

# Remove intermediate objects
rm(x.t, sd.t, r.t, s)

t <- list()
u <- list()

for(i in 1:k){
  for(j in 1:length(p)){
    t[[j]] <- sd[[i]][[1]]/sd[[i]][[j]]
  }
  names(t) <- c(1 - unlist(p))
  u[[i]] <- t
}
names(u) <- rep(1:k)

# Apply Lawley's correction -----------------------------------------------

t.r.correct   <- list() # Temporary list of corrected matrices
r.correct.mvt <- list() # List of corrected matrices

for (i in 1:length(x)) {
  for (j in 1:length(x[[i]])) {
    t.r.correct[[j]] <- correct_matrix_mvrr(
      Sigma_i = cov(x[[i]][[j]]),
      Sigma_xx_a = matrix(1, 1, 1),
      x_col = 1,
      standardize = TRUE
    )
  }

  names(t.r.correct) <- c(1 - unlist(p))

  r.correct.mvt[[i]] <- t.r.correct
}

names(r.correct.mvt) <- rep(1:k)
```

```r
ct1 <- list()
ct2 <- list()
ct3 <- list()

for (i in seq_along(r.correct.mvt)) {
  for (j in seq_along(r.correct.mvt[[i]])) {
    ct1[[j]] <- cbind(z,
                      h,
                      n,
                      k,
                      1 - p[[j]],
                      t(ltr(r.correct.mvt[[i]][[j]])))
  }
  ct2[[i]] <- do.call(rbind, ct1)
}

ct3[[z]] <- do.call(rbind, ct2)

correct.tri.mvt <- as.data.frame((do.call(rbind, ct3)))

names(correct.tri.mvt) <-
  c('z', 'h', 'n', 'k', 'sr', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34')

# Remove intermediate objects
rm(t.r.correct)

# Calculate the empirical sd of the corrected correlations
r.correct.sd <- list(round(apply(rbindlist(r.correct.mvt), 2, sd), 9))

# Prepare the data for RVE ------------------------------------------------

# Extract the lower right triangle of each corrected correlation matrix

for (i in 1:length(r.correct.mvt)) {
  for (j in 1:length(r.correct.mvt[[i]])) {
    t <-
      t(c(
        'tr' = (1 - (j - 1) / 10),
        matrix(c(ltr(
          r.correct.mvt[[i]][[j]]
        )), 1, 6),
        'u' = u[[i]][[j]][1, 1],
        matrix(c(ltr(r[[i]][[j]])), 1, 6)
      ))
    if (j == 1) {
      t.ltr <- t
    } else {
      t.ltr <- rbind(t.ltr, t)
    }
  }
  d <- data.frame('study' = i, t.ltr)
  if (i == 1) {
    rve.data <- d
  } else {
    rve.data <- rbind(rve.data, d)
  }
}

names(rve.data) <-
  c('study', 'sr', 'r12', 'r13', 'r14', 'r23', 'r24', 'r34', 'u',
    'ro12','ro13','ro14','ro23','ro24','ro34')
```

```r
# Change to long format
rve.data <- reshape(
  rve.data,
  idvar = c('study', 'sr'),
  times = c('r12','r13','r14','r23','r24','r34'),
  direction = 'long',
  varying = list(Observed = c(10:15), Corrected = c(3:8)),
  v.names = c('observed', 'corrected')
)

names(rve.data)[4] <- c('effect')

# Compute the ratio of corrected and observe correlation, necessary to compute
# the variance of corrected correlations
rve.data$ratio <- rve.data$corrected / rve.data$observed
rve.data$effect <- as.factor(rve.data$effect)

# Remove intermediate objects
rm(t, d, t.ltr)

# Calculate the expected standard error of each correlation
# For the unrestricted sample, use the standard variance of r
# Else use the squared standard error of corrected correlations

rve.data <- add_column(rve.data,
                       'vi' = ifelse(
                         rve.data$sr == 1,
                         rvar(rve.data$corrected, n),
                         ifelse(
                           (
                             rvar(rve.data$observed, n) *
                               rve.data$ratio *
                               crvar(r = rve.data$corrected, u = rve.data$u)
                           ) < 0,
                           0,
                           rvar(rve.data$observed, n) *
                             rve.data$ratio *
                             crvar(r = rve.data$corrected, u = rve.data$u)
                         )
                       ))

# Calculate the weight for each correlation, which is the average of all
# within-study variances

rve.data <- ddply(rve.data, c('study', 'sr'), mutate, wi = mean(vi))

# Apply RVE

for (i in 1:length(p)) {
  d <- subset(rve.data, sr == (1 - (i - 1) / 10))

  t.rve <-
    robu(
      formula = corrected ~ -1 + effect,
      var.eff.size = wi,
      studynum = study,
      modelweights = "CORR" ,
      rho = 0.8 ,
      small = TRUE ,
      data = d
    )

  out1[[i]] <- cbind(
```

```r
    'h'    = h,
    'k'    = k,
    'n'    = n,
    'sr'   = (1 - (i - 1) / 10),
    'rep'  = z,
    'corr' = t.rve[["reg_table"]][["labels"]],
    'est'  = t.rve[["reg_table"]][["b.r"]],
    'sde'  = t.rve[["reg_table"]][["SE"]],
    'cil'  = t.rve[["reg_table"]][["CI.L"]],
    'ciu'  = t.rve[["reg_table"]][["CI.U"]]
  )
}

out2[[z]] <- do.call(rbind, out1)

}

out <- as.data.frame(do.call(rbind, out2))

# Add the population correlations to the data
out$pop[out$corr == 1] <- p12
out$pop[out$corr == 2] <- p13
out$pop[out$corr == 3] <- p14
out$pop[out$corr == 4] <- p23
out$pop[out$corr == 5] <- p24
out$pop[out$corr == 6] <- p34

# Calculate whether or not the estimated correlation falls within the C.I.
out$acc[out$pop >= out$cil & out$pop <= out$ciu] <- 1
out$acc[out$pop <= out$cil | out$pop >= out$ciu] <- 0

# Aggregate the number of CIs covering the population correlation by condition
ci <- aggregate(out[c('acc')],
               by = out[c('sr', 'k', 'n', 'corr', 'h')],
               FUN = sum)

# Aggregate the empirical sd by condition
var <- aggregate(correct.tri.mvt[c('r12','r13','r14','r23','r24','r34')],
                 by = correct.tri.mvt[c('sr', 'k', 'h', 'n')],
                 FUN = sd)

# Store each run of the script
summary.ci  <- rbind(summary.ci, ci)
summary.var <- rbind(summary.var, var)
summary.results <- rbind(summary.results, out)
```

APPENDIX E: TABLE OF MEDIAN BIAS OF THE CORRECTION METHODS IN ALL SIMULATION CONDITIONS

Table E1

Median Bias in All Simulation Conditions

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
.00	30	0.1	0.010	0.003	0.012	0.007	**0.223**	**0.223**	**0.065**	**0.073**	**0.285**	**0.290**	**0.282**	**0.282**
		0.2	0.044	0.046	0.048	**0.054**	0.043	0.033	0.023	0.024	**0.085**	**0.086**	**0.111**	**0.115**
		0.3	0.007	0.010	0.004	0.009	0.036	0.034	-0.020	-0.011	0.046	0.041	**0.059**	**0.066**
		0.4	0.028	0.022	0.006	0.009	0.033	0.029	-0.009	-0.004	0.038	0.044	**0.051**	**0.051**
		0.5	0.011	0.015	0.006	0.007	0.012	0.011	0.003	0.001	0.023	0.022	0.024	0.028
		0.6	0.018	0.016	0.005	0.003	0.011	0.008	-0.007	-0.007	0.017	0.015	0.014	0.016
		0.7	-0.003	-0.001	0.002	-0.002	0.003	-0.002	0.002	0.001	0.009	0.008	0.001	0.002
		0.8	0.002	0.002	-0.008	-0.005	-0.003	-0.004	0.001	0.001	0.010	0.010	0.003	0.007
		0.9	0.002	0.001	0.006	0.005	-0.004	-0.006	0.001	0.000	0.004	0.003	0.002	0.005
	50	0.1	**0.154**	**0.157**	-0.005	0.002	**0.099**	**0.086**	-0.020	-0.027	**0.187**	**0.190**	**0.134**	**0.144**
		0.2	0.023	0.027	-0.009	-0.009	**0.054**	**0.060**	-0.006	-0.011	**0.074**	**0.075**	**0.056**	**0.059**
		0.3	0.011	0.011	-0.016	-0.017	0.026	0.023	0.002	0.003	0.040	0.039	0.028	0.030
		0.4	0.025	0.025	0.013	0.010	0.013	0.015	-0.001	-0.001	0.015	0.016	0.022	0.025
		0.5	0.013	0.015	0.011	0.004	0.009	0.005	0.002	0.003	0.016	0.017	0.025	0.026
		0.6	0.015	0.016	-0.006	-0.005	0.010	0.012	0.004	0.005	0.009	0.010	0.010	0.009
		0.7	-0.006	-0.005	-0.005	-0.006	-0.004	0.000	0.002	0.003	0.000	0.001	0.007	0.006
		0.8	0.006	0.008	0.000	0.000	-0.003	-0.002	0.002	0.002	0.004	0.005	0.003	0.005
		0.9	0.000	-0.002	0.000	0.000	-0.002	-0.003	0.000	0.000	0.000	0.001	0.003	0.003
	100	0.1	**0.100**	**0.101**	0.019	0.023	**0.056**	**0.055**	**0.050**	**0.050**	**0.120**	**0.121**	**0.095**	**0.097**
		0.2	**0.073**	**0.074**	-0.017	-0.019	0.021	0.024	0.005	0.004	0.049	**0.053**	0.022	0.023
		0.3	0.038	0.033	0.011	0.015	0.022	0.024	0.005	0.004	0.033	0.035	0.023	0.022
		0.4	0.030	0.029	0.013	0.010	0.012	0.010	-0.002	-0.004	0.018	0.019	0.018	0.015
		0.5	0.014	0.014	0.009	0.011	0.008	0.010	-0.001	-0.001	0.008	0.006	0.012	0.011
		0.6	0.005	0.006	0.010	0.009	0.008	0.006	-0.002	-0.003	0.003	0.002	0.009	0.008

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.7	0.006	0.007	0.006	0.007	-0.001	0.001	0.001	0.001	0.004	0.006	0.007	0.008
		0.8	0.006	0.008	0.003	0.002	-0.001	0.001	0.001	0.001	0.002	0.003	0.003	0.002
		0.9	0.004	0.004	0.002	0.002	0.000	0.001	0.001	0.001	0.000	0.001	0.000	-0.001
	300	0.1	-0.001	0.000	0.004	0.003	0.012	0.005	0.011	0.010	0.042	0.043	0.040	0.039
		0.2	0.008	0.010	0.009	0.009	0.001	0.005	-0.005	-0.004	0.026	0.027	0.013	0.014
		0.3	0.017	0.018	0.003	0.003	0.004	0.002	0.002	0.003	0.014	0.015	0.007	0.005
		0.4	0.000	0.000	-0.003	-0.002	-0.001	-0.003	0.001	0.001	0.007	0.007	0.004	0.003
		0.5	-0.004	-0.004	-0.008	-0.008	-0.001	-0.001	0.001	0.002	0.004	0.004	0.000	-0.001
		0.6	-0.002	-0.002	-0.004	-0.005	-0.003	-0.002	0.002	0.001	0.002	0.003	-0.002	-0.002
		0.7	-0.002	-0.003	0.001	0.001	-0.002	-0.002	0.002	0.002	0.000	0.000	-0.001	-0.001
		0.8	-0.002	-0.001	0.001	0.001	-0.001	0.001	0.002	0.002	-0.001	0.000	-0.001	-0.001
		0.9	0.000	0.000	0.000	0.000	0.000	0.001	-0.001	-0.001	0.000	0.000	0.000	0.000
.10	30	0.1	0.023	0.005	**0.269**	**0.262**	**0.151**	**0.162**	**0.174**	**0.169**	**0.271**	**0.273**	**0.302**	**0.302**
		0.2	-0.008	0.008	0.007	-0.003	**0.082**	**0.081**	**0.051**	**0.053**	**0.102**	**0.094**	**0.085**	**0.089**
		0.3	0.017	0.016	0.036	0.039	0.047	0.044	0.007	0.014	**0.052**	**0.056**	**0.053**	**0.057**
		0.4	0.029	0.026	0.009	0.008	0.029	0.027	-0.003	-0.007	0.030	0.031	0.029	0.028
		0.5	0.027	0.027	-0.008	-0.009	0.012	0.012	-0.002	-0.001	0.010	0.016	0.022	0.019
		0.6	0.004	0.003	0.005	0.011	0.000	-0.001	0.005	0.004	0.003	0.005	0.018	0.017
		0.7	0.000	-0.001	0.015	0.021	-0.003	-0.006	0.004	0.004	0.002	0.006	0.019	0.020
		0.8	-0.004	-0.008	0.004	0.001	-0.005	-0.006	0.000	0.001	0.000	0.000	0.005	0.006
		0.9	0.000	0.000	0.002	0.002	-0.004	-0.004	0.001	0.001	0.003	0.004	0.002	0.003
	50	0.1	0.003	0.002	**0.053**	**0.055**	**0.052**	**0.056**	-0.005	-0.001	**0.143**	**0.142**	**0.126**	**0.124**
		0.2	0.001	-0.002	0.025	0.023	0.041	0.034	0.016	0.013	**0.055**	**0.054**	**0.056**	**0.058**

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.3	0.011	0.006	-0.009	-0.011	-0.005	-0.006	0.006	0.002	0.025	0.026	0.013	0.007
		0.4	-0.004	-0.015	-0.009	-0.006	0.000	-0.004	0.010	0.010	0.019	0.019	0.011	0.007
		0.5	-0.005	-0.008	-0.006	-0.007	-0.002	-0.003	0.006	0.005	0.018	0.015	0.010	0.011
		0.6	0.006	0.003	0.009	0.006	-0.001	0.000	0.002	0.004	0.010	0.009	0.010	0.009
		0.7	0.004	0.002	0.006	0.005	0.006	0.002	0.001	0.000	0.007	0.008	0.005	0.004
		0.8	0.006	0.005	-0.002	0.000	-0.002	-0.006	0.002	0.001	0.000	0.001	0.004	0.003
		0.9	-0.001	-0.003	0.001	0.000	-0.003	-0.004	0.004	0.004	0.000	0.000	0.004	0.002
	100	0.1	**-0.052**	-0.049	0.010	0.007	-0.015	-0.018	-0.003	-0.006	0.022	0.021	**0.061**	**0.061**
		0.2	0.022	0.022	0.012	0.009	0.021	0.023	-0.004	-0.006	0.016	0.012	0.022	0.020
		0.3	0.001	0.000	0.009	0.007	0.008	0.006	-0.010	-0.010	0.013	0.014	0.015	0.016
		0.4	0.007	0.006	0.005	0.006	0.004	0.004	-0.004	-0.005	0.007	0.005	0.013	0.015
		0.5	-0.004	-0.005	0.003	0.005	0.003	0.004	-0.005	-0.005	0.004	0.004	0.008	0.008
		0.6	-0.005	-0.005	0.003	0.002	0.003	0.001	-0.005	-0.004	0.002	0.001	0.002	0.004
		0.7	-0.006	-0.006	-0.004	-0.008	0.001	-0.002	-0.003	-0.003	-0.001	-0.002	0.002	0.002
		0.8	-0.005	-0.006	0.002	0.001	-0.002	-0.001	-0.002	-0.002	-0.002	-0.001	0.002	0.003
		0.9	-0.002	-0.003	0.001	-0.001	-0.005	-0.005	0.000	0.000	-0.002	-0.002	0.002	0.002
	300	0.1	0.022	0.021	0.029	0.034	0.009	0.012	-0.004	-0.004	0.025	0.023	0.018	0.019
		0.2	0.012	0.014	-0.006	-0.004	0.010	0.008	0.008	0.009	0.009	0.011	0.007	0.007
		0.3	0.009	0.009	-0.003	-0.004	0.006	0.005	0.002	0.001	0.005	0.006	0.004	0.004
		0.4	0.011	0.009	0.002	0.003	0.005	0.006	0.002	0.002	0.005	0.005	0.002	0.001
		0.5	0.007	0.007	-0.004	-0.003	0.004	0.004	0.000	0.000	0.003	0.003	0.001	0.002
		0.6	0.010	0.012	0.005	0.004	0.003	0.003	0.002	0.001	0.003	0.003	0.002	0.002
		0.7	0.002	0.000	0.003	0.004	-0.002	0.000	0.001	0.001	0.003	0.003	-0.002	-0.002
		0.8	-0.001	-0.001	0.003	0.003	0.001	0.000	0.000	0.000	0.001	0.002	-0.001	-0.001

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.9	0.002	0.002	0.000	0.000	-0.002	0.000	0.000	0.000	0.001	0.001	-0.001	0.000
.30	30	0.1	**0.278**	**0.272**	**0.267**	**0.247**	**0.252**	**0.252**	**0.261**	**0.253**	**0.231**	**0.232**	**0.262**	**0.264**
		0.2	0.042	0.039	**0.080**	**0.068**	0.049	0.043	0.003	0.004	0.027	0.020	0.049	**0.050**
		0.3	0.024	0.024	0.033	0.037	0.025	0.021	-0.003	-0.003	0.025	0.018	0.011	0.009
		0.4	-0.003	-0.005	0.026	0.031	0.021	0.011	-0.011	-0.013	0.009	0.009	0.008	0.002
		0.5	0.011	0.009	0.025	0.026	0.010	0.000	-0.009	-0.014	0.002	0.001	0.000	-0.001
		0.6	-0.013	-0.008	0.012	0.014	0.008	0.003	0.002	0.002	-0.003	-0.004	0.005	0.005
		0.7	-0.015	-0.015	-0.002	-0.005	-0.001	0.000	0.004	0.001	-0.004	-0.007	0.001	0.001
		0.8	-0.014	-0.012	-0.002	-0.003	-0.005	-0.007	-0.002	-0.003	-0.002	-0.005	-0.002	-0.002
		0.9	-0.011	-0.009	-0.001	-0.002	-0.006	-0.005	0.000	0.002	-0.001	-0.002	0.002	-0.001
	50	0.1	0.032	0.031	**0.141**	**0.131**	**0.102**	**0.085**	-0.009	-0.011	**0.069**	**0.071**	**0.071**	**0.077**
		0.2	-0.010	-0.006	0.038	0.030	**0.064**	**0.062**	0.004	0.005	0.015	0.015	0.002	0.003
		0.3	0.007	0.002	0.025	0.022	0.018	0.010	-0.004	-0.007	-0.005	-0.005	0.002	0.003
		0.4	0.009	0.006	0.014	0.014	0.013	0.014	-0.008	-0.008	-0.011	-0.014	0.003	0.001
		0.5	0.007	0.006	0.012	0.003	0.013	0.009	-0.005	-0.005	-0.008	-0.005	0.006	0.005
		0.6	0.007	0.006	-0.007	-0.007	0.008	0.006	-0.005	-0.006	-0.006	-0.006	0.001	-0.003
		0.7	-0.010	-0.012	-0.002	-0.001	0.004	-0.003	-0.008	-0.006	-0.003	-0.003	0.001	0.001
		0.8	-0.003	-0.008	-0.002	0.000	0.000	-0.002	0.000	-0.003	-0.001	-0.001	0.003	0.002
		0.9	0.001	-0.001	-0.004	-0.006	-0.002	-0.002	0.002	0.001	-0.001	0.000	0.001	0.000
	100	0.1	0.031	0.029	-0.002	-0.017	0.031	0.030	0.011	0.003	0.038	0.040	0.010	0.011
		0.2	-0.014	-0.010	-0.011	-0.017	-0.001	-0.003	-0.024	-0.023	-0.005	-0.006	-0.003	-0.003
		0.3	-0.004	-0.009	0.018	0.015	0.009	0.010	-0.007	-0.003	-0.008	-0.007	-0.006	-0.006
		0.4	0.003	0.007	0.002	0.005	0.007	0.007	-0.005	-0.005	0.001	0.001	-0.002	-0.003

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.5	-0.002	0.000	-0.001	-0.007	0.003	0.002	-0.011	-0.013	-0.003	-0.004	0.000	-0.003
		0.6	-0.003	-0.005	-0.007	-0.006	0.000	0.001	-0.005	-0.007	-0.002	-0.002	-0.002	-0.002
		0.7	-0.004	-0.006	-0.004	-0.004	-0.001	-0.004	-0.002	-0.002	-0.004	-0.003	-0.001	-0.002
		0.8	-0.003	-0.003	0.003	0.001	0.002	0.001	-0.002	-0.002	0.001	0.001	0.001	-0.001
		0.9	0.001	0.003	0.001	0.000	0.000	-0.001	0.002	0.001	0.002	0.002	0.002	0.002
	300	0.1	0.023	0.022	-0.020	-0.023	0.001	0.003	-0.015	-0.014	-0.001	-0.001	-0.006	-0.007
		0.2	0.010	0.010	-0.016	-0.016	0.001	0.002	-0.007	-0.006	0.006	0.004	-0.009	-0.011
		0.3	0.006	0.003	-0.002	-0.004	0.006	0.007	0.001	-0.001	0.000	-0.002	-0.003	-0.003
		0.4	0.006	0.006	-0.009	-0.008	-0.005	-0.002	-0.002	-0.002	0.001	0.001	-0.006	-0.006
		0.5	0.002	0.002	-0.003	-0.001	-0.001	0.001	0.001	0.001	0.003	0.003	0.001	0.001
		0.6	0.002	0.001	-0.002	-0.001	0.001	0.001	0.000	0.001	0.002	0.002	-0.001	-0.002
		0.7	0.002	0.002	0.003	0.002	0.002	-0.001	0.000	0.001	0.002	0.003	0.002	0.003
		0.8	0.001	0.002	0.002	0.003	0.000	-0.001	0.001	0.001	0.000	0.001	0.001	0.001
		0.9	0.001	0.002	-0.001	-0.001	0.000	0.000	0.000	-0.001	0.001	0.001	0.001	0.000
.50	30	0.1	**0.067**	0.048	**0.097**	**0.085**	**0.081**	**0.078**	**0.196**	**0.190**	**0.085**	**0.093**	**0.145**	**0.146**
		0.2	**0.056**	0.045	0.030	0.041	0.006	0.002	-0.012	-0.016	-0.029	-0.037	-0.020	-0.034
		0.3	0.024	0.028	0.020	0.024	0.008	0.008	-0.008	-0.013	-0.009	-0.011	-0.007	-0.012
		0.4	0.004	0.000	0.000	-0.006	0.021	0.023	-0.014	-0.022	-0.019	-0.024	-0.015	-0.027
		0.5	0.006	0.001	-0.005	-0.017	0.014	0.018	-0.025	-0.028	-0.011	-0.011	-0.021	-0.026
		0.6	0.003	0.007	-0.014	-0.016	0.015	0.010	-0.015	-0.016	-0.010	-0.013	-0.009	-0.013
		0.7	-0.006	-0.010	-0.014	-0.018	-0.001	-0.005	-0.019	-0.024	-0.008	-0.012	-0.010	-0.012
		0.8	-0.008	-0.007	-0.014	-0.016	-0.006	-0.007	-0.009	-0.012	-0.002	-0.006	0.000	-0.004
		0.9	-0.006	-0.007	-0.008	-0.010	-0.004	-0.002	-0.001	-0.002	0.002	-0.001	0.002	-0.002

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = | bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
	50	0.1	**0.120**	**0.119**	**0.147**	**0.147**	**0.136**	**0.137**	0.035	0.035	**0.054**	**0.050**	**0.058**	**0.053**
		0.2	-0.002	0.005	0.018	0.015	0.021	0.015	-0.012	-0.017	0.003	-0.007	-0.015	-0.021
		0.3	0.015	0.007	0.018	0.021	0.022	0.007	-0.016	-0.015	-0.008	-0.019	-0.015	-0.019
		0.4	0.005	0.008	0.008	0.005	0.007	0.004	-0.007	-0.011	-0.014	-0.016	-0.009	-0.013
		0.5	0.001	-0.003	0.009	0.002	-0.002	-0.006	-0.013	-0.019	-0.007	-0.010	-0.002	-0.009
		0.6	-0.007	-0.006	0.002	0.002	-0.001	-0.003	-0.002	-0.004	-0.005	-0.007	-0.004	-0.003
		0.7	-0.002	-0.005	0.001	-0.003	0.001	-0.005	-0.004	-0.004	-0.001	-0.003	0.000	-0.003
		0.8	-0.007	-0.008	-0.003	-0.003	0.003	0.001	0.000	-0.003	0.004	0.001	0.000	-0.001
		0.9	-0.004	-0.005	-0.001	-0.003	-0.002	-0.002	0.001	0.000	0.003	0.002	0.003	0.000
	100	0.1	0.007	0.006	**0.060**	**0.058**	0.011	0.011	-0.011	-0.009	-0.019	-0.020	-0.030	-0.017
		0.2	0.008	0.010	0.018	0.021	0.005	0.002	-0.016	-0.019	-0.008	-0.013	-0.025	-0.024
		0.3	0.005	-0.002	0.018	0.014	-0.003	-0.002	-0.007	-0.009	-0.008	-0.007	-0.008	-0.006
		0.4	0.003	-0.001	0.013	0.008	0.004	0.005	-0.008	-0.008	-0.005	-0.003	-0.002	-0.003
		0.5	0.004	0.001	0.003	0.001	-0.001	-0.006	-0.007	-0.007	-0.007	-0.006	-0.003	-0.002
		0.6	0.003	0.000	0.000	0.001	-0.002	-0.003	-0.008	-0.010	-0.005	-0.006	-0.001	-0.001
		0.7	-0.003	-0.004	0.004	0.005	-0.003	-0.005	-0.005	-0.004	-0.001	0.000	0.001	0.001
		0.8	0.002	-0.003	0.000	-0.001	-0.001	-0.003	-0.003	-0.003	0.000	0.000	-0.002	-0.003
		0.9	-0.001	-0.002	0.001	0.003	0.001	-0.002	0.000	-0.002	0.002	0.000	0.001	-0.001
	300	0.1	-0.013	-0.013	0.015	0.013	0.005	0.007	-0.025	-0.026	-0.031	-0.030	-0.029	-0.030
		0.2	-0.002	-0.004	0.007	0.004	0.002	0.001	-0.010	-0.011	-0.008	-0.006	-0.015	-0.014
		0.3	-0.008	-0.013	0.004	0.003	0.003	0.002	-0.008	-0.007	-0.007	-0.007	-0.006	-0.006
		0.4	-0.005	-0.006	0.004	0.003	0.004	0.004	-0.005	-0.006	-0.004	-0.005	-0.001	-0.002
		0.5	-0.004	-0.006	0.001	0.001	0.001	0.000	-0.006	-0.006	-0.005	-0.005	-0.002	-0.003
		0.6	0.000	-0.001	0.000	-0.001	0.000	-0.005	-0.002	-0.002	-0.004	-0.004	-0.003	-0.003

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.7	-0.001	-0.001	0.000	0.001	0.000	-0.002	-0.001	-0.002	-0.004	-0.005	0.000	0.000
		0.8	-0.001	-0.002	0.000	0.000	-0.002	-0.002	0.000	-0.001	-0.001	-0.002	0.000	0.000
		0.9	0.000	0.001	0.001	0.001	0.000	-0.001	0.001	0.002	0.000	0.000	0.002	0.001
.70	30	0.1	**0.152**	**0.150**	**0.156**	**0.151**	**0.202**	**0.197**	**0.133**	**0.132**	**0.154**	**0.152**	**0.156**	**0.131**
		0.2	**0.070**	**0.068**	**0.063**	**0.058**	**0.071**	**0.068**	0.019	0.017	-0.032	-0.027	0.005	0.001
		0.3	0.024	0.025	0.032	0.032	0.048	0.040	-0.009	-0.015	-0.022	-0.028	0.003	0.002
		0.4	0.018	0.019	0.011	0.008	0.033	0.026	-0.008	-0.008	0.000	-0.003	0.004	0.001
		0.5	0.006	0.007	0.006	0.006	0.006	0.003	-0.006	-0.012	-0.008	-0.008	0.003	0.001
		0.6	-0.004	-0.003	0.007	0.009	0.013	0.008	-0.006	-0.011	0.004	0.000	0.002	0.003
		0.7	-0.008	-0.005	0.000	0.002	0.000	-0.004	-0.001	-0.004	-0.004	-0.011	0.000	-0.001
		0.8	-0.007	-0.009	-0.003	0.000	-0.002	-0.005	-0.003	-0.005	-0.002	-0.007	0.000	-0.003
		0.9	-0.008	-0.005	-0.005	-0.001	-0.005	-0.002	-0.002	0.000	-0.002	0.000	0.000	-0.002
	50	0.1	0.034	0.042	0.043	**0.055**	**0.051**	**0.059**	0.034	0.031	-0.004	-0.002	-0.048	**-0.051**
		0.2	0.023	0.026	0.019	0.022	0.028	0.023	-0.012	-0.012	-0.021	-0.014	-0.009	-0.009
		0.3	0.021	0.021	0.015	0.010	-0.003	-0.008	0.002	-0.003	-0.009	-0.013	-0.009	-0.018
		0.4	0.009	0.009	-0.004	-0.004	-0.007	-0.014	0.004	0.000	-0.017	-0.016	-0.014	-0.017
		0.5	0.007	0.003	-0.002	-0.006	-0.004	-0.008	0.001	0.001	-0.004	-0.005	-0.003	-0.005
		0.6	0.003	0.001	-0.002	-0.003	-0.001	0.002	0.002	-0.003	0.004	0.003	0.002	0.002
		0.7	0.000	-0.001	-0.005	-0.006	0.003	0.000	0.000	-0.003	0.004	0.002	0.006	0.002
		0.8	0.000	0.002	-0.004	-0.006	-0.004	-0.006	0.001	0.000	0.001	-0.002	-0.001	0.000
		0.9	-0.003	0.000	-0.003	-0.003	-0.007	-0.005	0.001	0.001	-0.001	-0.001	0.002	0.000
	100	0.1	0.015	0.005	0.022	0.019	0.049	0.044	-0.024	-0.027	-0.026	-0.029	-0.006	-0.013
		0.2	-0.001	-0.004	0.007	0.003	0.017	0.021	-0.007	-0.011	-0.011	-0.021	-0.020	-0.020

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.3	0.005	0.003	0.002	0.001	0.001	0.003	-0.010	-0.009	-0.007	-0.009	-0.013	-0.016
		0.4	0.000	-0.003	0.000	-0.001	-0.010	-0.012	-0.004	-0.006	-0.004	-0.007	-0.011	-0.014
		0.5	-0.001	-0.004	-0.002	-0.003	-0.008	-0.006	-0.004	-0.005	-0.005	-0.009	-0.006	-0.012
		0.6	-0.003	-0.004	0.001	0.002	0.001	-0.005	-0.002	-0.005	-0.004	-0.004	-0.001	-0.002
		0.7	-0.001	-0.003	0.000	0.000	-0.005	-0.004	-0.001	-0.002	-0.005	-0.007	-0.003	-0.004
		0.8	-0.001	-0.004	-0.001	-0.001	-0.003	-0.005	0.000	0.000	-0.002	-0.004	-0.003	-0.005
		0.9	0.000	-0.001	-0.002	-0.001	-0.002	-0.005	0.000	-0.001	-0.002	-0.004	-0.001	-0.004
	300	0.1	-0.002	-0.001	0.009	0.009	0.021	0.019	-0.017	-0.015	-0.020	-0.021	-0.013	-0.015
		0.2	0.000	0.005	0.010	0.008	0.002	-0.002	-0.007	-0.006	-0.006	-0.006	-0.009	-0.010
		0.3	0.004	0.006	0.004	0.002	-0.002	0.000	-0.001	-0.002	-0.005	-0.005	-0.008	-0.007
		0.4	0.002	0.003	0.003	0.004	-0.001	-0.002	0.000	-0.001	-0.004	-0.005	-0.005	-0.005
		0.5	0.001	0.000	0.004	0.004	-0.003	-0.004	0.001	0.001	-0.005	-0.004	-0.003	-0.003
		0.6	-0.002	-0.001	0.001	0.002	-0.002	-0.001	0.000	0.000	-0.004	-0.005	-0.004	-0.003
		0.7	0.000	0.000	0.002	0.002	-0.002	-0.002	0.001	0.000	0.001	0.001	-0.002	-0.002
		0.8	0.000	0.000	0.001	0.002	0.000	0.000	0.000	0.001	0.000	0.001	-0.001	0.000
		0.9	-0.002	0.000	-0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
.90	30	0.1	**0.055**	**0.054**	**0.055**	**0.054**	**0.057**	**0.064**	0.047	0.046	0.019	-0.002	0.049	0.029
		0.2	0.022	0.019	0.023	0.021	0.035	0.035	0.009	0.012	0.005	0.006	-0.022	-0.024
		0.3	0.011	0.011	0.011	0.012	0.029	0.028	0.007	0.007	-0.009	-0.003	0.005	0.006
		0.4	0.006	0.007	0.006	0.007	0.032	0.030	0.006	0.006	0.010	0.007	0.015	0.011
		0.5	0.004	0.005	0.003	0.003	0.003	0.000	0.004	0.004	0.002	0.000	0.009	0.005
		0.6	0.002	0.003	0.001	0.003	0.000	-0.007	0.002	0.002	0.003	-0.005	0.006	0.006
		0.7	0.000	0.001	-0.001	0.002	-0.002	-0.001	0.001	0.002	0.000	0.000	0.002	-0.001
		0.8	-0.001	0.002	-0.001	0.001	0.000	0.000	0.000	0.001	0.002	0.002	0.007	0.004

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

$\boldsymbol{\rho}$	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.9	-0.001	0.001	-0.002	0.001	-0.003	-0.003	0.000	0.002	0.001	-0.002	0.002	-0.002
	50	0.1	0.030	0.032	0.027	0.027	**0.061**	**0.067**	0.014	0.016	0.007	0.013	0.002	-0.003
		0.2	0.014	0.014	0.011	0.012	0.037	0.045	0.004	0.005	0.007	0.013	0.016	0.020
		0.3	0.007	0.007	0.003	0.004	0.004	0.002	0.001	0.001	-0.001	-0.008	0.001	-0.007
		0.4	0.001	0.004	0.001	0.002	0.001	0.003	0.001	0.001	0.000	-0.004	0.000	0.006
		0.5	0.001	0.002	0.000	0.001	0.011	0.002	-0.001	0.001	0.005	0.004	0.008	0.001
		0.6	0.000	0.002	0.000	0.001	0.004	0.000	0.000	0.001	0.002	-0.001	-0.004	-0.006
		0.7	-0.001	0.001	0.000	0.001	-0.002	-0.006	0.000	0.002	-0.005	-0.008	0.000	-0.002
		0.8	-0.001	0.001	0.000	0.001	-0.009	-0.012	0.000	0.001	-0.004	-0.007	-0.003	-0.009
		0.9	0.000	0.001	0.000	0.002	-0.002	0.000	0.000	0.001	0.001	0.000	0.001	0.000
	100	0.1	0.012	0.015	0.015	0.014	0.032	0.031	0.007	0.005	-0.023	-0.022	0.009	0.006
		0.2	0.005	0.005	0.002	0.003	0.022	0.024	-0.002	0.000	0.017	0.018	0.021	0.022
		0.3	0.002	0.003	0.002	0.003	0.025	0.025	0.000	0.001	0.017	0.015	0.027	0.026
		0.4	0.000	0.001	0.001	0.001	0.015	0.018	0.001	0.002	0.009	0.010	0.010	0.010
		0.5	0.001	0.001	0.000	0.001	0.009	0.012	0.001	0.002	0.004	0.008	0.008	0.006
		0.6	0.000	0.001	0.001	0.001	-0.002	-0.002	0.001	0.001	-0.001	-0.002	0.000	0.002
		0.7	0.000	0.001	0.000	0.001	0.001	-0.001	0.001	0.002	0.000	0.001	0.000	-0.001
		0.8	0.000	0.002	0.000	0.001	-0.004	-0.004	0.000	0.000	-0.005	-0.002	-0.004	-0.004
		0.9	0.000	0.001	-0.001	0.001	-0.004	-0.004	0.001	0.001	-0.005	-0.004	-0.004	-0.005
	300	0.1	0.003	0.004	0.006	0.006	0.018	0.017	0.001	0.001	0.016	0.018	0.010	0.008
		0.2	0.002	0.003	0.001	0.003	-0.003	-0.003	0.000	0.002	-0.002	0.000	0.004	0.002
		0.3	0.000	0.001	0.001	0.001	-0.001	0.000	0.000	0.000	-0.002	-0.003	-0.006	-0.003
		0.4	0.001	0.001	0.001	0.001	-0.004	-0.006	0.000	0.000	-0.006	-0.005	-0.008	-0.004

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = | bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
		0.5	0.001	0.001	0.000	0.000	-0.004	-0.003	0.000	0.000	-0.002	-0.001	-0.004	-0.003
		0.6	0.000	0.000	0.001	0.001	-0.002	-0.003	0.000	0.001	-0.001	-0.002	-0.003	0.000
		0.7	0.001	0.001	0.000	0.001	0.001	0.002	0.000	0.001	0.001	0.003	0.001	0.001
		0.8	0.000	0.001	0.000	0.001	0.002	0.001	0.000	0.001	-0.001	0.002	0.000	0.000
		0.9	0.000	0.001	0.000	0.000	0.000	-0.001	0.000	0.001	0.000	0.000	-0.001	-0.001
Klieger	30	0.1	**0.199**	**0.175**	**0.121**	**0.078**	**0.055**	0.042	**0.206**	**0.211**	**0.128**	**0.142**	**0.143**	**0.141**
		0.2	0.039	0.045	-0.004	0.005	0.049	0.044	**0.062**	**0.064**	-0.017	-0.032	-0.030	-0.028
		0.3	0.039	0.047	-0.017	-0.013	-0.004	-0.015	-0.012	-0.008	-0.022	-0.026	-0.017	-0.012
		0.4	0.033	0.024	-0.031	-0.026	-0.010	-0.012	-0.025	-0.024	-0.013	-0.015	-0.022	-0.025
		0.5	0.001	-0.001	-0.022	-0.020	0.008	0.009	-0.020	-0.020	-0.012	-0.017	-0.023	-0.023
		0.6	-0.012	-0.008	-0.009	-0.007	0.010	0.011	-0.012	-0.015	-0.008	-0.010	-0.019	-0.022
		0.7	0.003	0.006	-0.006	-0.009	0.003	0.001	-0.009	-0.014	-0.012	-0.008	-0.015	-0.015
		0.8	-0.001	-0.010	-0.006	-0.012	-0.001	-0.004	-0.003	-0.005	-0.005	-0.006	-0.009	-0.010
		0.9	-0.008	-0.010	-0.003	-0.005	-0.012	-0.013	-0.002	-0.002	-0.001	-0.003	-0.004	-0.005
	50	0.1	**0.087**	**0.085**	**0.142**	**0.153**	**0.085**	**0.075**	**0.061**	**0.072**	0.028	0.016	0.013	0.006
		0.2	-0.003	-0.004	0.018	0.023	0.043	0.044	0.003	0.003	-0.022	-0.032	-0.018	-0.015
		0.3	0.013	0.008	0.011	0.012	0.023	0.029	-0.006	-0.008	-0.020	-0.016	-0.015	-0.011
		0.4	-0.009	-0.010	0.005	-0.003	0.025	0.025	-0.011	-0.008	-0.005	-0.008	-0.010	-0.015
		0.5	0.007	0.006	0.013	0.013	0.011	0.014	-0.002	-0.003	-0.001	-0.002	-0.004	-0.005
		0.6	0.006	-0.002	0.001	0.001	0.003	0.003	-0.001	-0.001	-0.005	-0.009	-0.009	-0.011
		0.7	0.001	0.005	0.004	0.003	0.003	0.000	0.002	0.001	-0.010	-0.010	-0.009	-0.009
		0.8	0.000	0.001	0.000	-0.002	-0.001	-0.001	0.001	0.003	-0.002	-0.003	-0.004	-0.005
		0.9	-0.003	-0.004	-0.006	-0.008	0.000	-0.002	0.000	-0.003	-0.001	-0.001	-0.004	-0.004

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

Table E1

Median Bias in All Simulation Conditions Cont.

ρ	N	SR	Explicit Selection						Incidental Selection					
			r_{12}		r_{13}		r_{14}		r_{23}		r_{24}		r_{34}	
			UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT	UNI	MVT
	100	0.1	0.045	0.042	0.008	0.015	0.025	0.023	0.021	0.020	-0.016	-0.016	-0.010	-0.010
		0.2	0.022	0.025	-0.004	0.001	0.004	0.002	0.001	-0.001	0.000	-0.001	0.002	-0.005
		0.3	0.023	0.019	0.005	0.001	0.015	0.016	0.000	-0.001	-0.007	-0.007	-0.010	-0.011
		0.4	0.016	0.019	-0.007	-0.006	0.005	0.003	-0.006	-0.005	-0.001	-0.002	-0.013	-0.014
		0.5	0.006	0.010	-0.005	-0.005	0.000	0.000	-0.006	-0.005	-0.006	-0.004	-0.012	-0.012
		0.6	0.000	0.004	-0.014	-0.011	0.002	0.001	-0.005	-0.003	0.000	-0.002	-0.008	-0.008
		0.7	-0.001	0.004	-0.007	-0.012	0.004	0.006	-0.001	-0.002	0.001	0.002	-0.004	-0.004
		0.8	0.000	0.001	-0.008	-0.007	0.004	0.003	-0.003	-0.003	0.000	0.000	-0.003	-0.003
		0.9	-0.001	0.001	-0.005	-0.004	0.002	0.004	-0.001	-0.001	0.004	0.003	0.000	0.001
	300	0.1	-0.001	-0.003	-0.013	-0.009	0.008	0.009	-0.024	-0.021	-0.035	-0.030	-0.019	-0.021
		0.2	-0.011	-0.007	-0.006	-0.001	0.001	0.002	0.001	0.001	-0.011	-0.011	-0.018	-0.019
		0.3	-0.002	0.002	0.001	-0.002	0.000	0.002	-0.005	-0.003	-0.004	-0.004	-0.005	-0.005
		0.4	-0.006	-0.005	0.002	0.000	0.002	0.001	-0.005	-0.004	-0.006	-0.006	0.000	0.000
		0.5	0.003	0.002	0.001	0.004	0.000	-0.001	0.000	0.000	0.000	0.000	-0.001	0.000
		0.6	-0.002	0.001	0.004	0.003	0.001	0.000	-0.001	0.000	-0.001	-0.001	-0.002	-0.001
		0.7	-0.001	0.000	0.000	0.000	0.002	0.001	-0.001	0.000	0.001	0.002	0.000	0.000
		0.8	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.001	-0.001	0.000	0.000
		0.9	0.001	0.004	0.000	0.001	-0.001	-0.001	0.000	0.001	0.000	0.000	0.001	0.001

Note. SR = selection ratio, UNI = univariate correction, MVT = multivariate correction. Bold = $|$ bias $| \geq .05$.

APPENDIX F: EFFICIENCY IN THE CORRECTIONS

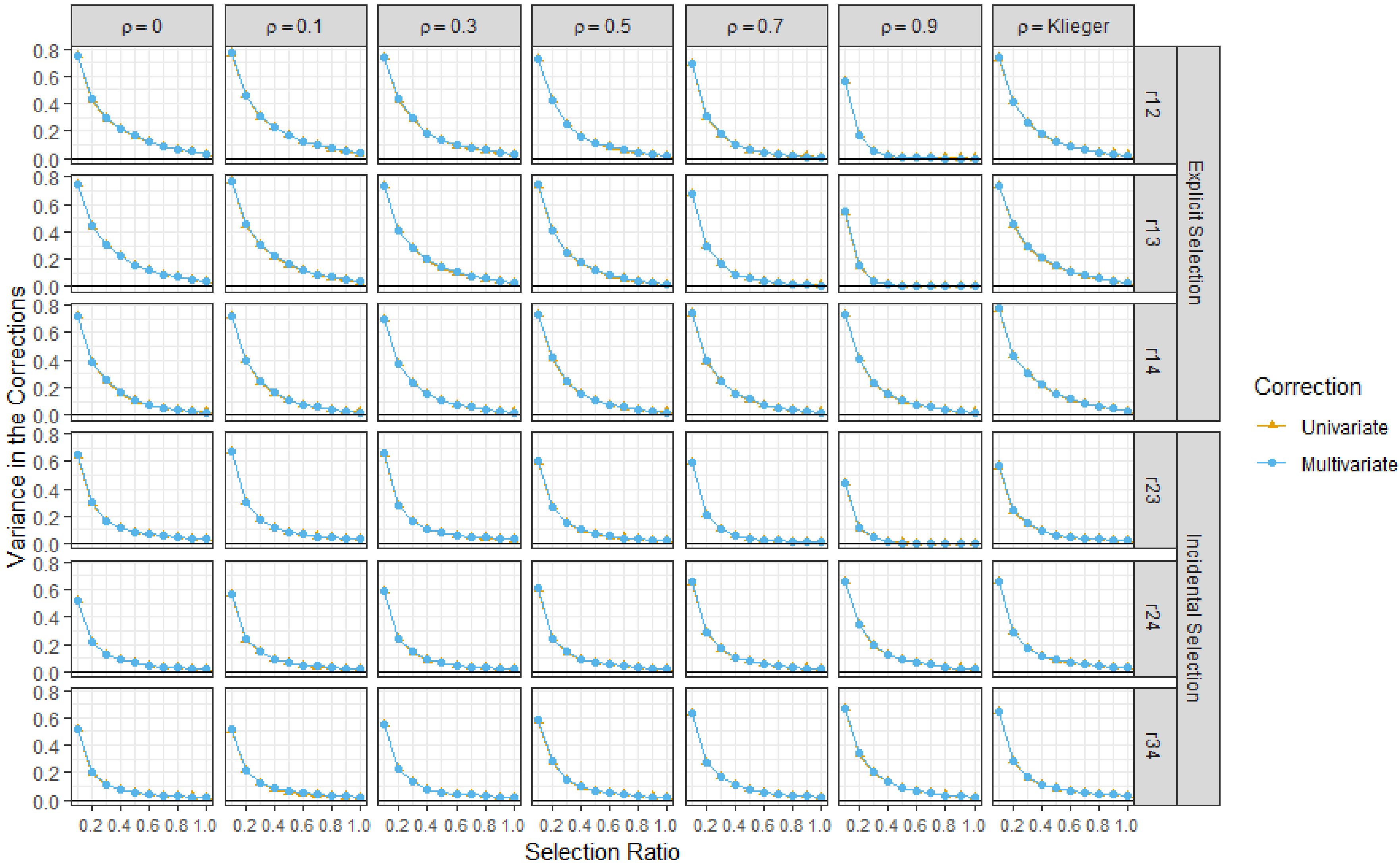

Figure F1. Variance of the corrected correlations by population correlation, ρ, and selection ratio; $N = 30$.

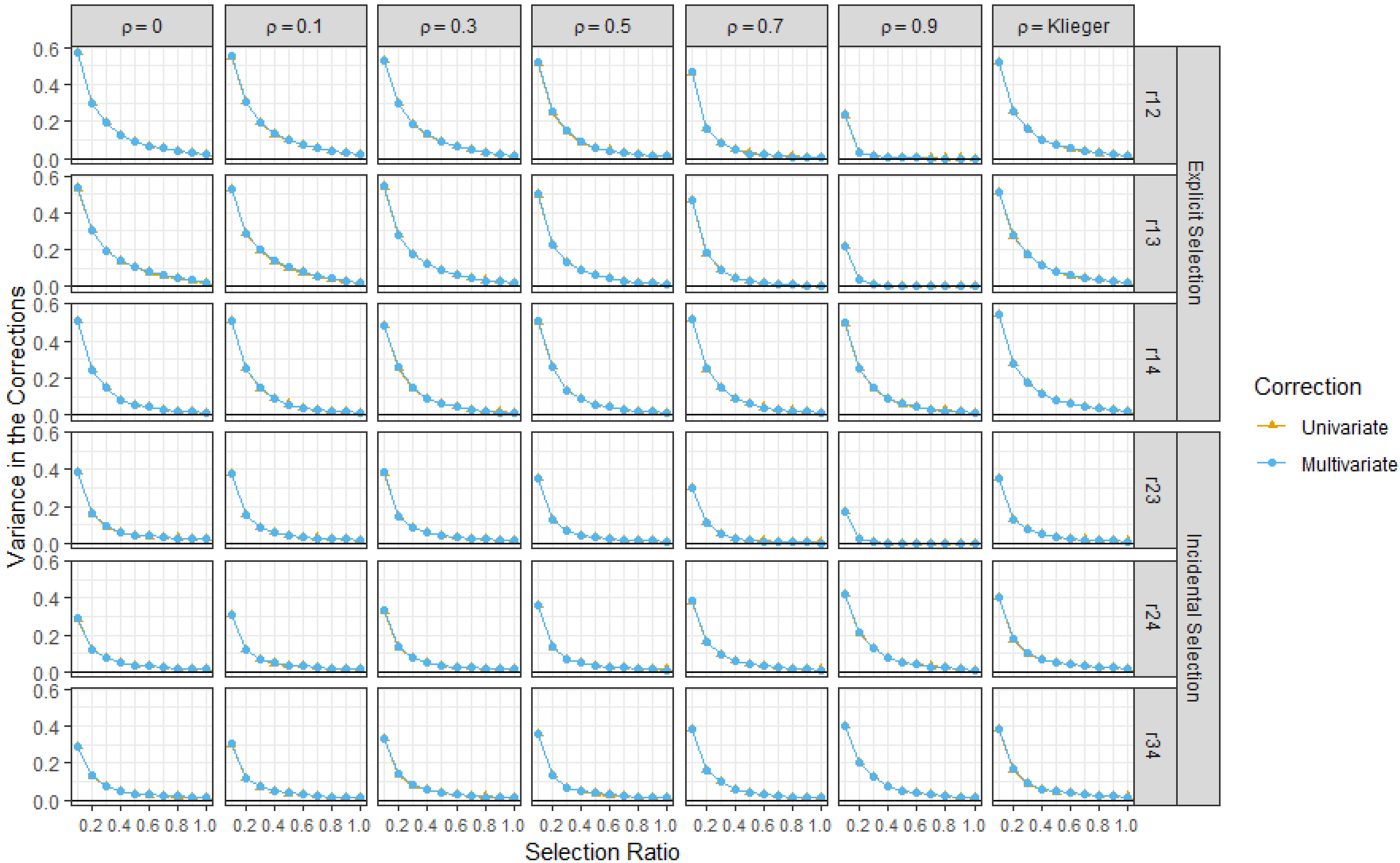

Figure F2. Variance of the corrected correlations by population correlation, ρ, and selection ratio; $N = 50$.

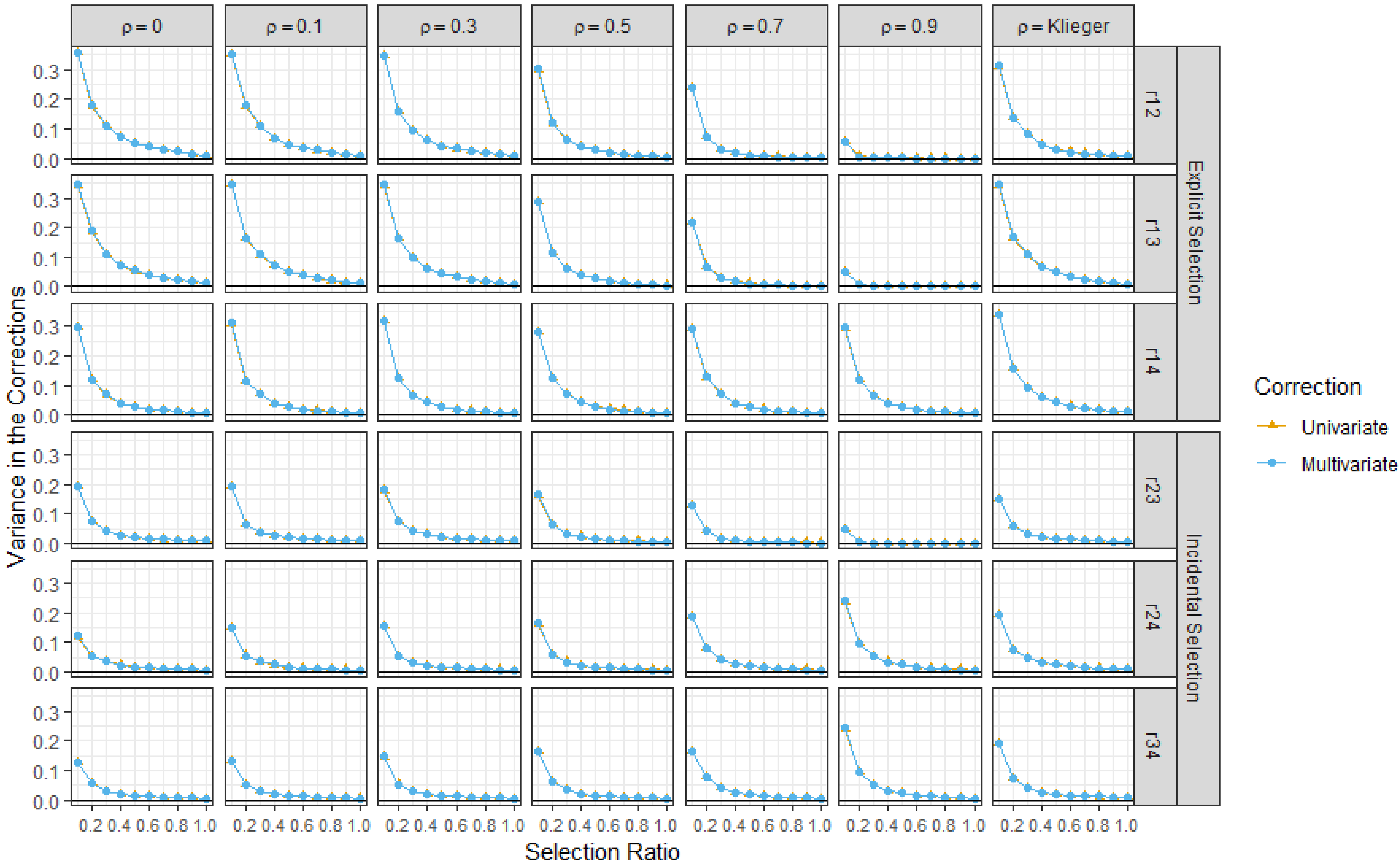

Figure F3. Variance of the corrected correlations by population correlation, ρ, and selection ratio; $N = 100$.

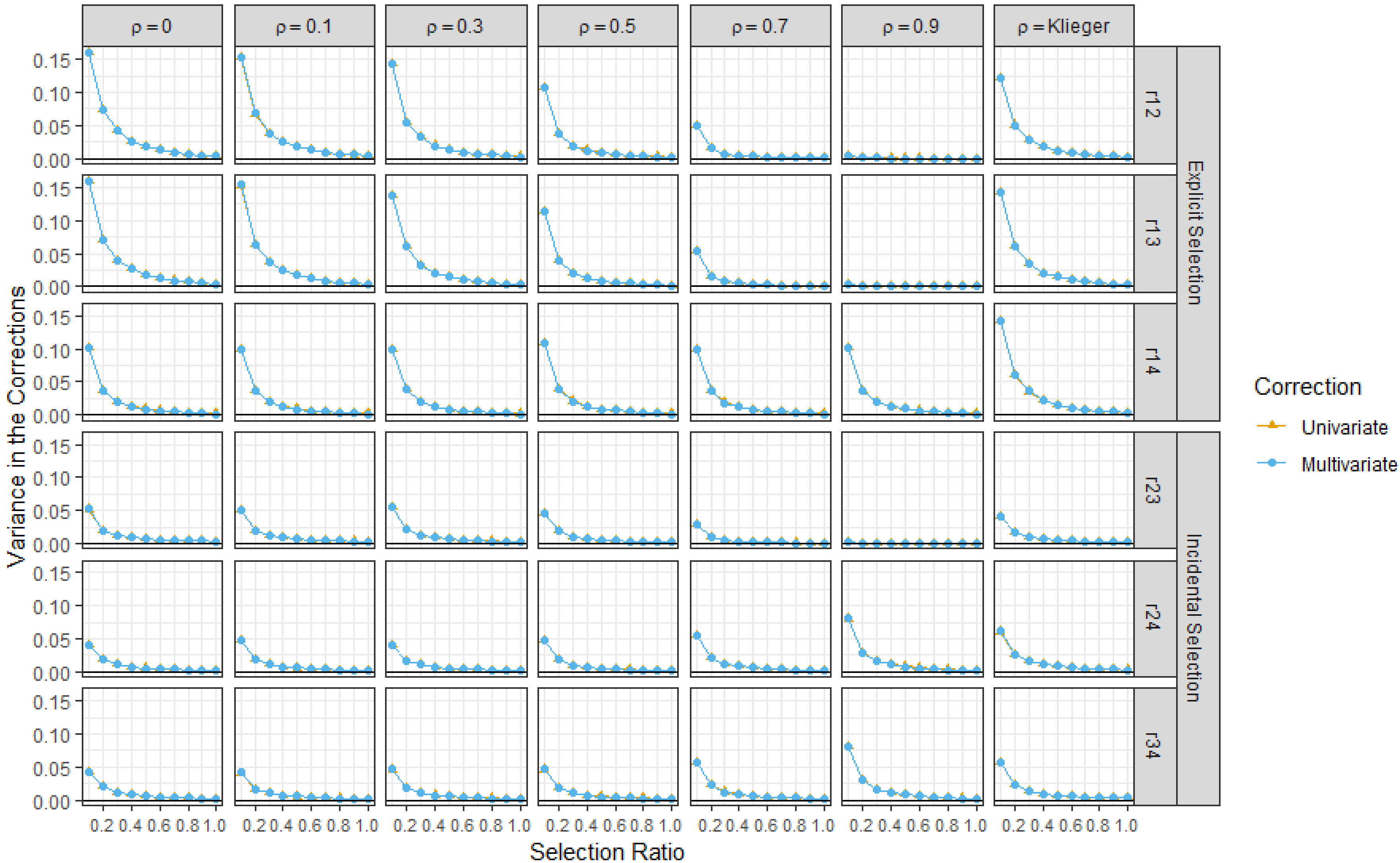

Figure F4. Variance of the corrected correlations by population correlation, ρ, and selection ratio; $N = 300$.

APPENDIX G: DISTRIBUTIONS OF RVE ESTIMATES BY CORRELATION

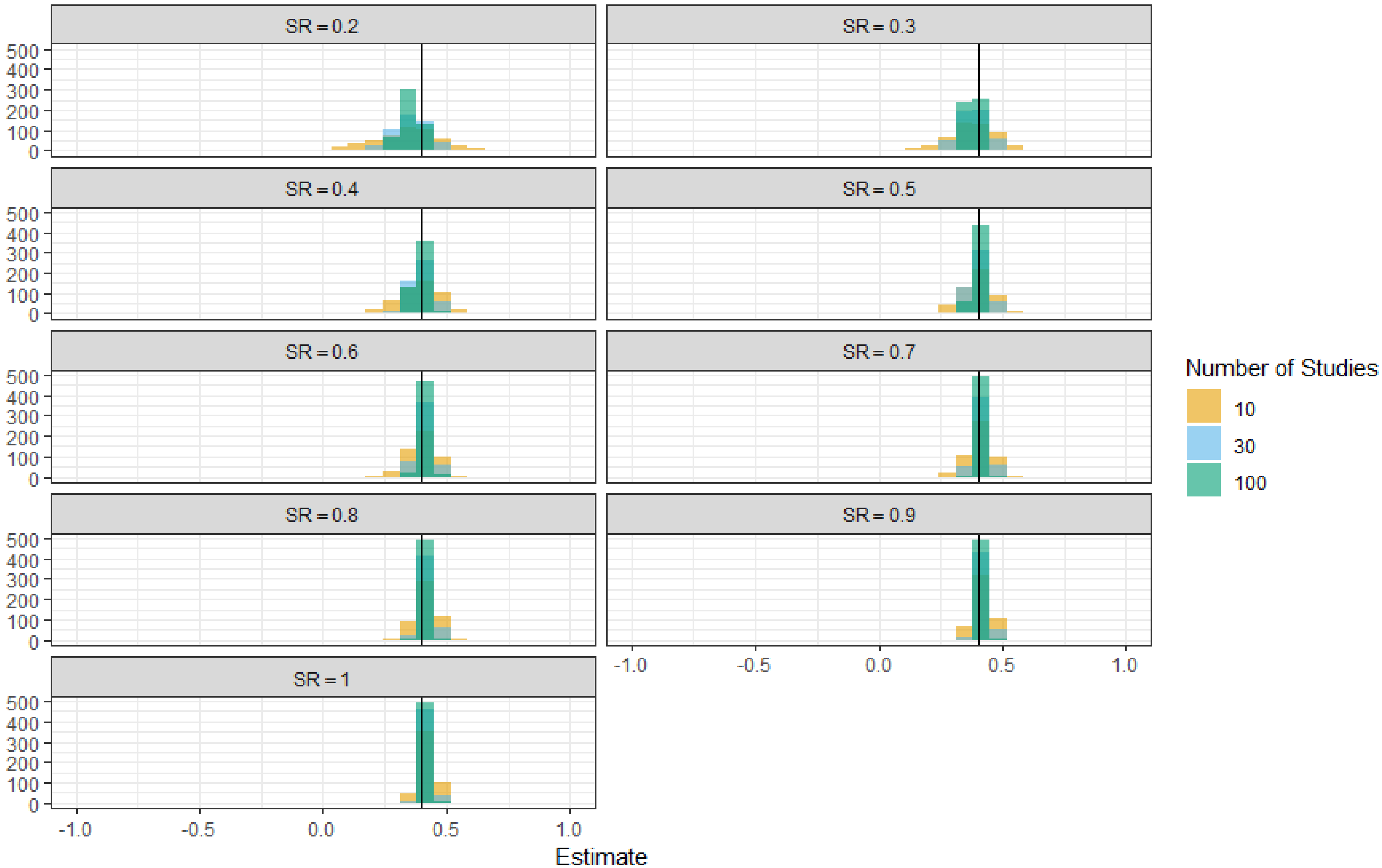

Figure G1. Distribution of the RVE estimates for r_{12} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

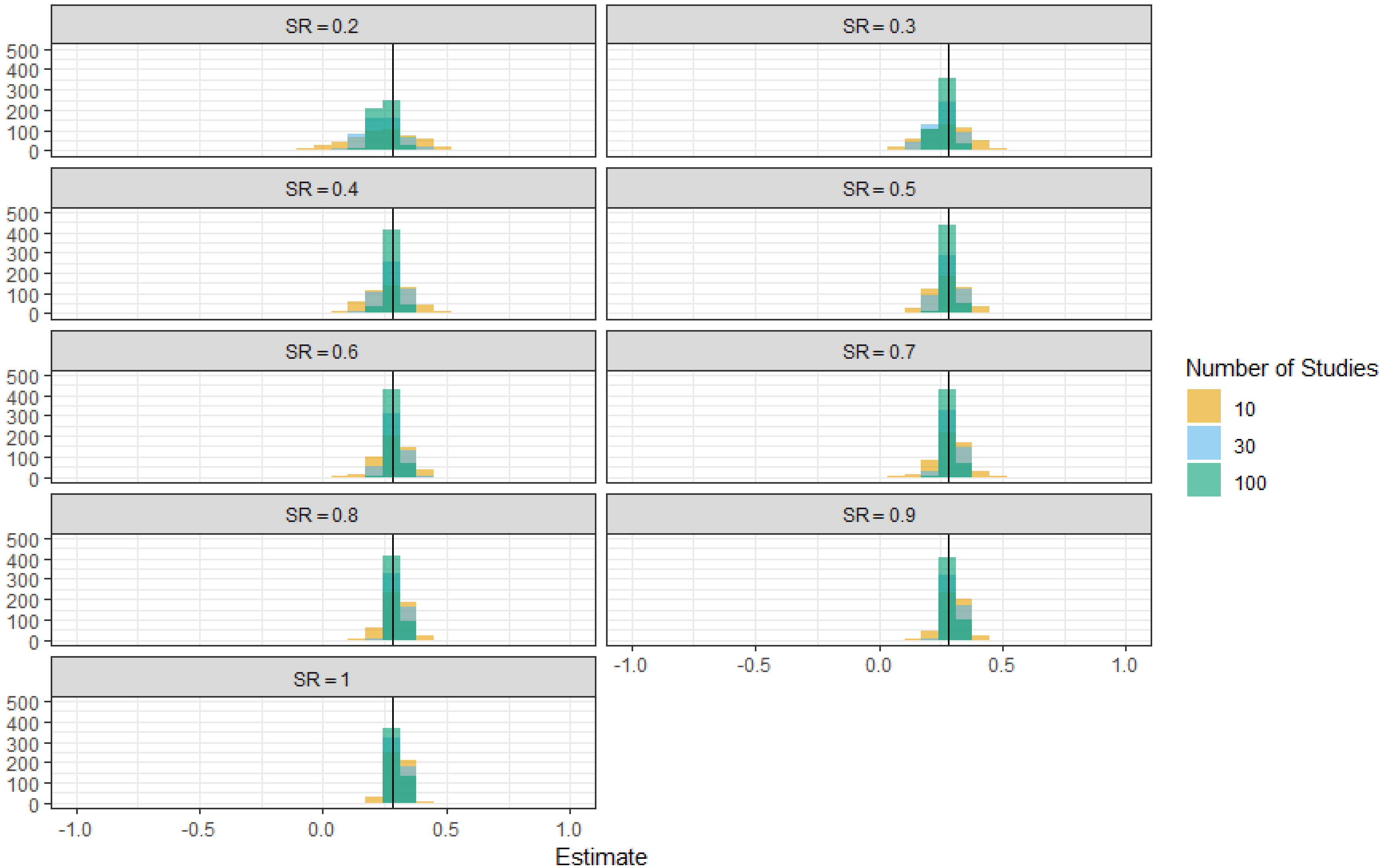

Figure G2. Distribution of the RVE estimates for r_{13} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

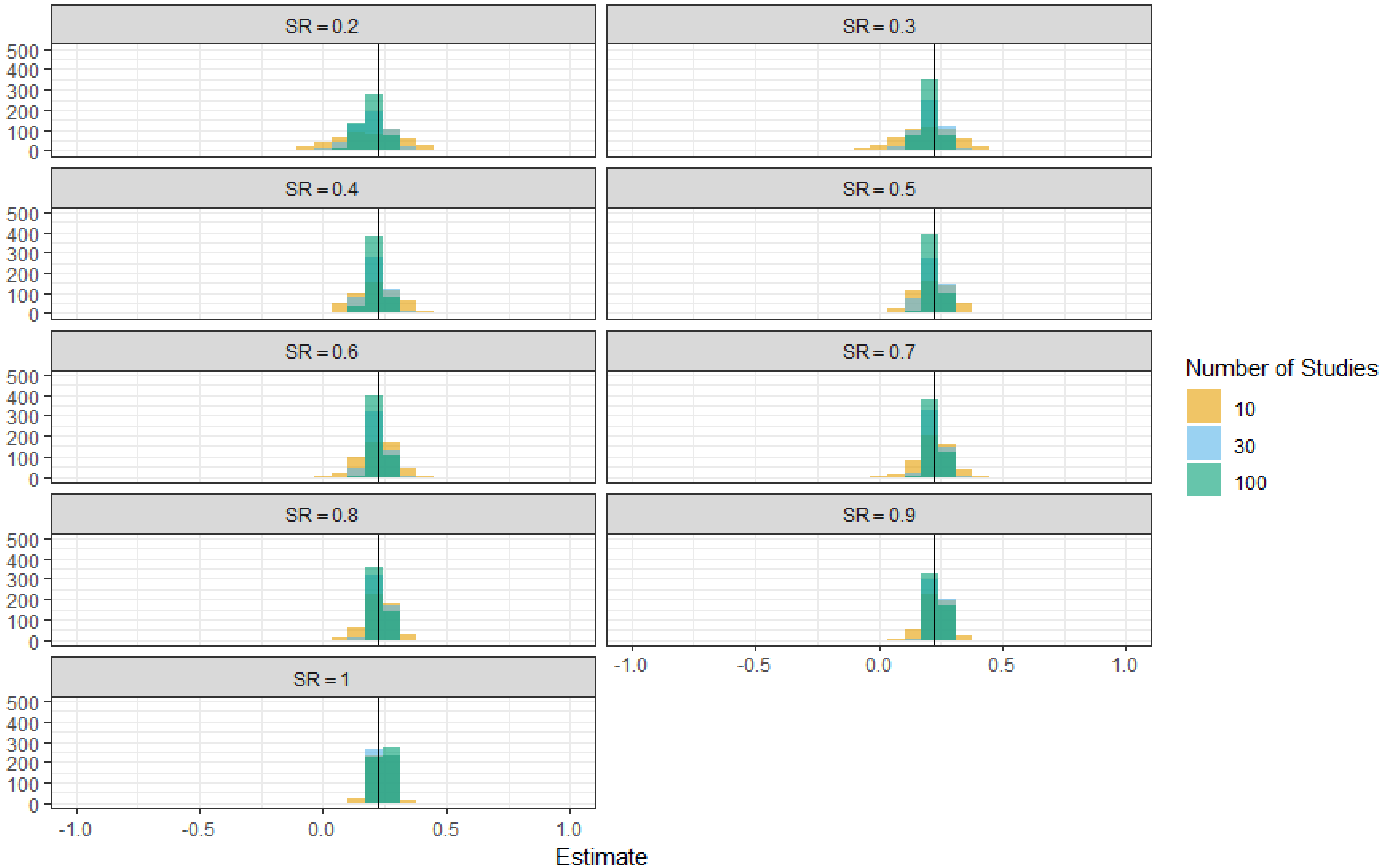

Figure G3. Distribution of the RVE estimates for r_{14} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

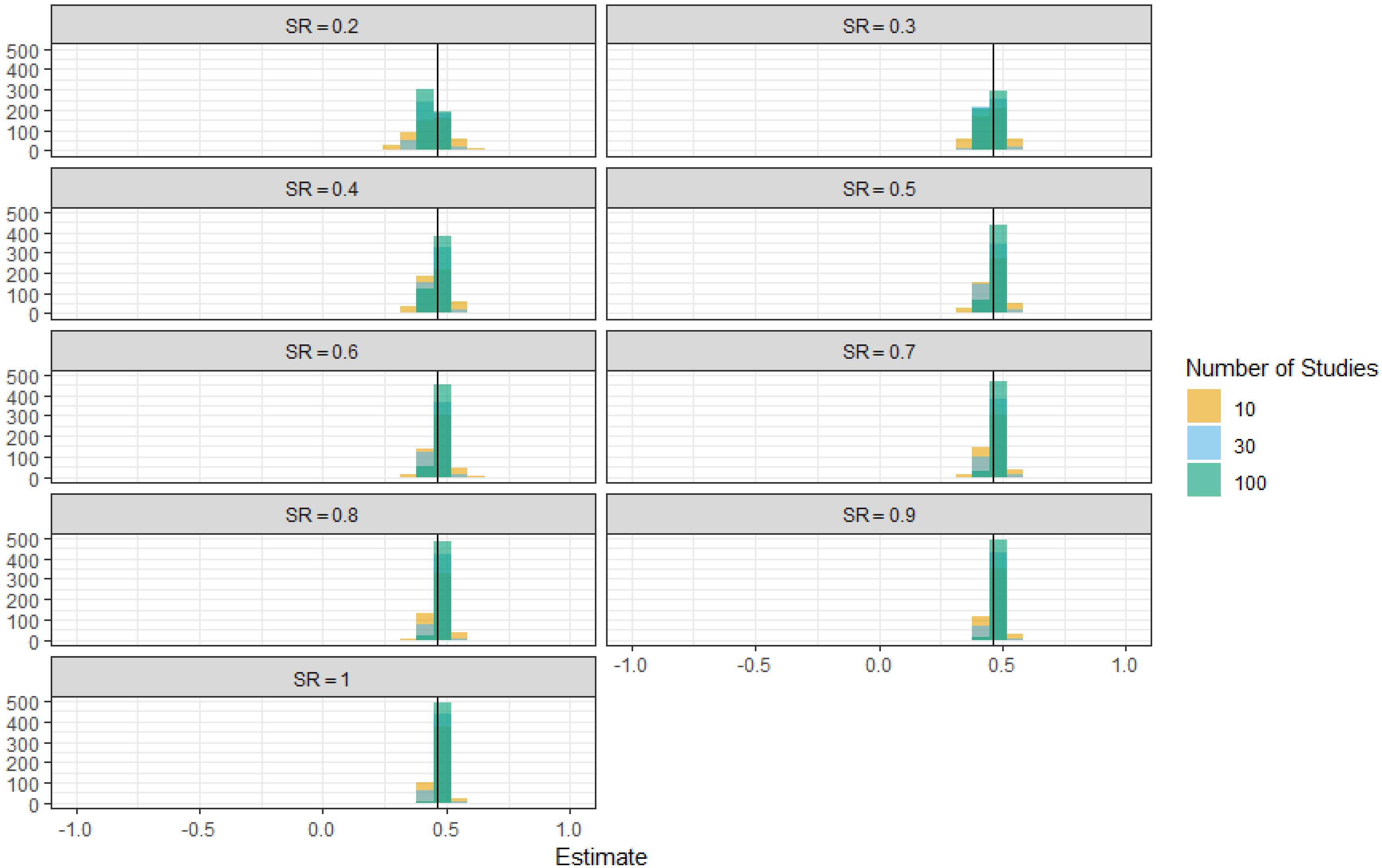

Figure G4. Distribution of the RVE estimates for r_{23} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

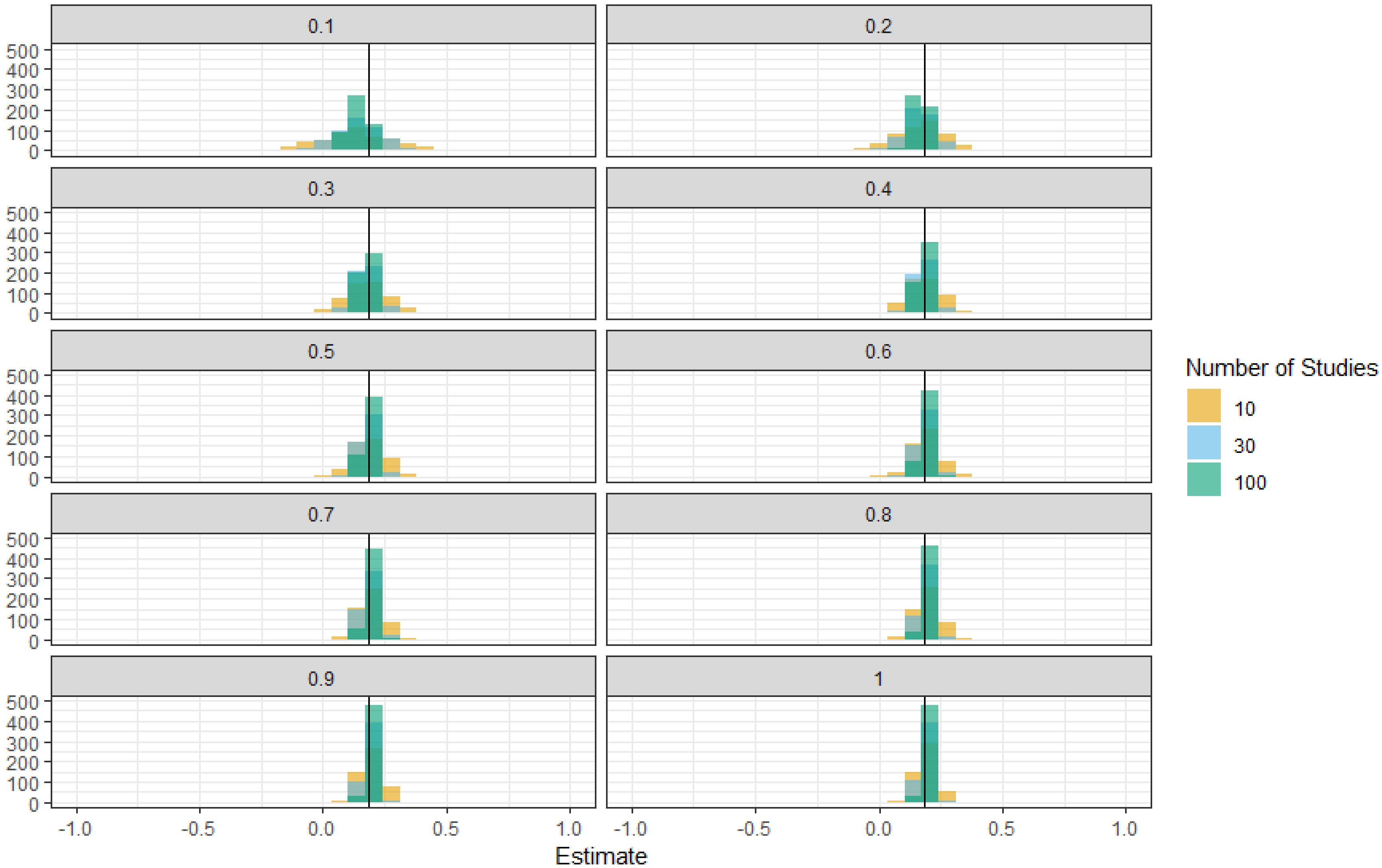

Figure G5. Distribution of the RVE estimates for r_{24} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

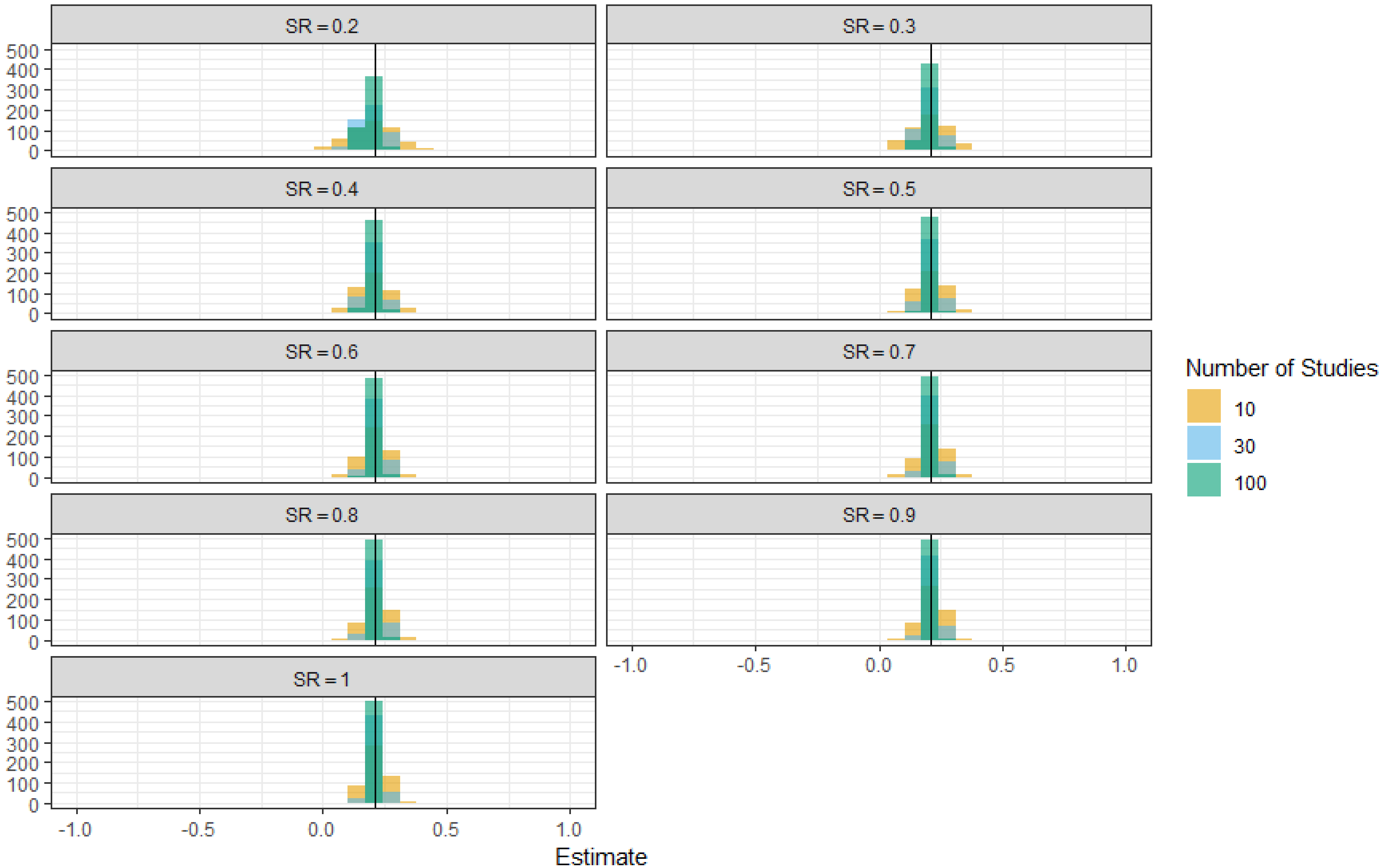

Figure G6. Distribution of the RVE estimates for r_{34} by selection ratio, SR; and number of studies, k; $\tau^2 = \frac{2}{3}v$; $N = 100$.

REFERENCES

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education. *Review of Educational Research*, *82*(4), 436–476. doi:10.3102/0034654312458162

Aitken, A. C. (1935). A note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, *4*, 106–110. doi:10.1017/s0013091500008063

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Allen, N. L. & Dunbar, S. B. (1990). Standard errors of correlations adjusted for incidental selection. *Applied Psychological Measurement*, *14*(1), 83–94. doi:10.1177/014662169001400109

Aloe, A. M. & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*, *37*(2), 278–297. doi:10.3102/1076998610396901

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, *99*(3), 388–399.

Baron, R. M. & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi:10.1037/0022-3514.51.6.1173

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics*, *17*(4), 341–362. doi:10.2307/1165128

Becker, B. J. (1995). Corrections to “using results from replicated studies to estimate linear models”. *Journal of Educational Statistics*, *20*(1), 100–102. doi:10.2307/1165390

Becker, B. J. (2001). Examining theoretical models through research synthesis. *Evaluation & the Health Professions*, *24*(2), 190–217. doi:10.1177/01632780122034876

Becker, B. J. (2007). Multivariate meta-analysis: Contributions of ingram olkin. *Statistical Science*, *22*(3), 401–406. doi:10.1214/07-sts239

Becker, B. J. (2009). Model-based meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 377–395). New York: Russell Sage Foundation.

Becker, B. J. & Fahrbach, K. R. (1994). *A comparison of approaches to the synthesis of correlation matrices.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

Becker, B. J. & Schram, C. M. (1994). Examining explanatory models through research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 357–381). New York: Russell Sage Foundation.

Becker, B. J. & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, *22*(3), 414–429. doi:10.1214/07-STS243

Birnbaum, Z. W., Paulson, E., & Andrews, F. C. (1950). On the effect of selection performed on some coordinates of a multi-dimensional population. *Psychometrika*, *15*(2), 191–204. doi:10.1007/bf02289200

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice* (3rd ed.). Cambridge, MA: MIT Press.

Bobko, P. (1983). An analysis of correlations corrected for attenuation and range restriction. *Journal of Applied Psychology*, *68*(4), 584–589. doi:10.1037/0021-9010.68.4.584

Bobko, P. & Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement*, *4*(3), 385–398. doi:10.1177/014662168000400309

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Campbell Collaboration. (2019). About us. Retrieved from https://campbellcollaboration.org

Chalmers, I. (2007). The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: The need for systematic reviews. In P. Rothwell (Ed.), *Treating individuals* (pp. 37–58). London: Lancet.

Cheung, M. W.-L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*, *13*, 182–202. doi:10.1037/a0013163

Cheung, M. W.-L. & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40–64. doi:10.1037/1082-989x.10.1.40

Cheung, S. F. & Chan, D. K.-S. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, *89*(5), 780–791. doi:10.1037/0021-9010.89.5.780

Cheung, S. F. & Chan, D. K.-S. (2008). Dependent correlations in meta-analysis. *Educational and Psychological Measurement*, *68*(5), 760–777. doi:10.1177/0013164408315263

Cheung, S. F. & Chan, D. K.-S. (2014). Meta-analyzing dependent correlations: An SPSS macro and an R script. *Behavior Research Methods*, *46*, 331–345. doi:10.3758/s13428-013-0386-2

Cheung, S. F., Chan, D. K.-S., & Sun, R. W. (2018). Meta-analyzing dependent correlations with correction for artifacts that multiplicatively attenuate the true correlation. *Behavior Research Methods*. doi:10.3758/s13428-018-1111-y

Cochrane Collaboration. (2019). About us. Retrieved from https://www.cochrane.org/

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis* (2nd ed.). NY: Russell Sage Foundation.

Cooper, H. & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, *14*(2), 165–176. doi:10.1037/a0015565

Credé, M. & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*(6), 425–453. doi:10.1111/j.1745-6924.2008.00089.x

Crocker, L. & Algina, J. (2008). *Introduction to classical & modern test theory*. Mason, Ohio: Cengage Learning.

Dahlke, J. A. & Wiernik, B. M. (2018). psychmeta: An R package for psychometric meta-analysis. *Applied Psychological Measurement*. [Online Only]. doi:10.1177/0146621618795933

Duan, B. & Dunlap, W. P. (1997). The accuracy of different methods for estimating the standard error of correlations corrected for range restriction. *Educational and Psychological Measurement*, *57*(2), 254–265. doi:10.1177/0013164497057002005

Elston, R. C. (1975). On the correlation between correlations. *Biometrika*, *62*(1), 133–140. doi:10.1093/biomet/62.1.133

Eysenck, H. J. (1965). The effects of psychotherapy. *International Journal of Psychiatry*, *1*, 97–178.

Fisher, Z. & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. Retrieved from https://arxiv.org/archive/stat

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: Robust variance meta-regression*. R package version 2.0. Retrieved from https://CRAN.R-project.org/package=robumeta

Glasener, K. M., Martell, C. A., & Posselt, J. R. (2019). Framing diversity: Examining the place of race in institutional policy and practice post-affirmative action. *Journal of Diversity in Higher Education*, *12*(1), 3–16. doi:10.1037/dhe0000086

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. doi:10.2307/1174772

Gleser, L. J. & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York: Russell Sage Foundation.

Greener, J. M. & Osburn, H. (1979). An empirical study of the accuracy of corrections for restriction in range due to explicit selection. *Applied Psychological Measurement*, *3*(1), 31–41. doi:10.1177/014662167900300104

Hafdahl, A. R. (2007). Combining correlation matrices: Simulation analysis of improved fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *32*, 180–205. doi:10.3102/1076998606298041

Hafdahl, A. R. (2008). Combining heterogeneous correlation matrices: Simulation analysis of fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *33*(4), 507–533. doi:10.3102/1076998607309472

Hartocollis, A. (2019, May 16). SAT's new 'adversity score' will take students' hardships into account. *The New York Times*. Retrieved from http://www.nytimes.com

Hedges, L. V. (1982). Fitting continuous models to effect size data. *Journal of Educational Statistics*, *7*(4), 245. doi:10.2307/1164634

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

Hedges, L. V. & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, *6*(3), 203–217. doi:10.1037/1082-989x.6.3.203

Hedges, L. V. & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, *9*(4), 426–445. doi:10.1037/1082-989x.9.4.426

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. doi:10.1002/jrsm.5

Hedges, L. V. & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504.

Held, J. D. & Foley, P. P. (1994). Explanations for accuracy of the general multivariate formulas in correcting for range restriction. *Applied Psychological Measurement*, *18*(4), 355–367. doi:10.1177/014662169401800406

Hong, C., Riley, R. D., & Chen, Y. (2017). An improved method for bivariate meta-analysis when within-study correlations are unknown. *Research Synthesis Methods*, *9*(1), 73–88. doi:10.1002/jrsm.1274

Hubert, L. J. (1972). A note on the restriction of range for pearson product-moment correlation coefficients. *Educational and Psychological Measurement*, *32*(3), 767–770. doi:10.1177/001316447203200315

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). New Jersey: Erlbaum. doi:10.4324/9781315211947-13

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

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction on meta-analysis methods and findings. *Journal of Applied Psychology*, *91*(3), 594–612. doi:10.1037/0021-9010.91.3.594

Jaschik, S. (2019, May 20). New SAT score: Adversity. *Inside Higher Education*. Retrieved from https://www.insidehighered.com/

Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2015). Why does self-reported emotional intelligence predict job performance? a meta-analytic investigation of mixed EI. *Journal of Applied Psychology*, *100*(2), 298–342. doi:10.1037/a0037681

Joseph, D. L. & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, *95*(1), 54–78. doi:10.1037/a0017286

Kaufman, R. L. (2014). *Heteroskedasticity in regression: Detection and correction*. Retrieved from https://doi.org/10.4135/9781452270128.n5

Kimbrel, L. A. (2019). Improving the teacher hiring process through the combination of teacher quality and employee selection research. *AASA Journal of Scholarship & Practice*, *15*(4). Retrieved from http://www.aasa.org/jsp.aspx

Klieger, D. M., Cline, F. A., Holtzman, S. L., Minsky, J. L., & Lorenz, F. (2014). New perspectives on the validity of the GRE® general test for predicting graduate school grades. *ETS Research Report Series*, *2014*(2), 1–62. doi:10.1002/ets2.12026

Kontopantelis, E. (2018). A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. *Research Synthesis Methods*. doi:10.1002/jrsm.1303

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, *25*, 120–148. doi:10.1016/j.edurev.2018.10.001

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181. doi:10.1037/0033-2909.127.1.162

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business and Psychology*, *28*(3), 251–261. doi:10.1007/s10869-013-9285-x

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical*, *62*, 28–30. doi:10.1017/s0080454100006385

Lawson, G. M., Hook, C. J., & Farah, M. J. (2017). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*(2), e12529. doi:10.1111/desc.12529

LeBreton, J. M., Scherer, K. T., & James, L. R. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology*, *7*(4), 478–500. doi:10.1111/iops.12184

Linn, R. L. (1968). Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin*, *69*(1), 69–73. doi:10.1037/h0025263

Lipsey, M. W. (1997). Using linked meta-analysis to build policy models. In W. J. Bukoski (Ed.), *Meta-analysis of drug abuse prevention programs* (pp. 216–233). National Institute on Drug Abuse. doi:10.1037/e495592006-012

López-López, J. A., den Noortgate, W. V., Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A monte carlo simulation. *Research Synthesis Methods*, *8*(4), 435–450. doi:10.1002/jrsm.1245

Mendoza, J. L., Hart, D. E., & Powell, A. (1991). A bootstrap confidence interval based on a correlation corrected for range restriction. *Multivariate Behavioral Research*, *26*(2), 255–269. doi:10.1207/s15327906mbr2602_4

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, *93*(5), 1042–1052. doi:10.1037/0021-9010.93.5.1042

Millsap, R. E. (1989). Sampling variance in the correlation coefficient under range restriction: A monte carlo study. *Journal of Applied Psychology*, *74*(3), 456–461. doi:10.1037/0021-9010.74.3.456

Murphy, K. R. (Ed.). (2003). *Validity generalization: A critical review*. NJ: Lawrence Erlbaum Associates.

National Center for Education Statistics. (2017). Digest of education statistics: 2017. Retrieved from https://nces.ed.gov/programs/digest/d17/

Olkin, I. (1981). Range restrictions for product-moment correlation matrices. *Psychometrika*, *46*(4), 469–472. doi:10.1007/bf02293804

Olkin, I. & Finn, J. D. (1990). Testing correlated correlations. *Psychological Bulletin*, *108*(2), 330–333. doi:10.1037/0033-2909.108.2.330

Olkin, I. & Siotani, M. (1976). Asymptotic distributions of functions of a correlation matrix. In S. Ikeda (Ed.), *Essays in probability and statistics : A volume in honor of professor junjiro ogawa* (pp. 235–251). Tokyo: Shinko Tsusho.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2017). Realizing the full potential of psychometric meta-analysis for a cumulative science and practice of human resource management. *Human Resource Management Review*, *27*(1), 201–215. doi:10.1016/j.hrmr.2016.09.011

Paris, B. (2019, May 20). The new SAT score: So many questions. *Inside Higher Education*. Retrieved from https://www.insidehighered.com/

Park, S. & Beretvas, S. N. (2018). Synthesizing effects for multiple outcomes per study using robust variance estimation versus the three-level model. *Behavior Research Methods*, *51*(1), 152–171. doi:10.3758/s13428-018-1156-y

Patsopoulous, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *Journal of the American Medical Association*, *293*(19), 2362–2366.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. XI. on the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of*

the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, *200*, 1–66. doi:10.1098/rsta.1903.0001

Pearson, K. & Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. on the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *191*, 229–311. doi:10.1098/rsta.1898.0007

Powers, D. E. (2004). Validity of Graduate Record Examinations (GRE®) general test scores for admissions to colleges of veterinary medicine. *Journal of Applied Psychology*, *89*(2), 208–219. doi:10.1037/0021-9010.89.2.208

Premack, S. L. & Hunter, J. E. (1988). Individual unionization decisions. *Psychological Bulletin*, *103*(2), 223–234. doi:10.1037/0033-2909.103.2.223

Pustejovsky, J. E. (2018). The multivariate delta method. Retrieved from https://www.jepusto.com/multivariate-delta-method/

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Raju, N. S. & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement*, *27*(1), 52–71. doi:10.1177/0146621602239476

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295–315). New York: Russell Sage Foundation.

Raudenbush, S. W., Becker, B. J., & Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin*, *103*(1), 111–120. doi:10.1037/0033-2909.103.1.111

Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, *79*, 298–301. doi:10.1037//0021-9010.79.2.298

Riley, R. D. (2009). Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(4), 179–811. doi:10.1111/j.1467-985x.2008.00593.x

Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, *340*(c221), 1–7. doi:10.1136/bmj.c221

Riley, R. D., Thompson, J. R., & Abrams, K. R. (2007). An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, *9*(1), 172–186. doi:10.1093/biostatistics/kxm023

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? a meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. doi:10.1037/0033-2909.130.2.261

Robbins, S. B., Oh, I.-S., Le, H., & Button, C. (2009). Intervention effects on college performance and retention as mediated by motivational, emotional, and social control factors: Integrated meta-analytic path analyses. *Journal of Applied Psychology*, *94*(5), 1163–1184. doi:10.1037/a0015738

Roh, H., Chun, K., Ryou, Y., & Son, J. (2018). Opening the black box: A meta-analytic examination of the effects of top management team diversity on emergent team processes and multilevel contextual influence. *Group & Organization Management*, *44*(1), 112–164. doi:10.1177/1059601118799475

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, *85*(1), 185–193. doi:10.1037/0033-2909.85.1.185

Rosenthal, R. & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences*, *1*(03), 377. doi:10.1017/s0140525x00075506

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, *135*(1), 1–22. doi:10.1037/a0013978

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. doi:10.1037/0021-9010.92.2.538

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, *56*(4), 302–318. doi:10.1037/0003-066x.56.4.302

Sackett, P. R. & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Pscychology*, *85*(1), 112–118. doi:10.1037//0021-9010.85.1.112

Schmidt, F. L. (1992). What do data really mean? research findings, meta-analysis, and cumulative knowledge in psychology. *American Pscyhologist*, *47*(10), 1173–1181. doi:10.1037/0003-066x.47.10.1173

Schmidt, F. L. & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 529–540. doi:10.1037/0021-9010.62.5.529

Schmidt, F. L. & Hunter, J. E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1977–2002. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31–66). Mahwah, NJ: Lawrence Erlbaum Associates.

Schmidt, F. L. & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). SAGE Publications. doi:10.4135/9781483398105

Schmidt, F. L., Le, H., & Oh, I.-S. (2009). Correcting for the distorting effects of study artifacts in meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 317–336). New York: Russell Sage Foundation.

Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, *59*(2), 281–305. doi:10.1111/j.1744-6570.2006.00065.x

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*(6), 1479–1497. doi:10.1037/a0016810

Schram, C. M. (1995). *A comparison of methods for correcting multivariate data for attenuation with application to synthesizing correlation matrices.* (Doctoral dissertation). Retrieved from ProQuest Dissertation & Theses Global (9605938).

Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues. *Psychological Methods*, *1*(1), 47–65. doi:10.1037/1082-989x.1.1.47

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth, Cengage Learning.

Shadish, W. R. & Lecy, J. D. (2015). The meta-analytic big bang. *Research Synthesis Methods*, *6*(3), 246–264. doi:10.1002/jrsm.1132

Smith, M. L. & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. doi:10.1037/0003-066x.32.9.752

Stanley, J. C. & Wang, M. D. (1969). Restrictions on the possible values of r12, given r13 and r23. *Educational and Psychological Measurement*, *29*(3), 579–581. doi:10.1177/001316446902900304

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, *15*(3), 335–352. doi:10.1207/s15327906mbr1503_7

Stewart, L. A. & Tierney, J. F. (2002). To IPD or not to IPD? advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions*, *25*(1), 76–97. doi:10.1177/0163278702025001006

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. doi:10.1007/s40865-016-0026-5

The College Board. (2019). Environmental context dashboard. Retrieved from https://professionals.collegeboard.org

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York: John Wiley & Sons.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. doi:10.1037/met0000011

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods*. doi:10.1002/jrsm.1339

Tracz, S. M., Elmore, P. B., & Pohlmann, J. T. (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educational and Psychological Measurement*, *52*(4), 879–888. doi:10.1177/0013164492052004007

Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with s* (4th ed.). ISBN 0-387-95457-0. New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Viswesvaran, C. & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, *48*(4), 865–885. doi:10.1111/j.1744-6570.1995.tb01784.x

Waschull, S. B. (2018). Improving developmental education reform in florida. *New Directions for Community Colleges*, *2018*(182), 75–83. doi:10.1002/cc.20303

Wei, Y. & Higgins, J. P. (2012). Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in Medicine*, *32*(7), 1191–1205. doi:10.1002/sim.5679

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores,

high school grades, and SES. *Educational Assessment*, *20*(1), 23–45. doi:10.1080/10627197.2015.997614

Wexler, N. (2019, May 26). Why outrage over the new SAT 'adversity score' is misplaced. *Inside Higher Education*. Retrieved from https://www.forbes.com/

Whiteside, M. F. & Becker, B. J. (2000). Parental factors and the young child's postdivorce adjustment: A meta-analysis with implications for parenting arrangements. *Journal of Family Psychology*, *14*(1), 5–26. doi:10.1037//0893-3200.14.1.5

Wothke, W. (1993). Nonpositive definite matrices in structural equation modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 256–293). Newbury Park, CA: SAGE Publications.

Zwick, R. & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, *48*(2), 101–121. doi:10.1111/j.1745-3984.2011.00136.x