



eCOMMONS

Loyola University Chicago  
Loyola eCommons

---

Dissertations

Theses and Dissertations

---

2012

## Using Robust Standard Errors to Combine Multiple Regression Estimates with Meta-Analysis

Ryan T. Williams  
*Loyola University Chicago*

Follow this and additional works at: [https://ecommons.luc.edu/luc\\_diss](https://ecommons.luc.edu/luc_diss)

 Part of the [Quantitative Psychology Commons](#)

---

### Recommended Citation

Williams, Ryan T., "Using Robust Standard Errors to Combine Multiple Regression Estimates with Meta-Analysis" (2012). *Dissertations*. 405.

[https://ecommons.luc.edu/luc\\_diss/405](https://ecommons.luc.edu/luc_diss/405)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).  
Copyright © 2012 Ryan T. Williams

LOYOLA UNIVERSITY CHICAGO

USING ROBUST STANDARD ERRORS TO COMBINE MULTIPLE  
REGRESSION ESTIMATES WITH META-ANALYSIS

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE GRADUATE SCHOOL  
IN CANDIDACY FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

PROGRAM IN RESEARCH METHODOLOGY

BY

RYAN T. WILLIAMS

CHICAGO, IL

AUGUST 2012

Copyright by Ryan T. Williams, 2012  
All rights reserved.

## ACKNOWLEDGEMENTS

To Adrienne, I owe more than I can express. Your endured mellifluous effect has enabled me to reach those goals that once seemed so vastly unattainable. You continue to be the source of my curiosity and you are the love of my life, already making me more accomplished than this paper ever will.

I owe thanks to my parents, Charles and Helen. You patiently answered my questions, however bizarre or unusual. You responded to my first hypotheses and you never discouraged my interests. Most importantly, you fostered intellectual freedom. You provided all the conditions necessary for me to truly think.

I owe much to my committee members, Meng-Jia Wu and Beth Tipton. I am thankful for all I have learned from Drs. Wu and Tipton throughout this process and throughout my graduate studies. I am sure these two remarkable scientists will continue to teach me many new things throughout my career.

I want to thank all of my colleagues at Loyola University Chicago, Rehabilitation Institutes of Chicago (RIC), and American Institutes for Research (AIR) who have challenged my thinking, in such wonderful ways, throughout my studies including (but not limited to) Josh Polanin, Allen Heinemann, Jared Eno, Andrew Swanlund, Kelly Hallberg, Shazia Miller, Larry Friedman, and Arie van der Ploeg.

For translating the obscure and the abstract into something palpable, I owe a great deal of thanks to Timothy O'Brien, the best statistics teacher I have ever had.

I owe the greatest thanks to Terri Pigott. Terri embodies the most important characteristics of a mentor. She was patient. She listened to my many, many unformulated questions and hypotheses and helped me turn them into something tangible and productive. Terri was the authoritative parent throughout my studies. She allowed me to wander in my thought and when I ran into trouble, panicked, or for whatever reason lost my footing, she was there to help me find the right answer. Terri also provided me with many generous opportunities to become part of her community of scholars. Most importantly, Terri always saw the big picture and I cannot thank her enough for helping me see it, too. While this chapter of my life ends, Terri will always be my mentor.

If we set out to be methodologists, we set out to be experts in problems and, hopefully,  
inventors of solutions.

*Donald T. Campbell*

To Adrienne

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
ABSTRACT .....	xi
CHAPTER ONE: INTRODUCTION .....	1
CHAPTER TWO: LITERATURE REVIEW .....	3
Meta-analysis .....	3
Effect sizes .....	6
Multivariate models and meta-analysis .....	13
Summary .....	47
CHAPTER THREE: METHODS .....	50
Robust variance estimation and meta-analysis .....	50
Analyses .....	55
Applications .....	55
Simulation studies .....	59
CHAPTER FOUR: RESULTS .....	68
Phase I results .....	68
Model-based example .....	68
Focal slope example .....	70
Phase II Results .....	72
Model-based approach .....	72
Focal slope approach .....	80
CHAPTER FIVE: DISCUSSION .....	89
Conclusions .....	89
Limitations .....	91
Future directions .....	93
APPENDIX A: SIMULATION CODE .....	94
REFERENCES .....	103
VITA .....	108



## LIST OF TABLES

1. Table 1. Regression Model Results with Orthogonal Predictors	17
2. Table 2. Regression Model Results with Correlated Predictors	18
3. Table 3. Link and Mulligan (1986) Sample Characteristics	69
4. Table 4. Meta-Regression Analysis of Link and Mulligan (1986)	70
5. Table 5. Focal Slope Meta-Analysis Study Characteristics	71
6. Table 6. Meta-Regression Analysis of Link and Mulligan (1986)	72
7. Table 7. Model-Based Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 3$	73
8. Table 8. Model-Based Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 6$	74
9. Table 9. Model-Based Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 9$	74
10. Table 10. Model-Based Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 3$	75
11. Table 11. Model-Based Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 6$	76
12. Table 12. Model-Based Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 9$	76
13. Table 13. Model-Based Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 3$	77
14. Table 14. Model-Based Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 6$	78
15. Table 15. Model-Based Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 9$	79
16. Table 16. Focal Slope Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 3$	81
17. Table 17. Focal Slope Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 6$	82
18. Table 18. Focal Slope Approach: $\tau^2 = 0 \times v$ ; $I^2 = 0$ ; $k = 9$	82
19. Table 19. Focal Slope Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 3$	83

20. Table 20. Focal Slope Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 6$	84
21. Table 21. Focal Slope Approach: $\tau^2 = .5 \times v$ ; $I^2 = .33$ ; $k = 9$	84
22. Table 22. Focal Slope Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 3$	85
23. Table 23. Focal Slope Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 6$	86
24. Table 24. Focal Slope Approach: $\tau^2 = 1 \times v$ ; $I^2 = .5$ ; $k = 9$	86

## LIST OF FIGURES

1. Figure 1. Number of Meta-Analytic Research Papers in PsychInfo, 1995-2010	5
2. Figure 2. Number of Meta-Analytic Research Papers in PubMed, 1995-2010	6
3. Figure 3. Venn Diagram of Collinearity in Focal Predictor	16
4. Figure 4. Distributions of Focal Slope Estimates Under Different Collinearity Conditions	19
5. Figure 5. Exclusionary Steps in Focal Slope Meta-Analysis Example	58
6. Figure 6. Simulation Data Generation and Analysis Steps	67

## ABSTRACT

Combining multiple regression estimates with meta-analysis has continued to be a difficult task. A variety of methods have been proposed and used to combine multiple regression slope estimates with meta-analysis, however, most of these methods have serious methodological and practical limitations. The purpose of this study was to explore the use of robust variance estimation for combining commonly specified multiple regression models and for combining sample-dependent focal slope estimates from diversely specified models. A series of Monte-Carlo simulations were conducted to investigate the performance of a robust variance estimator for each of these approaches. Key meta-analytic parameters were varied throughout the process. Also, two small scale, examples were conducted to illustrate the use of the robust variance estimator in each of these two approaches. In general, the robust variance estimator performed well. Robust confidence interval parameter recovery was close to the specified 95% under almost all conditions. Only when there were a larger number of slope estimates and a small number of study samples did the robust standard errors noticeably lose efficiency. Combining sample-dependent focal slope estimates provides biased point estimates, however, the results of this paper suggest that the robust standard errors are still accurate.

## CHAPTER ONE

### INTRODUCTION

Methods for meta-analysis have made great strides over the past 30 years. The focus of this work has been primarily on synthesizing univariate and bivariate effect size estimates. Increasingly, however, multivariate models such as multivariate linear models, factor analyses, and latent trait models are used to describe relationships in the social sciences and in medicine. Meta-analytic methods to combine effect sizes from these kinds of models are often cumbersome to implement or nonexistent.

Multiple regression is one of the most commonly used statistical methods in the social sciences and in medicine. Because of the multivariate nature of multiple regression coefficients, they have historically been very difficult to combine with meta-analysis. Becker and Wu (2007) outline three key difficulties in combining multiple regression slope estimates. First, all model outcomes must be measured on a common scale. Second, the slope estimate of interest (focal slope) is measured on a common scale across studies. Finally, each study estimates the partial relationship between the focal slope and the outcome using the model (i.e. includes an identical set of additional predictors). Maintaining these assumptions in any given synthesis will almost always be impossible.

These stringent assumptions have not deterred many researchers from combining multiple regression estimates with meta-analysis. The majority of this work has focused on univariate meta-analysis of focal slopes (i.e. slope estimates of a particularly

interesting relationship). A diverse set of approaches have been used, including ordinary least squares modeling of focal slope estimates and weighted least squared modeling of focal slope estimates. Difficulties arise when focal slope estimates arise from diversely specified models, which lead to heterogeneous parameter estimation. Other researchers have used more elegant multivariate methods to summarize full, but commonly specified regression models (i.e. each model contains the same set of predictor variables). The information needed to conduct this type of multivariate synthesis, however, is almost never available.

The purpose of this study was to explore using robust variance estimation for meta-analyzing multiple regression estimates. The proposed estimator obviates traditionally required information about the covariance structure of the dependent effect size estimates, making it a potentially flexible method for conducting meta-analyses of regression estimates.

A series of applied examples were conducted using studies from a database of multiple regression studies estimating the partial effects of per-pupil expenditures on student achievement. Also, a series of Monte Carlo simulations were conducted to explore the performance of the robust variance estimator under different meta-analytic conditions.

## CHAPTER TWO

### LITERATURE REVIEW

This chapter reviews methods used for combining regression estimates with meta-analysis in micro economics, epidemiology, ecology, psychology, and education. A historical context is provided, highlighting the strengths and the weaknesses of these methods.

#### **Meta-analysis**

Researchers from various academic fields have worked, at times simultaneously, to develop and refine methods for combining scientific findings. Meta-analysis is set of statistical methods used to combine results from independent primary studies.

Throughout this paper, in keeping with the Glassian tradition, “meta-analysis” is used synonymously with “research synthesis”, “systematic review”, and “quantitative review”.

Gene Glass coined the term “meta-analysis” in his 1976 address to the American Educational Research Association (Glass, 1976). While Glass was the first to define the basic quantitative methods for combining independent study results, Robert Rosenthal, John Hunter, and Frank Schmidt were also working to develop methods for research synthesis at that time.

In the late 1970's and early 1980's, meta-analysis was viewed with optimism, passive skepticism, and even derision. Most vocally from the opposition was Hans Eysenck who published a 1978 paper entitled *An Exercise in Mega-Silliness*. In its inception, criticisms of the methodological integrity of meta-analysis were not uncommon (e.g. Slavin, 1986) and continue today (e.g. Bonett, 2009).

Despite the initial pessimism surrounding meta-analytic research methods, several persistent methodologists including (but certainly not limited to) Larry Hedges, Ingram Olkin, Frank Schmidt, John Hunter, Harris Cooper and Robert Rosenthal launched meta-analysis into the 21st century. A small group of researchers in Great Britain were also working to advance the practice of meta-analysis in medicine during the same period.

The popularity of meta-analysis in the social, behavioral, and biomedical sciences has grown over the past 35 years. Research activity in meta-analysis has surged in recent history (Cooper & Hedges, 2009). A simple search of PsychInfo shows a dramatic increase in the number of publications with “research synthesis”, “research review”, “systematic review”, or “meta-analysis” in their title from the years 1995 to 2010. Figure 1 depicts this increase and Figure 2 presents the results of a parallel search in PubMed. In both databases, the number of meta-analytic activity is increasing exponentially. The value of and demand for meta-analytic research is greater than ever before, underscoring the transformative capacity of the enterprise.



Figure 1. Number of Meta-Analytic Research Papers in PsychInfo, 1995-2010

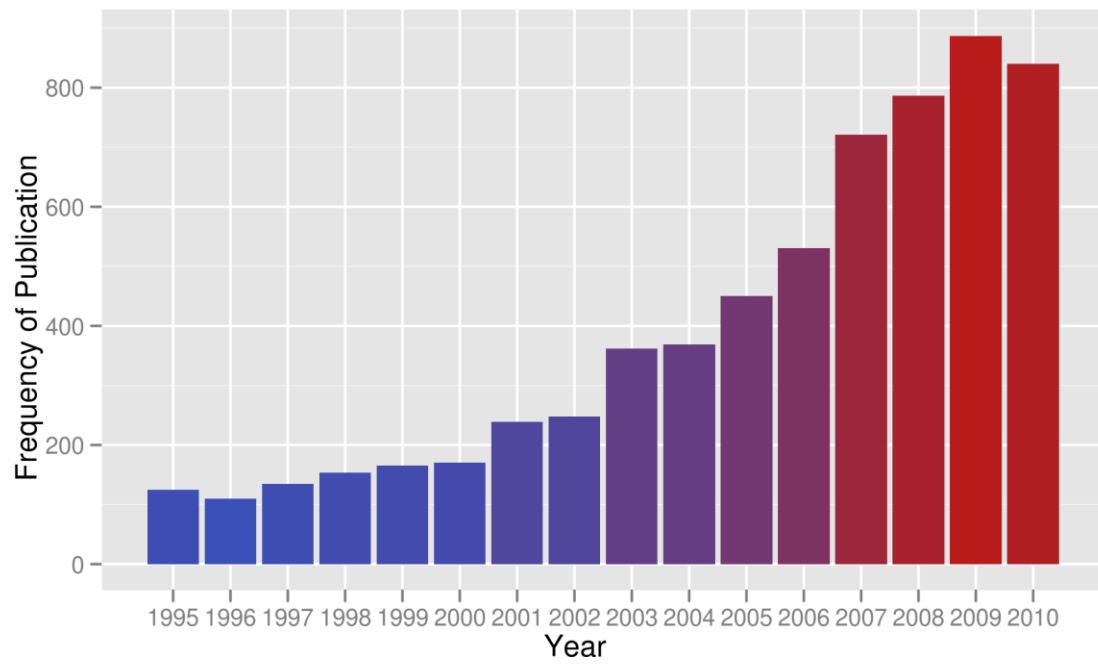
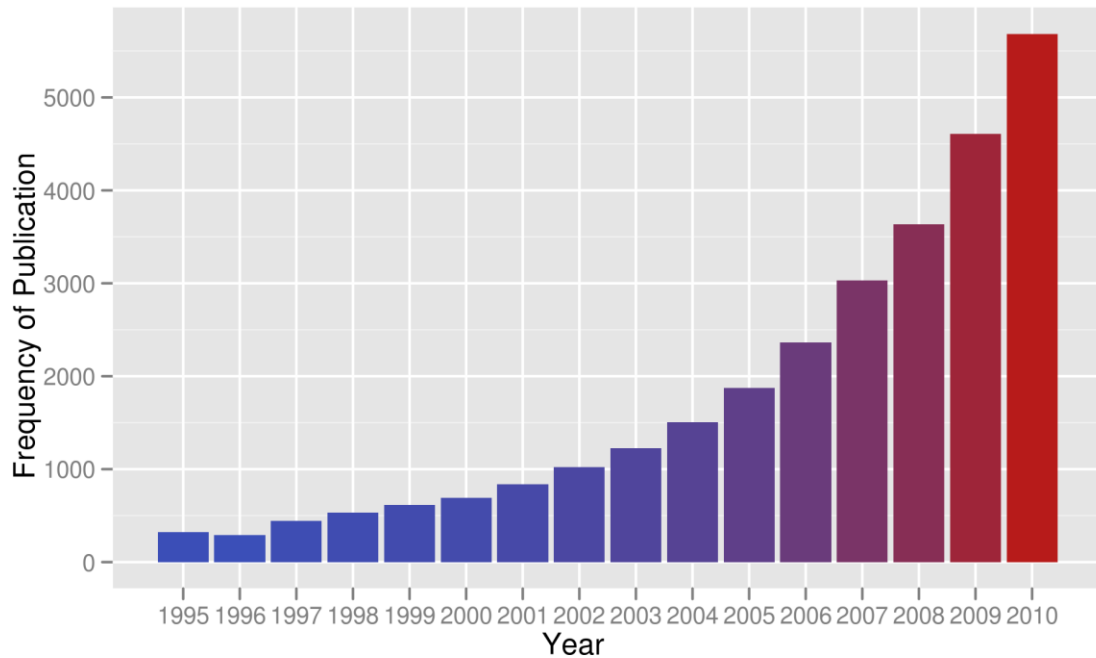


Figure 2. Number of Meta-Analytic Research Papers in PubMed, 1995-2010



**Effect sizes.** A hallmark attribute of meta-analysis is the effect size estimate, which summarizes the direction and magnitude of association between independent and dependent variables. Glass, McGaw, and Smith (1981) originally defined the effect size using the following linear model formulation

$$T_k = \theta + e_k, \quad (1)$$

where  $T$  is the  $k$ th effect size estimate of a common parameter,  $\theta$ , and  $e_k$  is the estimation residual such that  $e \sim N(0, \sigma_T^2)$  and thus  $T \sim N(\theta, \sigma_T^2)$ , where  $\sigma_T^2$  is an unknown variance.

Hedges and Olkin (1985) first proposed weighting the effect size estimate  $T$  by the inverse of its variance. Inverse variance weighting of effect size estimates is standard practice, especially in the social and behavioral sciences.

The effect size is the key to synthesizing quantitative research findings. Studies are rarely replicated in a manner that they share common variables and scales with which to measure the constructs of interest. Simply ignoring the incomparability of measurement scales across studies would yield meaningless and almost certainly incorrect results. Meta-analysis uses the effect size as a means of standardizing study findings so that quantitative results are consistent across all variables and measures involved (Lipsey & Wilson, 2001). In this way, effects sizes representing comparable study constructs, with different operationalizations may be combined. For example, many crime and justice researchers study the effects of different social programs on recidivism. Recidivism may be operationalized as re-arrest, conviction, probation violations, and so on. Assuming each of these potential outcomes represent the same construct, recidivism in the example above, effect size estimation makes possible a meta-analysis of program effects.

In addition to providing a means of comparing variables across studies, effect sizes are well-suited for meta-analysis because, unlike statistical tests, effect sizes are unit-independent. Large effect size estimates may be derived from studies with small samples and conversely, small effect size estimates may be derived from studies with large samples. Error terms and confidence intervals in effect size estimates, however, will vary with sample size and for this reason effect size estimates are weighted by their precision in most meta-analytic research (e.g. by the inverse of their variance). Meta-

analysis, as presented by Hedges and Olkin (1985) fundamentally relies on weighted least squares (WLS) estimation.

Meta-analytic research by necessity constrains itself to a specific set of relationships because the distribution of effect sizes used is assumed known and constant across all estimates. For example, one meta-analysis may investigate the effects of beta blockers on blood pressure. In this scenario, both the independent and dependent variables of interest have fairly clear definitions<sup>1</sup>. Another meta-analysis may investigate the relationship between socioeconomic status and academic achievement. The constructs of interest in this study, socioeconomic status and academic achievement, may be considerably more nebulous than those in the first example. For example, socioeconomic status may be operationalized as family income, eligibility for a free or reduced-price school lunch, average neighborhood home value and so on. Furthermore, academic achievement may also be operationalized as test score performance, grade point averages, graduation, or college enrollment to name a few. In the social and behavioral sciences especially, there is often at least some variability in dependent and independent variables within a meta-analysis. While measures may commonly vary across studies, the constructs themselves should not, which then leads to, as much as is possible, a homogenous set of effect size estimate with a common distribution.

Thus far, I have discussed only as a set of statistical procedures for combining primary study results. Undertaking a meta-analysis is, without doubt, a much larger

---

<sup>1</sup> Of course, blood pressure could be measured many ways. Additionally, many drugs are classified as beta-blockers and dosages for each of these vary.

endeavor. Cooper (2009) outlines nine steps intended to increase the quality of a meta-analysis: problem formulation; systematic search of the literature base; gathering information from studies; evaluating the quality of studies; analyzing and interpreting the evidence; and presenting the results. Cooper's steps distinguish meta-analysis from traditional narrative reviews. And, when followed, each of these steps increases the transparent and systematic nature of meta-analysis.

Social scientists have historically used meta-analysis to combine the effects of social, educational and psychological interventions in education and psychology and medical researchers have historically used meta-analysis to combine findings from randomized controlled clinical trials (RCTs) in the medicine. This comes without surprise given that the origin of the method is rooted in intervention research (Shadish, et al., 1993; Smith, Glass, & Miller, 1992; Smith & Glass, 1977). Many effect size estimators, however, are available for meta-analysis, depending on the nature of the research question and the structure of the primary data. Effect size estimates may represent group differences (dependent or independent) or even bivariate correlations. Meta-analysis of univariate effect sizes such as the arithmetic mean or a sample proportion may also be appropriate for certain research questions.

In hopes of providing a small foundation for this line of research, the following two sections outline some basic methods for meta-analyzing group mean differences for continuous data as well as bivariate correlations.

**Combining mean differences.** One of the most common measures of effect for continuous data is the mean difference. Borenstein (2009) shows that the unstandardized mean difference between two independent groups (e.g. treatment and control) is estimated with

$$D = \bar{x}_1 + \bar{x}_2 , \quad (2)$$

where  $\bar{x}_1$  and  $\bar{x}_2$  are the mean values of the study outcome for groups one and two, respectively. The variance estimate for  $D$  is

$$\hat{\sigma}_D^2 = \frac{n_1 + n_2}{n_1 n_2} \hat{\sigma}_p^2 \quad (3)$$

where  $\hat{\sigma}_p^2$  is the estimated pooled variance between groups one and two such that

$$\hat{\sigma}_p^2 = \frac{(n_1 - 1) \hat{\sigma}_1^2 + (n_2 - 1) \hat{\sigma}_2^2}{n_1 + n_2 - 2} . \quad (4)$$

When outcome measures vary in scale, a standardized mean difference is appropriate and may be estimated using

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_p} . \quad (5)$$

Because  $d$  is upwardly biased, Hedges (1981) provided an unbiased estimator: Hedges'  $g$ , which is computed with

$$g = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_p} . \quad (6)$$

where

$$\hat{\sigma}^* = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}}. \quad (7)$$

To adjust for the remaining upward bias in (6) and (7), Hedges (1981) proposed the following correction for  $g$

$$g^* = 1 - \left( \frac{3}{4n - 9} \right) g. \quad (8)$$

These mean difference effect size estimators assume independent samples; however, effect size estimates can certainly be derived from dependent or matched samples. Furthermore, effect size estimates for binary outcome measures may also be estimated via the odds ratio, risk ratio, or difference of proportions. The details about these mean difference effect size estimators are covered extensively by Borenstein (2009), as well as Lipsey and Wilson (2001). Fleiss and Berlin (2009) cover extensively methods for estimating group differences with binary outcomes.

***Combining correlations.*** While meta-analysis is frequently used to study treatment effects, correlational (observational) data are also viable effect size estimates (Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Rosenthal, 1991). Correlations are standardized covariance estimates and have an absolute value ranging from zero to one. A common effect size estimate in meta-analytic research, the Pearson product moment correlation, is defined as

$$r = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}, \quad (9)$$

where  $\sigma_{xy}^2$  is the covariance between  $x$  and  $y$  and  $\sigma_x$  and  $\sigma_y$  are the standard deviations of  $x$  and  $y$ , respectively. The variance of  $r_{xy}$  is given by

$$\hat{\sigma}_r^2 = \frac{(1 - r^2)^2}{n - 1}. \quad (10)$$

Because the distribution of the correlation coefficient is known, one can combine multiple coefficients in a meta-analysis. Just as with meta-analysis of mean differences, it is assumed that each study uses the same two variables in the estimation of the bivariate correlation of interest. And, it is not uncommon for independent and dependent variable measures to vary. It is assumed, however, that the constructs represented by each effect size estimate are invariant. When bivariate correlations are available, they provide convenient and even optimal effect size estimates that typically require little adjustment.

The sampling distribution of the Pearson product-moment correlation is negatively skewed as  $r$  approaches 1. As such, Fisher's  $Z$  transformation is commonly used as a variance stabilizing logarithmic transformation, which is:

$$r_z = .5 \log \frac{(1 + r)}{(1 - r)}, \quad (11)$$

with variance

$$\hat{\sigma}_{r_z}^2 = \frac{1}{n - 3} \quad (12)$$

where  $n$  is the sample size.



The estimators presented above are appropriate only for continuous data.

Tetrachoric and polychoric correlations may also be used, but as pointed out by Hunter and Schmidt (2004), these estimators will tend to underestimate the population value.

**Multivariate models and meta-analysis.** Multivariate statistical methods are increasingly common. In education for example, hierarchical linear modeling (HLM; e.g. Raudenbush & Bryk, 2002) has quickly become a favored modeling technique. In psychology, structural equation modeling and factor analysis are popular techniques. The meta-analytic methods presented in the above sections are appropriate only for bivariate relationships, leaving a gap in the methods available to synthesize a literature base.

Meta-analysis of multivariate relationships is exceedingly complex. Effect size estimates derived from multivariate data represent partial effects because the relationship between any two variables is, almost always, at least partially dependent on other variables included in the model. The next section introduces the use the multiple regression coefficient and discusses the difficulty of combining these multivariate estimates in a meta-analysis. The issues discussed below generalize to many multivariate modeling techniques.

***Combining multiple regression coefficients.*** Meta-analysts have made use of simple and multiple regression models. Simple regression models, estimate the change in outcome for each unit change in a single predictor variable. Multiple regression is a multivariate statistical technique that examines the change in outcome for each unit

change in a predictor variable, holding constant the effects of additional predictors.

Ordinary least squares (OLS) regression takes the basic form of

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots \beta_k X_{ki} + \varepsilon_i , \quad (13)$$

where  $X_{1i}$  to  $X_{ki}$  are independent variables predicting outcome  $Y$  and  $\varepsilon_i$  is the estimate residual from the  $i$ th predicted value of  $Y$ . Many terms exist to describe independent and dependent variables in multiple regression analyses such as exogenous and endogenous variables, predictor and outcome variables, predictor and response variables, and so on. Throughout the remainder of this paper, the term “predictor” will be used synonymously with exogenous and independent variables and the term “outcome” will be used synonymously with endogenous, response, and dependent variables.

Standardized simple regression coefficients may be combined with bivariate correlations in a meta-analysis. For example, education researchers may be interested in meta-analyzing the relationship of per-pupil expenditure (PPE) on academic achievement. This relationship may be represented by a bivariate correlation, as discussed above, or the relationship could be represented in a simple regression model such as:

$$Y_i = \beta_0 + \beta_1 PPE_i + \varepsilon_i , \quad (14)$$

where  $Y_i$  is the  $i$ th student’s achievement score,  $\beta_1$  is the slope estimate for PPE (i.e. the change in achievement for each unit increase in PPE). When  $\beta_1$  is standardized it is equivalent to the bivariate correlation.

The focal relationship, such as PPE and academic achievement, may commonly be estimated in the context of a more complex multivariate linear model. For example, the relationship between PPE and academic achievement may be included in a model that also accounts for teacher experience:

$$Y_i = \beta_0 + \beta_1 PPE_i + \beta_2 TeachExp_i + \varepsilon_i, \quad (15)$$

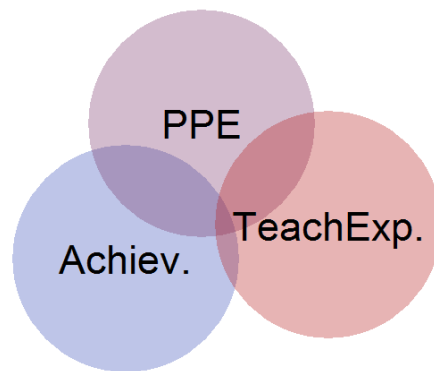
where  $\beta_1$  is now the slope estimate of PPE, holding constant the effects of teacher experience and  $\beta_2$  is the slope estimate for teacher experience holding constant the effects of PEE. This simple example illustrates how, by including even one additional predictor variable, the meaning of the effect size estimate can change. And it is not uncommon for multiple regression models to include many more than two predictor variables. This is one reason that meta-analysis of multiple regression coefficients is particularly difficult.

The structure and interpretation of multiple regression estimates has two important implications for meta-analyses of regression coefficients. First, the scale of the predictor variable may vary across samples. For example, in meta-analyzing the relationship between PPE and academic achievement, academic achievement may be scaled differently across all studies. Standardization is often not a viable option in this context. Slope estimates that are not reported in their standardized form may not be easily standardized by the meta-analyst. And, even if standardized slope estimate are reported, there is little guarantee that they will provide meaningful information to end-users (Greenland, Schlesselman, & Criqui, 1986). For this reason, standardized regression

coefficients are still controversial (this idea is briefly revisited later). Second, and perhaps most importantly, when the focal predictor is correlated with additional predictors in the model, or the outcome, its distribution will differ from its simple bivariate, or zero-order, correlation with the outcome. Figure 3 graphical presents an example of this scenario.

This scenario is not uncommon in most social sciences regression models, and most multiple regression models include more than two predictor variables, which further complicates the covariance structure of the parameter estimates.

Figure 3. Venn Diagram of Collinearity in Focal Predictor



It is important to distinguish that the incomparability of regression slope estimates across studies from having incomparable study constructs and operationalizations. Slope estimates are frequently incomparable because of their statistical behavior in the presence of diversely specified regression models, not because the underlying constructs and their operationalizations across studies are truly different.

This behavior is further illustrated with a simple example. Assume a dataset of student achievement, PPE, and teacher experience has the following covariance matrix

$$\begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp} \\ 10.1241 & 7.6481 & -0.0721 \\ 7.6481 & 10.9043 & 0.0602 \\ -0.0721 & 0.0602 & 9.8872 \end{bmatrix}$$

and therefore the following correlation structure

$$\begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp} \\ 1.0000 & 0.7379 & -0.0072 \\ 0.7379 & 1.0000 & 0.0058 \\ -0.0072 & 0.0058 & 1.0000 \end{bmatrix}$$

Here, PPE is correlated with achievement. And teacher experience is neither correlated with achievement nor PPE. The results of regressing student achievement on PPE and teacher experience with this covariance structure (with a sample of  $N = 1000$ ) are presented in Table 1.

Table 1. Regression Model Results with Orthogonal Predictors

	Slope Estimate	Standard Error	<i>t</i>	p-value
Achiev.	0.05517	0.0693	0.7960	0.4260
PPE	0.70145	0.02092	33.5270	< 0.0001
TchExp	-0.01157	0.02197	-0.5260	0.5990

The regression model parameter estimates here are roughly equivalent to the bivariate correlations presented above. Now consider the following, perhaps more common, covariance structure where PPE is correlated with both student achievement and teacher experience

$$\begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp.} \\ 9.3071 & 6.7052 & 1.9705 \\ 6.7051 & 9.8281 & 2.2402 \\ 1.9705 & 2.2402 & 9.9455 \end{bmatrix}$$

and correlation matrix

$$\begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp.} \\ 1.0000 & 0.7011 & 0.2048 \\ 0.7011 & 1.0000 & 0.2266 \\ 0.2048 & 0.2266 & 1.0000 \end{bmatrix}$$

This covariance matrix produces the following regression results in Table 2.

Table 2. Regression Model Results with Correlated Predictors

	Slope Estimate	Standard Error	t	p-value
Achiev.	.0131	.0688	.1910	.8489
PPE	.6716	.0225	29.8280	.0001
TchExp	-.0468	.0224	2.0940	.0365

Notice how the PPE slope estimate has changed, it is no longer equal to its bivariate correlation. This was the result of PPE having a correlation of only .20 with teacher experience. If we were to repeat these processes 1000 times for data where teacher experience is correlated with PPE and not student achievement:

$$\mathbf{A} = \begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp.} \\ 1.0000 & 0.2000 & 0.0000 \\ 0.2000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

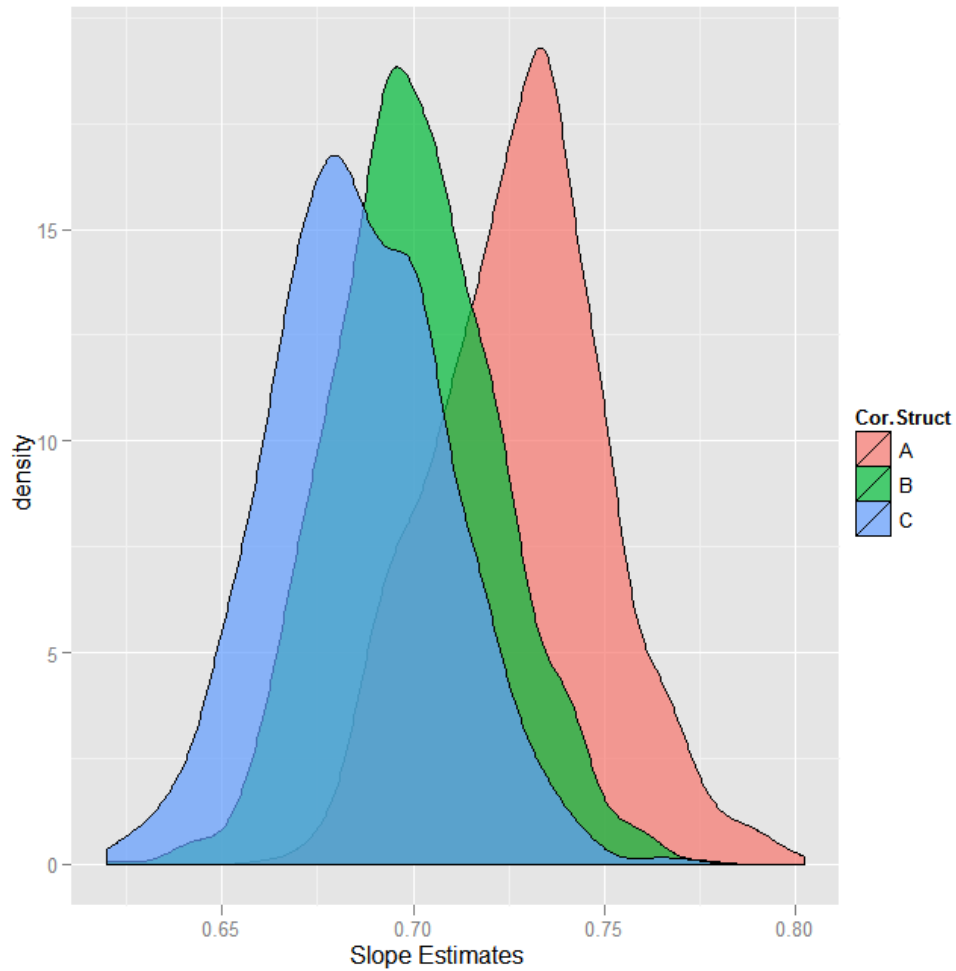
; where teacher experience is correlated with achievement and not PPE

$$\mathbf{B} = \begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp.} \\ 1.0000 & 0.0000 & 0.2000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.2000 & 0.0000 & 1.0000 \end{bmatrix}$$

; and where teacher experience is correlated with both student achievement and PPE

$$\mathbf{C} = \begin{bmatrix} \text{Achiev.} & \text{PPE} & \text{TchExp} \\ 1.0000 & 0.2000 & 0.2000 \\ 0.2000 & 1.0000 & 0.0000 \\ 0.2000 & 0.2000 & 1.0000 \end{bmatrix}$$

Figure 4. Distributions of Focal Slope Estimates Under Different Collinearity Conditions



Only when PPE is orthogonal to teacher experience (**B**) does it tend toward its bivariate correlation with student achievement (.70).

If each study in a meta-analysis specifies the same model (e.g. regressing academic achievement on PPE and teacher experience), one may proceed in applying traditional meta-analytic methods to those slope estimates (Hunter and Schmidt, 2004). This approach is considered at length later on.

The example presented above, where a focal predictor covaries to some degree with at least one other model covariate, is common. As such, meta-analysis of multiple regression estimates is exceedingly difficult. Researchers in various disciplines have nonetheless continued their work in this area and many meta-analyses using multiple regression estimates have been conducted. The following sections present some of the methods researchers have used for this problem.

*Microeconomic meta-regression analysis.* Economists have developed several approaches for meta-analyzing regression coefficients ranging from syntheses of  $t$  statistics to complex multilevel modeling techniques. The following section provides a brief history of meta-analysis in the econometric literature. Methods developed by Stanley and Jarrell (1989) are discussed followed by examples of applications of these methods. Adaptations of the methods outlined by Stanley and Jarrell (1989) are presented.

Meta-analysis in the late 1980's was burgeoning in the social, behavioral and educational sciences. Stanley and Jarrell (1989) extrapolated the methods developed by Hedges and Olkin (1985) to micro economics. Because econometric models almost never involve the comparison of randomized treatment and control groups, (Colegrave & Giles,



2008), economists are generally restricted to using meta-analysis to summarize multiple regression estimates. To apply meta-analytic techniques to econometric regression models, Stanley and Jarrell (1989) delineated what they called meta-regression analysis (MRA).

MRA in this context is not synonymous with meta-regression in the mainstream meta-analytic literature (Cooper, Hedges, and Valentine, 2009; Hedges and Pigott, 2004) where a weighted mean effect size is estimated and variation in effect size estimates is modeled with study-level predictors. Rather, this set of methods uses the following OLS approach to model variation in the observed slopes from a multiple regression model taking the form of

$$b_j = \beta_0 + \beta_1 X_{j1} + \beta_2 X_{j2} + \cdots \beta_k X_{jk} + e_j , \quad (16)$$

where  $b$  is the  $j$ th primary study's focal slope estimate,  $\beta_{j1}$  through  $\beta_{jk}$  are the estimated meta-regression coefficients of  $X_{j1}$  through  $X_{jk}$ , which are study-level predictors referencing variation in primary study model specifications or other relevant study-level characteristics such as sample size or measures of research quality. This is a fundamental characteristic of MRA in microeconomics.

We could apply this approach to the simple example of PPE and student achievement presented above. If we have a pool of regression models regressing student achievement on PPE, but some studies also include teacher experience, taking the approach from (17) we may want to model the variation in the focal slope estimates by indicating the presence or absence of teacher experience by

$$b_j(PPE) = \beta_0 + \beta(TeachExp)_j + e_j, \quad (18)$$

where *TeachExp* is a binary variable indicating the presence absence of teacher experience in the *j*th study model.

Given the variance in model specification across a pool of primary studies, Stanley and Jarrell (1989) note that model error terms using (19) will likely be heteroskedastic. The authors propose using the *t* statistic,

$$t_j = \frac{b_j}{\hat{\sigma}_b}, \quad (20)$$

where  $\hat{\sigma}_b$  is the estimated standard error of  $b_j$ .

The authors provide three arguments for using *t* statistics as effect size estimates in meta-analysis. First, the *t* statistic is often the focal statistic within primary research (i.e. the value of *t* dictates the rejection or retention of the null hypothesis). Second, the *t* statistics, unlike a regression coefficient is metric-independent; making irrelevant variable scales of measurement for *X* or *Y*. Third, since *t* is the result of the ratio of  $b_j$  to its standard error, each estimate in a meta-analysis would automatically be precision-weighted.

A major limitation in the use of *t* as an effect size index in meta-analysis is that it is not unit-independent (Becker & Wu, 2007). If used as an effect size, *t* will increase in values as its standard error decreases. The standard error of the slope will decrease as sample size increases or when observed variation in regression residuals is low. An additional hazard of employing *t* as a measure of effect (although cited as a

methodological attribute by Stanley and Jarrell (1989)) is that it relates directly to the value of  $p$  in significance testing.

Meta-regression analysis has developed as a valued econometric method. Stanley (2001), points to 16 examples of MRA studies published after Stanley and Jarrell (1989). Colegrave and Giles (2005), for example, used MRA to combine slope estimates for the relationship between optimal school size and student achievement. The authors, rather than synthesizing  $t$  statistics, extracted the raw slope estimates from 22 models, constructing a MRA model with 10 meta-predictor variables. Because the authors used observed slope estimates, they tested for the presence of heteroskedasticity with a  $\chi^2$  test. Based on this test, there was no evidence of heteroskedasticity, however it is unclear how powerful the test statistic was in their study.

Jarrell and Stanley (2004) (updated from Jarrell & Stanley, 1990; Stanley & Jarrell, 1998) used MRA methods to combine estimates of the effects of gender on wage. They uncovered a total of 104 regression estimates and combined the slope estimates from a diverse set of models. The authors used heteroskedasticity-robust standard errors (White, 1980) but give little detail as to how they applied those standard errors to their research.

Card and Krueger (1995) modeled log-based transformations of  $t$  statistics of regression slope estimates of the relationship between minimum wages and employment. The authors found a strong negative correlation between the observed  $t$  statistics in their meta-analysis and study sample sizes.

Variations of the methods established by Stanley and Jarrel (1989) continue to be used in microeconomics research. Some researchers have modified this set of methods to investigate various substantive and methodological areas such as publication bias (Doucouliagos, 2005; Stanley & Doucouliagos, 2010).

Weichselbaumer and Winter-Ebmer (2005) used MRA to combine 1,535 slope estimates from 788 international studies (two estimates per study on average) of the relationship between gender and salary. While the combination of multiple dependent effect size estimates will bias cumulative effect size estimate standard errors (Hedges, Tipton, & Johnson, 2010), the authors state that they used a weighting and clustering technique to adjust their standard errors for underestimation. In a two-step process, the authors divided each of the dependent slopes by the total number of estimates extracted from that study and used a “clustering approach” to further correct for upward bias. The authors cite Froot (1989) who proposed the use of cluster- and heteroskedastic-robust standard error estimation in analyzing time series data when alternative methods are impractical (i.e. generalized least squares). It is, however, unclear how these methods were applied to their meta-analysis. The methods used by Weichselbaumer and Winter-Ebmer (2005) may have promising implications for the field of meta-analysis of regression estimates but the authors do not provide sufficient detail around their methods for them to be practically useful to other researchers. Furthermore, the statistical performance of the estimators they used is unknown.

Doucouliafos and Laroche (2009) used a derivative of MRA to synthesize the effects of labor unions on profit. The authors identified the partial correlation as the target effect size estimate and model observed variation in these effect such that

$$r_{jw} = \sum N_j \frac{r_j}{\sum N_j}, \quad (21)$$

where  $r$  the  $j$ th partial correlation between the focal slope and the outcome and  $N$  is the  $j$ th study's sample size. Doucouliafos and Laroche (2009) used the following equation to model variation in the mean effect size

$$r_j = \beta_0 + \sum \beta_k X_{jk} + \sum \beta_n K_{jn} + \varepsilon_j, \quad (22)$$

where  $X$  is the  $k$ th binary indicator representing variations in model specification or dichotomous study characteristics and  $K$  is the  $n$ th covariate representing variation in continuous study characteristics.

Doucouliafos & Laroche (2009) address multiply dependent effect size estimates (e.g. multiple slope estimates per study) by using a two-level random intercepts model such that intercept estimates from (20) enter

$$\beta_0 = \gamma_{00} + u_{0i} + \varepsilon_{ij}, \quad (23)$$

where  $u_{0j}$  is the estimated between-study variance of the level-two model and  $\varepsilon_{ij}$  is the level-one model residual. The authors also used a fixed effects model with clustered standard errors to present the results alongside those from.

Microeconomists have expanded on the original work of Stanley and Jarrell (1989) that proposed the use of meta-regression analysis. Researchers in this field have

used a variety of techniques for combining regression results including simple OLS with model specification indicators, meta-analysis of  $t$  statistics, and multilevel models. At least two primary concerns are largely unaddressed. Many of the methods proposed by microeconomists are attractive given their simplicity, however, a reliance on  $t$  statistics along with common effect size dependencies (i.e. collinearity between the focal slope estimate and additional predictors) are limitations that pervade this area of research.

*Dose-response models in epidemiology.* The next section outlines methods used to synthesize epidemiologic studies. In contrast to the social sciences, binary outcomes (e.g. disease contraction) are commonplace in epidemiology. Greenland (1987) outlined meta-analytic methods for epidemiologic research. Epidemiologists typically encounter far fewer studies to meta-analyze compared to the social and behavioral research fields. Additionally, epidemiological studies are rarely eligible for an experimental design of treatment effects and thus, case-control and dose-response models with nonequivalent and dependent control groups are used frequently. Unfortunately, these study designs are not easily combined with meta-analytic methods.

Epidemiologists often study the relationship between human exposure to foreign substances, diseases, or drugs and related health outcomes. Typically, investigations like these use log-linear dose-response models to explain the relationship between predictor and outcome variables. And, most epidemiological studies investigate the relationship between multiple exposure or dosage levels and health outcomes (Greenland & Longnecker, 1992). For example, someone may conduct a study looking at the effect of

consuming different amounts of coffee on risk of myocardial infarction (heart attack).

This study may include individuals who consumed less than one or fewer cups of coffee a day, two to three cups of coffee a day, and four or more cups of coffee a day. The group of coffee drinkers who consumes one or fewer cups of coffee may be the reference group and the other groups of coffee drinkers would be compared to this reference group in terms of their risk for myocardial infarction. That is, each non-reference category effect shares the same non-reference group.

Meta-analysis of dose-response estimates require an initial within-study summary of the effects of dosage levels on the outcome, which is generally a log-based average of the multiple within-study dose effect estimates. Inverse variance weighted dose effect estimates may be obtained using

$$\hat{\beta} = \frac{\sum w_j x_j y_j}{\sum w_j x_j^2}, \quad (24)$$

where  $w_j$  is a precision weight (e.g. inverse variance),  $y_j$  is the  $j$ th log relative risk., and  $x_j$  is the exposure level (e.g. four or more cups of coffee a day). Each of  $j$  studies in a dose-response meta-analysis takes at least the simple form of

$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \mathbf{e}, \quad (25)$$

where  $\mathbf{Y}$  is a vector of estimated beta coefficients in the form of log odds, relative risk, or rate ratios;  $\mathbf{X}$  is a design matrix specifying the various non-reference group exposure levels and additional model covariates within study  $j$  and the reference group assumes a value of zero, therefore relieving the need for an intercept term;  $\boldsymbol{\beta}$  is a vector of unknown

regression coefficients; and  $\mathbf{e}$  is a vector of random errors such that  $\mathbf{e} \sim N(0, \mathbf{\Sigma})$

(Greenland & Longnecker, 1992; Orsini, Bellocco, & Greenland, 2006). Estimation of  $\beta$  is accomplished using

$$\hat{\beta} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{Y}, \quad (26)$$

and covariance estimator

$$\mathbf{V} = \widehat{\text{Cov}}(\beta) = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}, \quad (27)$$

where  $\mathbf{\Sigma}$  is the variance-covariance matrix of the vector of observed dose effects  $\mathbf{Y}$ . The off-diagonal elements of  $\mathbf{\Sigma}$  are presumed to be non-zero (i.e. the dose effect estimates covary). Steps for estimating the off-diagonal elements of  $\mathbf{\Sigma}$  are covered extensively by Orsini, Bellocco, and Greenland (2006).

Combining dose-response models in a meta-analysis is problematic if the estimated regression coefficients for each of the dosage levels are not independent; an assumption that is violated when a common (shared) reference group is used across dosage levels. In epidemiological research, this scenario is not uncommon. Greenland & Longnecker (1992), however, present a set of methods to combine dose-response models in meta-analysis using GLS estimation. Using (24) both fixed and random effects models can be used to combine dose-response estimates from primary research. Equation (24) can be used as a fixed effects dose-response meta-regression model where the summarized dose effect estimates are weighted by  $\mathbf{\Sigma}$  (Orsini, Bellocco, & Greenland, 2006). A random effects model may take the form of

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\eta + \mathbf{e}, \quad (28)$$



where  $\mathbf{X}$  is a matrix of model covariates;  $\mathbf{Z}$  is a vector of dosage or exposure levels; and  $\eta$  is the model random effect estimate such that  $\eta \sim N(0, \tau^2)$  where  $\tau^2$  is the between-study variance component with larger values indicating greater heterogeneity in summarized dose effect estimates.

There are many examples of the use of the above methods to combine dose-response studies in meta-analytic research. Longnecker (1994) used meta-analysis to combine relative risk estimates for alcoholic beverage consumption and breast cancer. Similarly, Corrao, Bagnardi, Zambon, and Arico (1999) meta-analyzed summarized dose-response studies of the relationship between alcoholic beverage consumption and six different cancer types. Greenland (1993) meta-analyzed the relationship between coffee consumption and myocardial infarction.

While the methods outlined by Greenland and Longnecker (1992), Berlin, Longnecker, and Greenland (1993), and Orsini, Bellocco, and Greenland (2006) account for the dependencies observed within-study summarized dose effect estimates, meta-analysis of full dose-response models (dose effects and specific model covariates) would require an extension of GLS (e.g. Becker & Wu, 2007). However, if the purpose of the meta-analysis is to synthesis dose effect estimates from effectively uniform dose-response models, extracting and correcting summarized focal dose effect estimates would be an appropriate method to use, just as in the case of diversely specified OLS models described above. For example, in the example of coffee consumption, if each study that examined the effects of the same levels of consumption on risk of myocardial infarction,

the log-based average of each study's dose effect estimates could be combined. But, to the extent that additional, and different, predictors are added to the dose-response models, the log-based average dose effects will be less comparable across studies.

*Meta-analysis of regression coefficients in psychological research.* The following section discusses methods developed by psychological researchers to combine regression estimates with meta-analysis. The first part of this section discusses validity generalization approaches and the second part of this section discusses imputing correlation coefficients based on regression slope estimates.

Validity generalization. Psychological researchers have a long history of applying meta-analytic techniques to regression models. One popular application has been in the realm of psychometric meta-analysis with a specific focus on validity generalization (e.g. Pearlman, Schmidt, & Hunter, 1980; Schmidt & Hunter, 1977). This line of inquiry sprang out of the industrial-organizational psychology literature on the generalizability of test validities (e.g. test-criterion correlations) which generally concluded that predictive job performance measures were only locally valid (Hunter & Schmidt, 2004). That is, the correlation between aptitude measures (test) and job performance (criterion) varied substantially across studies of the same job, performance measure, and criterion.

At the same time Glass was developing methods to combine empirical evidence that spoke to the effectiveness of psychotherapy research, John Hunter and Frank Schmidt were developing methods for synthesizing test validity coefficients. Their efforts culminated with the emergence of a new domain of meta-analytic research that focuses

on correcting for statistical measurement artifacts in primary research studies which, in the case of aptitude measure validity coefficients, explained away the previously observed between study variation.

As presented by Schmidt & Hunter (1977) and Raju, Pappas, and Williams (1989), the conventional approach to validity generalization meta-analysis synthesizes correlations taking the following structure

$$r = \rho \sqrt{r_{xx}} \sqrt{r_{yy}} \frac{u}{[1 + (1 - u^2)\rho^2 r_{xx} r_{yy}]^{\frac{1}{2}}} + e, \quad (29)$$

Where  $r$  is the observed correlation (validity coefficient) between test  $x$  and criterion  $y$ ;  $\rho$  the unrestricted and unattenuated population correlation;  $r_{xx}$  and  $r_{yy}$  are the population reliability parameters for test  $x$  and criterion  $y$ , respectively;  $u$  is the ratio of the restricted standard deviation of text  $x$  to the unrestricted standard deviation of test  $x$ ; and  $e$  is sampling error. Raju, Pappas, and Williams (1989) note that using these procedures leads to an estimation of total variance partitioned into the test variance, criterion variance, range restrictions, and sampling error. Applied under a meta-analytic framework, (21) leads to the estimation of the mean and variance of the population validity coefficient  $\rho$ :

$M_r$  and  $V_r$ . The weighted estimated mean of  $\rho$  is

$$M_r = \frac{\sum_{i=1}^k N_i r_i}{m}, \quad (30)$$

where  $N$  is the sample size of study  $i$ , and  $r$  is the observed validity coefficient from study  $i$ ; and  $m$  is the number of studies examining the relationship between test  $x$  and criterion

y. Each observed  $r$  from (27) are assumed to be independently estimated. The estimated variance of  $\rho$  is

$$V_r = \frac{\sum_{i=1}^k \frac{n_i(1 - r_i^2)^2}{n_i - 1}}{m}. \quad (31)$$

Using the estimates from (27) and (28) Raju and Burke (1983) provide the procedures for a Taylor series approximation of the “true” population mean and variance.

While test validity coefficients most typically take the form of a Pearson correlation, they can also be represented in a linear regression model as shown by Raju, Fralicx, and Steinhaus, (1986). Specifically the regression model for a validity generalization study takes the basic form of

$$b_{xy} = \beta_{yx}\rho_{xx} + e, \quad (32)$$

where one is interested in the observed coefficient  $b_{xy}$  of criterion  $y$  regressed on test  $x$ ;  $\beta_{xy}$  is the unattenuated and unrestricted population regression coefficient and  $e$  is sampling error. The authors also show that estimating the population validity coefficient from a model-based estimation procedure takes the form of

$$M_b = \frac{\sum_{i=1}^k n_i b_i}{m}, \quad (33)$$

and variance

$$V_b = \frac{\sum_{i=1}^k n_i b_i^2}{m} - M_b^2. \quad (34)$$

These estimates lead to the following corrected mean and variance parameter estimates

$$\hat{M}_B = \frac{M_b}{M_{r_{xx}}}, \quad (35)$$

and

$$\hat{V}_B = \frac{V_b - V_e - \left( \frac{M_b}{M_{r_{xx}}} \right)^2 V_{r_{xx}}}{V_{r_{xx}} + M_{r_{xx}}}. \quad (36)$$

The methods discussed in this section are most tenable when the regression estimates from (29) are derived from simple (single predictor) models and the test and criterion measures are harmonious across each of the studies in the validity generalization meta-analysis. Of the few studies using the validity generalization approach to combining regression slopes, none have satisfied these basic assumptions.

In one example, Crouch (1995; 1996) investigated 80 studies of price elasticities of international tourism demand. Unsurprisingly, there was a great diversity among these studies with over ten thousand relevant price elasticity regression coefficients reported in total. The author used the same validity generalization procedures outlined in this section to combine these estimates and then modeled their variation based on a number of study-level characteristics including model specification (e.g. Stanley & Jarrell, 1989). With such a large number of regression estimates in a relatively small number of studies, there were undoubtedly dependencies among those estimates. The author argues that the dependence issue will only increase the standard errors in the mean estimate and such an approach leads to a “conservative” (undercorrection for sampling error) combined estimate.

Szymanski, Troy, and Bharadwaj (1995) used the validity generalization methods outlined in this section to combine a total of 64 unstandardized regression coefficients of the relationship between order of entry and market share across 23 different studies. Like Crouch (1995; 1996), there was substantial variation in model specification among the 64 estimates used in the meta-analysis which ultimately explained a non-trivial percentage of the variance around the mean estimate.

While validity generalization meta-analyses of regression slopes have historically been rare, Aguinis, Culpepper, and Pierce (2010) called for the “revival of test bias research in pre-employment testing.” In this work, the authors conducted a large simulation study of the effects of statistical and methodological artifacts on regression slope-based differences in validity coefficients among minority and non-minority groups. Evidence was presented that challenges the established understanding of unbiased pre-employment screening.

Imputation of correlation coefficients given beta. One of the most recent methodological proposals for making use of regression coefficients in meta-analytic research has come from psychology. Peterson and Brown (2005) proposed the use of regression imputation using observed standardized regression coefficients to impute the bivariate correlations between  $x$  and  $y$ . The goal of this process is to estimate what the bivariate correlation between a focal predictor and an outcome would be if the multiple regression coefficient of that relationship was known. The authors show, that with two predictors the standardized regression coefficient is

$$\hat{\beta}_{x_1y} = \frac{r_{x_1y} - r_{x_1x_2}r_{x_1y}}{1 - r_{x_1x_2}^2}, \quad (37)$$

here  $x_1$  and  $x_2$  are predictors one and two; and  $y$  is the model outcome. They also show that the correlation between  $y$  and some target slope from (34) can be estimated using

$$r_{x_1y} = \hat{\beta}_{x_1y} + r_{x_1x_2}(r_{x_2y}) - \hat{\beta}_{x_1y}r_{x_1x_2}. \quad (38)$$

The authors use a large dataset of 1,700 standardized slope estimates and correlations from published studies. Inclusion criteria for the authors' investigation included: regression models were linear in their parameters; both standardized slope and correlation coefficients were reported; and the slopes were estimated with single models (i.e. multi-stage estimation equation were excluded). The slopes and corresponding regression coefficients were coded for sign, sample size, number of model covariates, and the presence of interaction terms. Using this information the authors constructed various models predicting  $r$  given  $\hat{\beta}$ . The authors derived the following best-fitting equation for their data:

$$\hat{r}_i = .98\beta_i + .05\lambda, \quad (39)$$

where  $\hat{r}$  is the predicted correlation coefficient;  $\beta$  is a vector of observed regression coefficients and  $\lambda$  is an indicator variable that takes the value of one if  $\beta$  is negative, and zero otherwise. To explore the viability of this model and some of its derivatives, the authors constructed the following “convenience” model which is a reduced form of (36) such that

$$\hat{r}_i = \beta_i + .05\lambda. \quad (40)$$

The authors also constructed a three parameter model that additionally takes into account the collinearity of primary study covariates which took the form of

$$\hat{r}_i = .99\beta_i + .04\lambda + .02\eta . \quad (41)$$

where  $\eta$  is an indicator variable taking the value of one when the mean intercorrelation of primary study model covariates is .18 or more, zero otherwise.

Observed correlation coefficients from their data set were randomly deleted and the authors compared the effectiveness of each of these three models as well as mean imputation to predict those missing values. They found that the reduced two-parameter or “convenience” model (37) performed just as well as the fuller three parameter model that included a collinearity parameter (38) as well as the model including regression weights derived from all observations (36).

The methods and results outlined in Peterson and Brown (2005) are interesting however their approach was studied in an isolated scenario. While there are advantages to using empirical data (as opposed to simulation) the later allows for richer investigation of a multitude of potential scenarios. For example, at what correlation between  $r$  and  $\beta$  does the utility of the regression imputation equations the authors provide diminish? Further, and most importantly, these methods yield approximations of the observed correlation coefficients and one would have to carefully consider the tradeoff between imprecision and practicality (see Kock & Gemünden, 2009).

Lux, Crook, and Woehr (2011) applied the methods outlined in Peterson and Brown (2005) to a meta-analysis of predictors and outcomes of corporate political



activity. The authors explore political, market, and corporate-level antecedents of corporate political contributions as well as the relationship between political contributions and later corporate-level economic performance. While correlations are a natural effect size for this type of a meta-analysis, many of the relationship the authors were interested in synthesizing were in the form of multiple regression coefficients. They used Peterson and Brown's (2005) “convenience” formula to impute the missing correlations given the observed values of  $r$ , holding constant the sign of the observed regression coefficient. Because the Peterson and Brown (2005) equation was created based on a very specific dataset, it is unclear how widely applicable it would be to other contexts.

*Meta-analysis of regression coefficients in ecological research.* Meta-analysis has been growing in popularity within the field of ecology. Stewart (2010), however, points out that less than 300 meta-analyses or systematic reviews have been undertaken in the ecological sciences over the past 16 years; far less than the medical and social science disciplines. Despite the small number of completed reviews, there are several examples of innovative meta-analytic designs in ecology.

One study, by Bini, Coelho, and Diniz-Filho (2001) used a weighted least squares approach to combine 74 independent regression slope estimates of the effect of population density on body weight for mammals and birds. Specifically, the authors used

$$b_c = \frac{\sum_{i=1}^k w_i b_i}{\sum_{i=1}^k w_i}, \quad (42)$$

where  $w$  is the inverse of the variance of the  $i$ th slope estimate. The variance of (39) is estimated using

$$\hat{\sigma}_{b_c}^2 = \frac{1}{\sum w_i}, \quad (43)$$

This study does not discuss the distribution of model specification. It may be that each of the 74 effect size estimates were drawn from a common model but this is impossible to know based on the published information. In such case, the sampling distribution of the estimates from (39) will be conditional upon each additional model covariate.

In another example, Paul, Lipps, & Madden (2006) used individual unit data (i.e. individual participant data) meta-analytic techniques to synthesize 126 independent estimates of the relationship between fusarium head blight and deoxynivalenol content of wheat. These authors were given access to the raw primary data of 126 different studies that included the relevant data for their research question and were able to apply the same simple regression model to each set of data. A mixed model was then used to synthesize the results of each of these regression analyses.

Using individual unit or participant data is perhaps one of the most promising approaches to combining multiple regression estimates in meta-analysis. Access to primary study data provides ultimate flexibility in estimating various relationships. There are, however, inferential issues such as Simpson's paradox and ecological bias that must be considered (Cooper & Patall, 2009). Unfortunately, many research fields, including education, have not yet adopted a data-sharing culture of research and practice.

Methods developed in the industrial-organizational psychology literature have also been applied in ecological meta-analytic work. Root et al. (2003) extracted regression slopes representing the relationship between temperature shift and biological changes in various species. The authors cite Raju, Fralicx, and Steinhaus (1986) in their methods but importantly note that the within-study sampling variance necessary to estimate the mean regression coefficient variance was not used. The published report was brief and few details were provided about the specific methods used to combine the regression estimates, the distribution of model specification, or model form. Additionally, the authors used a significance test vote-counting procedure to supplement the results of their meta-analysis of regression slopes.

*Meta-analysis of regression coefficients in education research.* The following section discusses some approaches developed by education researchers for combining to results of regression studies with meta-analysis. The first part of this section discusses the methods used by Greenwald, Hedges, and Laine (1996) in their education production function synthesis. Then GLS is discussed in the context of education research, and lastly a method for combining semi-partial correlations is discussed.

From the education production functions debate. Educational researchers have worked to develop methods for combining regression-based estimates in meta-analysis for the past two decades. This section will review a series of proposed methods for using regression estimates in meta-analytic research including vote-counting procedures, use of half-standardized regression coefficients, partial and semi-partial correlations ( $f^2$  effect

size estimates), combined significance tests, multivariate GLS, and the use of effect size estimates from regression models.

In the late 1980's and early 1990's a vibrant debate on the effects of school resources on student outcomes took center stage. Erik Hanushek (1989) published a vote-counting meta-analysis that dismissed the importance of expenditures on student achievement. Given that much of the literature Hanushek was analyzing was econometric in nature, most of the relationships between school inputs and student achievement outcomes were summarized in the form of regression equations. Hanushek extracted the regression slope estimates of interest and coded them for statistical significance as well as for direction (i.e. positive or negative). In all, there were 38 different studies which totaled 187 different p-values.

Hedges, Laine, and Greenwald (1994) identified several critical flaws in Hanushek's (1989) study. First, and foremost, vote-counting does not, and cannot, illustrate the magnitude of a relationship. Secondly, the statistical power of any one study (i.e. the probability of correctly retaining the null hypothesis) is a function of the statistical tests used, a predetermined alpha level and of course the size of the relationship. Power varies (often wildly) across the studies that meet the inclusion criteria for a synthesis, and, as Hedges, Laine, and Greenwald point out, it is arguably misleading not to acknowledge that each study does not have an equal opportunity to reject the null hypothesis.

The methods and results presented by Hanushek (1989) prompted Hedges and his colleagues to reanalyze the same set of data taking two separate meta-analytic approaches: combined significance meta-analysis and the use of half-standardized regression coefficients to represent effect magnitude. Hedges, Laine, and Greenwald (1994) first used methods to statistically combined the p-values from the 187 observed regression slope estimates to a) test the hypothesis that there was no positive relationship between per-pupil expenditure (PPE) and academic achievement; and b) to test a second hypothesis that there is no negative relationship between PPE and academic achievement. This strategy differs markedly from the vote-counting procedure used by Hanushek (1994), however it too does not provide an average effect size estimate. The regression model inputs that Hedges et al. (1994) were interested in were all on a uniform scale (i.e. dollars), however the achievement outcomes were not. For this reason, the authors half-standardized the slope estimates used in the meta-analysis by dividing each by the standard deviation of the model outcome such that

$$b_H = \frac{b}{\hat{\sigma}_y}. \quad (44)$$

A one dollar increase would translate into number of standard deviations change in outcome with this approach.

Problems still exist with the modified education production function meta-analysis conducted by Hedges and colleagues. While the approach largely addresses the issue of inconsistent scales of measurement across studies, it does not solve the problem of heterogeneous regression models used across each of the independent primary studies.

Generalized least squares. Becker and Wu (2007) provide one of the most viable methods for combining multiple regression studies in meta-analysis. The authors outline many of the drawbacks with currently methods for combining regression coefficients in a meta-analysis. They argue that while many investigators are interested in extracting and combining specific regression coefficients, others may be interested in combining full multiple regression models.

One important statistical advantage of meta-analyzing full regression models is that it obviates the issue of heterogeneous primary study models. For example, educational researchers may be interested in meta-analyzing studies that use the following model

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \varepsilon_{ij} , \quad (45)$$

where  $Y$  is the  $i$ th student's Scholastic Aptitude Test (SAT) score in sample  $j$ ,  $X_1$  is per-pupil expenditure,  $X_2$  is teacher education, and  $X_3$  is class size. The effect size in this example effectively becomes the linear combination of each of the three predictor variables in contrast to a single regression coefficient which ignores the presence and influence of additional model covariates.

A problem that arises with this approach is non-zero covariance among slope estimates (Becker & Wu, 2007). This point is perhaps best described as a clustering problem. Slope estimates within a given study are dependent, to the degree that they covary with other slope estimates. This problem leads to the underestimation of standard

errors. To alleviate this problem, Becker and Wu (2007) propose the use of a multivariate GLS approach.

The GLS approach that Becker and Wu promote starts with a pool of studies that use the same predictor and outcome variables; a uniform model across studies. Each slope and intercept estimate from each study model, assumed to be independent, are stacked such that

$$\mathbf{b}_i = (b_{i0}, b_{i1}, \dots, b_{iP}) = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{Y}_i, \quad (46)$$

where  $b_{i0}$  is the  $i$ th study's intercept and  $b_{i1}$  to  $b_{iP}$  are the slope estimates from study  $i$ .

The variance-covariance structure of  $\mathbf{b}_i$  is estimated using

$$\mathbf{\Sigma}_i = \text{Cov}(\mathbf{b}_i) = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \sigma_i^2, \quad (47)$$

where  $\sigma_i^2$  is estimated using the  $i$ th study's mean square error (MSE), as noted by Becker and Wu (2007),  $\mathbf{b}_i \sim N(\beta_i, \mathbf{\Sigma}_i)$ . Each of  $k$  vectors of slope and intercept estimates are stacked into single vector such that

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad (48)$$

and each of  $k$  covariance estimates are block-diagonally stacked where

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Cov}(\mathbf{b}_1) & 0 & 0 & 0 \\ 0 & \text{Cov}(\mathbf{b}_2) & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \text{Cov}(\mathbf{b}_k) \end{bmatrix}. \quad (49)$$

The full GLS model then becomes

$$\mathbf{b} = \begin{bmatrix} b_{10} \\ b_{11} \\ \vdots \\ b_{1P} \\ \vdots \\ b_{k0} \\ \vdots \\ b_{kP} \end{bmatrix} = \mathbf{W}\boldsymbol{\beta} + \mathbf{e} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{bmatrix} + \mathbf{e}, \quad (50)$$

where  $\mathbf{W}$  is a design matrix indicating the presence of model covariates. In the example discussed by Becker and Wu,  $\mathbf{W}$  is a stack of  $k$  identity matrices, given uniform model specification across  $k$  samples. The authors note that GLS estimates can be produced using (44) by modifying  $\mathbf{W}$ . Slope estimates from models of varying specification may not, however, share a common distribution.

From (44)  $\boldsymbol{\beta}$  and its covariance is estimated using

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}\mathbf{W}\boldsymbol{\Sigma}^{-1}\mathbf{b}, \quad (51)$$

and

$$\text{Cov}(\hat{\boldsymbol{\beta}}^*) = (\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W})^{-1}. \quad (52)$$

where  $\boldsymbol{\Sigma}$  is unknown and estimated using a matrix of model MSEs, which Becker and Wu call  $\mathbf{V}$ , and which updates (45) and (46) to the estimators

$$\hat{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}\mathbf{b}, \quad (53)$$

and

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}. \quad (54)$$

Asymptotically,  $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \text{Cov}(\hat{\boldsymbol{\beta}}))$ . The authors show that when all studies in a meta-analysis specify the same multiple regression model, and each provides an estimate of



$\sigma_i^2$ , one may use a blockwise diagonal matrix  $(\mathbf{X}^*)$  containing  $(\mathbf{X}_i' \mathbf{X}_i)^{-1}$  instead of  $\mathbf{V}$  in (55) and (56).

Becker and Wu (2007) note that a fundamental shortcoming with the GLS is that it requires primary studies to report  $\text{Cov}(\mathbf{b}_i)$ . Primary study researchers generally have little interest in knowing the covariance structure of their regression model coefficients and therefore rarely estimate it and even more rarely report it. In the absence of knowing the covariance structure of each multiple regression model, the authors note that estimates can be assumed to be independent or a common correlation among slopes can be assumed (e.g. .20). The first of these options is more tenable, as the authors mention, when there is no multicollinearity and the model has been properly specified. Unfortunately, diagnostic procedures such as multicollinearity analyses are generally omitted from final reports.

Combining  $R^2$  estimates. It may be useful in some scenarios to extract and synthesize from regression analyses squared multiple correlation ( $R^2$ ). If a common model is of interest, investigators may use  $R^2$  as a measure of effect. The multiple correlation coefficient is simply the proportion of the variance in the dependent variable explained by the linear combination of independent variables. Specifically

$$R^2 = 1 - \frac{SS_{residual}}{SS_{regression}}, \quad (57)$$

where

$$SS_{regression} = \sum (\hat{Y}_i - \bar{Y})^2$$

and

$$SS_{residual} = \sum (Y_i - \hat{Y}_i)^2.$$

Combining  $R^2$  estimates is often just as difficult to combine in meta-analysis as regression slope estimates. When model specification varies across studies, a common effect size distribution cannot be assumed.

Even when a common model is used across all studies in a meta-analysis, authors may instead be interested in a specific bivariate relationship embedded in each of those models; a focal slope. The use of  $R^2$  in that scenario will not be useful. Whereas  $R^2$  provides an overall estimate of model fit, or the proportion of variance in the outcome explained by the predictors, one may be interested in knowing the change in  $R^2$  when specific predictors are added to a model, such a partial or semi-partial correlation.

Aloe (2009), however, proposed the use of the semi-partial correlation coefficient as an effect size estimator. His estimator,  $r_{sp}$  can be computed using

$$r_{sp} = \frac{t_p \sqrt{(1 - r_Y^2)}}{\sqrt{(n - p - 1)}}, \quad (58)$$

with variance

$$\begin{aligned} \text{Var}(r_{sp}) = 1/n(R^2 - R_j^2) & \left[ R^2(1 - R^2)^2 + R_j^2(1 - R_j^2)^2 - 2R \right. \\ & \left. \times R_j \left[ \frac{1}{2(2R_* - R \times R_j)(1 - R^2 - R_j^2 - R_*^2)} + R^3 \right] \right] \end{aligned} \quad (59)$$

where  $t_p$  is the test statistic of a  $p$ th model slope estimate,  $r_Y^2$  is model the  $R^2$  estimate,  $p$  is the number of predictors included in the model,  $R_j$  is difference between  $R^2$  and  $r_{sp}$  for the  $j$ th predictor variable, and  $n$  is the sample size.

Models with different predictor variables can be analyzed using this approach but each effect size estimate may be estimating a different parameter. Because the semi-partial correlation represents the unique proportion of variance in the outcome explained by a focal predictor,  $X_f$ , the magnitude of  $r_{sp}$  depends on the amount of collinearity among model predictors. And, while the  $r_{sp}$  estimator could be used in combination with bivariate correlations in a meta-analysis, the parameter estimates of each of these effect sizes is different. Aloe (2009) cautions against the use of both semi-partial correlations and bivariate correlations in the same meta-analysis until the statistical behavior of these estimators better understood.

## Summary

The previous sections presented a survey of some methods used to combine regression estimates with meta-analysis. Researchers across different fields are clearly eager to make use of multiple regression estimates in meta-analytic fashion. Some of the methods in use are intuitive, others are considerably more complex. Many of the methods, however, fail to address the fundamental problems of combining multiple regression estimates: collinearity among model predictors and diverse model specification. Other methods, such as GLS and individual participant data meta-analytic

techniques are promising but often the data needed to use those methods are unavailable. Flexible methods are still needed to meta-analyze regression estimates.

Most of the methods discussed thus far are univariate methods focusing on synthesizing single coefficients. Some of these approaches ignore problems that arise with heteroskedastic standard errors (e.g. Stanley and Jarrell, 1989) while others (e.g. Bini, Coelho, & Diniz-Filho, 2001) use WLS to alleviate this concern. Assuming that each slope estimate is independent combining focal slope estimates and modeling variation based on sample and model specification characteristics may be appropriate in some circumstances. Often, however, multiple models are used with the same sample. And, those models may each contain the focal slope estimate. Assuming that each effect size estimate is independent may not be tenable. Meta-analysts would generally be restricted to selecting one estimate per sample or averaging within sample.

Multiple regression models in the social sciences and in medicine commonly use variables that are, at least to some degree, correlated. Effect size dependencies in this case may be induced by collinearity in model covariates. As such, the multivariate GLS approach used by Becker and Wu (2007), accounts for the covariance among slope estimates and provides a pooled model with correct standard errors. In many ways the approach outlined by Becker and Wu (2007) is the most attractive approach for combining multiple regression estimates. However, one major shortcoming of this approach is that it requires information about the covariance structure of the model predictors; something that is almost never reported. In the absence of this information,

meta-analysts must assume a covariance structure for their data. But ultimately, these assumptions would be arbitrary and potentially overly conservative (assuming a high level of covariation among slope estimates) or liberal (assuming independent estimates).

Chapter three details the robust variance estimator proposed by Hedges, Tipton, and Johnson (2010). The estimator the authors propose may be used to combine weighted dependent focal slope estimates, using meta-regression to model variation in the focal estimates. And, the estimator may be used analogously to multivariate GLS, estimating a pooled model. However, as will be shown, knowledge of the covariance among predictors is not required for this approach.

## CHAPTER THREE

### METHOD

The methods presented above fail to cover several looming problems with meta-analyzing regression estimates. This chapter introduces robust variance estimation as an alternative method for combining multiple regression estimates.

#### **Robust variance estimation and meta-analysis**

Because the covariance structure of correlated effect size estimates is almost never report (or even explored) in primary research, Hedges, Tipton, and Johnson (2010) discuss in detail the use of robust standard errors in meta-regression with dependent effect size estimates. Their approach has been used with standardized mean differences and effect size estimates for binary outcomes. It has not yet been applied to multiple regression studies.

The procedures to follow assume that dependencies occur through correlated error terms. As such, a correlated effects modeling approach is presented. Modeling multiple regression estimates takes the following linear model form

$$\mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} , \quad (60)$$

where  $\mathbf{b}$  is a vector of  $n$  (from  $n = 1$  to  $k$ ) clustered partial effects in  $m$  studies (clusters),  $\mathbf{X}$  is a design matrix of  $(k_j \times p)$  meta-regression covariates (e.g. model specification indicators),  $\boldsymbol{\beta}$  is a vector of unknown regression coefficients, and  $\mathbf{e}$  is a vector of residuals. In the context of multiple regression estimates,  $\mathbf{b}$  is a vector of stacked model

slope (and intercept) estimates. In the case of combining commonly specified models  $\mathbf{X}$  may be an identity matrix and

$$\mathbf{b} = \begin{bmatrix} b_{01} \\ b_{11} \\ \vdots \\ b_{1p} \\ \vdots \\ b_{k0} \\ \vdots \\ b_{kp} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_p \end{bmatrix}.$$

The weighted least squares estimate of  $\boldsymbol{\beta}$  is

$$\hat{\boldsymbol{\beta}} = \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{b}_j \right), \quad (61)$$

with variance

$$V = \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left( \frac{1}{m} \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \Sigma_j \mathbf{W}_j \mathbf{X}_j \right) \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1}, \quad (62)$$

where  $\mathbf{W}$  is a weight matrix for the  $m$ th study such that  $\mathbf{W} = \text{diag}(\mathbf{W}_1 \dots \mathbf{W}_m)$  such that  $\mathbf{W}_j = \text{diag}(w_{j1}, \dots, w_{jkj})$ ,  $\Sigma$  is the  $m$ th study's  $\text{Cov}(\mathbf{b})$  such that  $\Sigma = \text{diag}(\Sigma_1, \dots, \Sigma_m)$ . Hedges et al. show that every element of  $\Sigma_j$  does not need to be estimated but rather the average of linear combinations of  $\Sigma_j$  need be estimated, specifically

$$\left( \frac{1}{m} \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \Sigma_j \mathbf{W}_j \mathbf{X}_j \right). \quad (63)$$

Because study report information about  $\Sigma_j$  is almost never provided, an empirical estimate of  $\Sigma_j$  is needed. Hedges, Tipton, and Johnson (2010) show that the cross product of within-study residuals may be used as a crude approximation of  $\Sigma_j$  so that the robust variance estimator is

$$\mathbf{V}^R = \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left( \frac{1}{m} \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{e}_j \mathbf{e}_j' \mathbf{W}_j \mathbf{X}_j \right) \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1}, \quad (64)$$

where  $\mathbf{e}_j \mathbf{e}_j'$  is the  $j$ th study's matrix of cross products of within-study residuals.

Specifically,

$$\mathbf{e}_j = \mathbf{b}_j - \mathbf{X}_j \boldsymbol{\beta}. \quad (65)$$

The asymptotic distribution of  $m$  as it approaches  $\infty$  is

$$\sqrt{m} (\bar{\mathbf{V}})^{-\frac{1}{2}} (\mathbf{b} - \boldsymbol{\beta}) \rightarrow N_p(\mathbf{0}, \mathbf{I}_p), \quad (66)$$

where  $\bar{\mathbf{V}} = m \mathbf{V}^R$ . As  $m$  rises,  $\mathbf{b} \sim N(\boldsymbol{\beta}, \mathbf{V}^R)$ . The estimated robust standard error of  $b_j$  is then

$$S_j^R = \sqrt{\frac{m v_{jj}^R}{m-p}}, \quad (67)$$

where  $v_{jj}^R$  is the  $j$ th diagonal element of  $\mathbf{V}^R$  and  $\sqrt{\frac{m}{m-p}}$  is a finite population correction. A

robust significance test of  $\beta_j$  follows a  $t$  distribution, specifically

$$t_j^R = \frac{b_j}{S_j^R}, \quad (68)$$

with  $m-p$  degrees of freedom. And, a robust confidence interval may be obtained using



$$b_j - c_{\alpha/2} S_j^R \leq \beta_j \leq b_j + c_{\alpha/2} S_j^R, \quad (69)$$

where  $c_\alpha$  is the  $1-\alpha$  point on the  $t$  distribution with  $m-p$  degrees of freedom.

Hedges, Tipton, and Johnson (2010) identify three important features of this estimator. First, and most importantly, the covariance structure of effect size estimates is not needed. Second, parameter estimates converge on the target parameter as the number of studies, not the number of cases within studies, rises. And, the authors show that accurate standard errors are produced with as few as 10 to 20 studies. Third, this estimator is unbiased for any set of weights.

The authors show the calculation of the mean effect size estimate is

$$b_1 = \frac{\sum_{j=1}^m w_{1j} T_{1j}}{\sum_{j=1}^m w_{1j}}, \quad (70)$$

and with identical weighting procedures used for within-study estimates, the robust variance estimate becomes

$$V^R = \frac{(\sum_{j=1}^m w_{1j} \bar{b}_j - b_1)^2}{(\sum_{j=1}^m w_j)^2}, \quad (71)$$

where  $\bar{b}$  is the  $j$ th cluster's unweighted mean effect size estimate,  $b_1$  is the weighted mean effect size estimate from (70), and  $w_j$  is the weight assigned to the  $j$ th cluster of effect size estimates; assumed to be uniform (i.e. common across within-study effect size estimates). When uniform weights are used  $V^R$  is equal to  $(m-1)/m^2$  times the value of the typical variance. And, the robust standard error estimate  $S^R$  is equal to  $1/m$  times the typical variance of  $\bar{b}_j$  when uniform weights are used.

Weighting within-study estimate and cluster (study) total weights warrants consideration. Because between-study variance is likely, in most meta-analyses, to be much greater than within-study variance, an optimal weighting procedure is used. As such assigning each within-study estimate an equal weight would mean

$$w_{ij} = \frac{1}{k_j \bar{v}_{j*}} = \frac{1}{\sum_{i=1}^k v_{ij}}, \quad (72)$$

where  $\bar{v}_{j*}$  is the average of the effect size estimate variances in study  $j$ . The authors note that little precision is lost with this approach so long as the covariance among within-study estimates is reasonably high. If the covariance structure of the estimates is known or estimable, it could be used in could be used with (72) to improve efficiency, however, the advantage of this approach is that it requires no information about the covariance structure to produce accurate standard errors.

It is also possible to estimate between-study variances with this approach. Using fixed effects weights described above, the weighted residual sum of squares homogeneity statistic is

$$Q_E = \sum_{j=1}^m \mathbf{b}_j' \mathbf{W}_j \mathbf{b}_j - \left( \sum_{j=1}^m \mathbf{b}_j' \mathbf{W}_j \mathbf{b}_j \right) \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{b}_j \right)^{-1} \left( \sum_{j=1}^m \mathbf{X}_j' \mathbf{W}_j \mathbf{b}_j \right), \quad (73)$$

and the estimated residual variance component is

$$\hat{\tau}^2 = \frac{Q_E - m + \text{tr} \left[ \mathbf{V} \left( \sum_{j=1}^m \frac{w_j}{k_j} \mathbf{X}_j' \mathbf{X}_j \right) \right] + \rho * \text{tr} \left[ \mathbf{V} \left( \sum_{j=1}^m \frac{w_j}{k_j} [\mathbf{X}_j' \mathbf{J}_j \mathbf{X}_j - \mathbf{X}_j' \mathbf{X}_j] \right) \right]}{\sum_{j=1}^m k_j w_j - \text{tr} [\mathbf{V} (\sum_{j=1}^m w_j^2 \mathbf{X}_j' \mathbf{J}_j \mathbf{X}_j)]}, \quad (74)$$

where  $\mathbf{J}_j$  is a  $k_j \times k_j$  matrix of 1's,  $\rho$  is the within study correlation of effects, and  $\mathbf{V}$  is the inverse of the  $\mathbf{X}'\mathbf{W}\mathbf{X}$  matrix specified as

$$\mathbf{V} = \left( \sum_{j=1}^m w_j \mathbf{X}_j' \mathbf{X}_j \right)^{-1}. \quad (75)$$

Knowing  $\rho$  may seem contradictory to this general approach, since we presume the covariance structure of dependent estimates is unknown. However, for this purpose, estimating a between study variance component,  $\rho$  may be roughly approximated without severe penalty so long as  $m$  is reasonably large. Ishak, Platt, Joseph, & Hanley (2008) simulated meta-regression analyses under various scenarios that demonstrate this point, and Hedges, Tipton, and Johnson (2010) reinforce this point in their work.

## Analyses

What follows are two phased sets of analyses. Phase I is an applied analytical example of using robust standard errors to combine regression results with meta-regression. Phase II is a series of Monte Carlo simulation studies examining the performance of the robust variance estimator in the context of meta-analysis (specifically, meta-regression) of multiple regression estimates.

**Applications.** Phase I of this work used a subset of a large database containing educational production function models. This dataset was compiled over three years and contains the same 60 studies included in Greenwald, Hedges, and Laine (1996). Over

6000 slope estimates were coded for these 60 studies. A total of 566 regression models of various types and form were coded. This dataset was tremendously diverse and as a result, two small subsets were used for illustration purposes in Phase I of this research.

Phase I was comprised of two tasks. The first task is a meta-analysis of a subset of commonly specified models. Link and Mulligan (1986) provide a nice example for this kind of meta-analysis. The authors collected data on a national random sample of students in grades three through six for the 1976-1977 school year. Analyses were disaggregated by grade level and ethnicity (African American, Latino, and White students), providing a total of 12 independent samples. The authors specified the following regression model for each sample:

$$\begin{aligned} \text{Math}_{ij} = & \beta_0 + \beta_1(\text{BooksInHome})_{ij} + \beta_2(\text{CompsEd})_{ij} + \beta_3(\text{MathPre})_{ij} \\ & + \beta_4(\text{LowClassAch})_{ij} + \beta_5(\text{TeachExp})_{ij} + \beta_6(\text{HoursMath})_{ij} \\ & + \beta_7(\text{HoursMath}^2)_{ij} + \beta_8(\text{Male})_{ij} + \beta_9(\text{MotherEd})_{ij} + e_{ij} \end{aligned}$$

where *Math* is the *i*th student's mathematics score on the Comprehensive Test of Basic Skills in sample *j*; *BooksInHome* is the estimated number of books the *i*th student has in their home; *CompsEd* indicates whether or not a child was in need of compensatory education according to their teacher; *MathPre* is the *i*th student's Math pretest; *LowClassAch* indicates if a student was in a large class of mostly low-achieving students; *TeachExp* indicates if the *i*th student's teacher had worked for at least six years; *HoursMath* is the number of hours the *i*th student spend working on mathematics each week; *Male* indicates the *i*th student's gender; and *Mother Ed* indicates if the *i*th student's mother graduated from college.

The formal meta-regression model used to estimate a mean regression model was

$$\mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} ,$$

where  $\mathbf{b}$  is a vector of fixed-effects weighted slope estimates from (68),  $\mathbf{X}$  is an identity matrix indexing each of the model covariate types,  $\boldsymbol{\beta}$  is a vector of unknown regression coefficients, and  $\mathbf{e}$  is a vector of residuals, which in this case, is a vector of deviations between each observed slope and its meta-regression slope parameter estimate. A robust standard error from (60) and robust confidence interval from (65) were computed for each meta-regression coefficient.

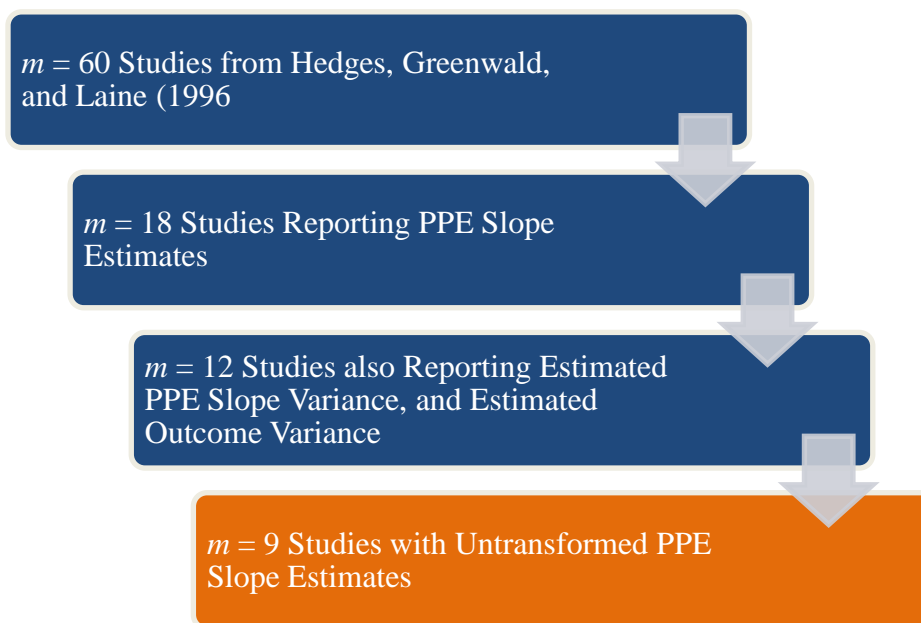
The second task for Phase I is a meta-analysis of nine studies regressing academic achievement on per-pupil expenditure (PPE), a focal slope meta-analysis. Measures of academic performance varied from study to study. Some studies used measures of basic skills, others used letter grades, and others used standardized assessments. Because of this diversity in outcome measurement, we half-standardized each of the PPE slope estimates by dividing them by the standard deviation of the outcome just as Greenwald, Hedges, and Laine (1996) did. The slope estimates then represent the increase in standardized units of student achievement for each dollar increase in PPE. While 19 studies included PPE in at least one study model, only six studies contained the necessary information to conduct a meta-analysis. The information required for this synthesis was:

- The outcome was student academic performance
- The outcome standard deviation was provided
- The outcome was regressed on PPE

- The estimated slope variance (e.g. the estimated standard error) was reported

While only a small subset of studies met the above criteria, a total of 20 effect size estimates were included. The following figure presents the exclusionary steps that were used for these data.

Figure 5. Exclusionary Steps in Focal Slope Meta-Analysis Example



Because this dataset was tremendously diverse, model predictors were categorized to ease analysis. Four broad categories were created: covariates related to students (e.g. prior achievement, intelligence, grade level, etc.); covariates related to teachers (e.g. teacher experience, teacher education, etc.); covariates related to schools (e.g. school-level poverty, school-level achievement, etc.); and covariates related to families (e.g. parental education, parental income, number of siblings, etc.). Admittedly, this is a crude method for combining slope estimates; however, it still presents the practical illustration of robust variance estimation in this type of meta-analysis.

A modified version of the linear model used for the model-based approach in Task A was used for Task B:

$$\mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} ,$$

where  $\mathbf{b}$  is a vector of inverse mean variance weighted slope estimates from (68),  $\mathbf{X}$  is a matrix of  $p$  model specification indicators,  $\boldsymbol{\beta}$  is a vector of unknown regression coefficients, and  $\mathbf{e}$  is a vector of residuals.

While Task A was substantively concerned with each of the meta-regression model coefficients, the focal slope approach is concerned with the estimated intercept (the unconditional focal slope estimate) and its standard error. A robust standard error from (60) and robust confidence interval from (65) was computed for each meta-regression coefficient, including the intercept.

To estimate the between studies variance component,  $\tau^2$ , in Tasks A and B, an initial estimate of correlation among slope estimates is needed. For Tasks A and B, .80 was. Without a good understanding of what a plausible value of the correlation among slopes would be, Hedges, Tipton, and Johnson (2010) recommend a sensitivity analysis using a range of values to see how they affect the between-studies variance component,  $\tau^2$ . A sensitivity analysis was used both Tasks.

**Simulation studies.** Phase II focused on evaluating the performance of the robust variance estimator proposed by Hedges, Tipton, and Johnson (2010) in a series of simulation studies. Phase II was comprised of two tasks. The first task evaluated the robust variance estimator's performance in combining commonly specified model (i.e.

analogous to Phase I, Task A). The second task evaluated the robust variance estimator's performance combining multiply dependent focal slope estimates (i.e. multiple models using the same focal covariate from the same sample).

*Parameter values.* Important meta-analytic parameters were varied to explore the robust variance estimator's performance under different conditions. Specifically, the number of units per study,  $n$ , the number of studies in a meta-analysis,  $m$ , the number of slope estimates per study,  $k$ , the correlation among primary units and effect size estimates,  $\rho$ , and the between study variance component,  $\tau^2$  (and subsequently  $I^2$ ) will be manipulated in these simulations.

A primary focus of these simulations is to explore the performance of the robust variance estimator under less desirable conditions (e.g. small numbers of studies, small sample sizes, and high correlations among estimates) and under relatively common conditions in the social sciences and medicine. As such, the following parameter values were selected for these simulations:

$$\begin{bmatrix} n & m & k & \rho & I^2 \\ 30 & 12 & 3 & .20 & .00 \\ 60 & 20 & 6 & .50 & .33 \\ 100 & 40 & 9 & .90 & .50 \end{bmatrix}$$

Each unique combination of parameter values represents one condition, for a total for 243 conditions for each meta-analytic approach.

These parameter values were selected for a couple of different reasons. Many of these values are similar, but not identical, to other, recent, simulation work on using robust standard errors in meta-analysis of dependent effect size estimates (e.g. Hedges,



Tipton, and Johnson, 2010; Tipton, 2011). For example, Hedges, Tipton, and Johnson (2010) used values of 10, 20, and 40 for  $m$ . Values of 12, 20, and 40 are used in this work primarily to facilitate meta-regression with a small number of studies and a large number of potential covariates with the focal slope approach. And, because the parameter values used in this study do not deviate dramatically from related simulation studies, the usefulness of this estimator is placed into a more general context.

***Data generation.*** The following section outlines the procedures used to generate primary study data for the model-based approach and the focal slope approach.

*Model-based approach.* This approach involved generating  $n$  primary study units from a  $k$ -variate normal distribution, each having a mean of 0 and variance of 1, for each of  $m$  studies included in each meta-analysis. A focus of these simulations was on the covariance structure among the slope estimates: the source of effect size dependency with this approach.

The covariance structure of each of the  $k$  model covariates was structured so that the average correlation between any two slope estimates was equal to  $\rho$ , which take the value of .20, .50, and .90 in these simulations. Nine slope correlation matrices were generated for these simulations: three conditions for  $k$  by three conditions for  $\rho$ . Each of these matrices yielded an average correlation among slope estimates equal to the specified value of  $\rho$ , which took the value of .20, .50, and .90 in these simulations. The values of these population slope correlation matrices were:

$k = 9, \rho = .90$ 

$$\mathbf{R}_1 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 \\ 1.0000 & 0.9227 & 0.8917 & 0.8723 & 0.8567 & 0.8414 & 0.8231 & 0.7954 & 0.7339 \\ 0.9227 & 1.0000 & 0.9664 & 0.9454 & 0.9285 & 0.9119 & 0.8921 & 0.8621 & 0.7954 \\ 0.8917 & 0.9664 & 1.0000 & 0.9783 & 0.9608 & 0.9436 & 0.9231 & 0.8921 & 0.8231 \\ 0.8723 & 0.9454 & 0.9783 & 1.0000 & 0.9821 & 0.9645 & 0.9436 & 0.9119 & 0.8414 \\ 0.8567 & 0.9285 & 0.9608 & 0.9821 & 1.0000 & 0.9821 & 0.9608 & 0.9285 & 0.8567 \\ 0.8414 & 0.9119 & 0.9436 & 0.9645 & 0.9821 & 1.0000 & 0.9783 & 0.9454 & 0.8723 \\ 0.8231 & 0.8921 & 0.9231 & 0.9436 & 0.9608 & 0.9783 & 1.0000 & 0.9664 & 0.8917 \\ 0.7954 & 0.8621 & 0.8921 & 0.9119 & 0.9285 & 0.9454 & 0.9664 & 1.0000 & 0.9227 \\ 0.7339 & 0.7954 & 0.8231 & 0.8414 & 0.8567 & 0.8723 & 0.8917 & 0.9227 & 1.0000 \end{bmatrix}$$

 $k = 9, \rho = .50$ 

$$\mathbf{R}_2 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 \\ 1.0000 & 0.6754 & 0.5226 & 0.4244 & 0.3510 & 0.2903 & 0.2357 & 0.1824 & 0.1232 \\ 0.6754 & 1.0000 & 0.7737 & 0.6283 & 0.5197 & 0.4298 & 0.3490 & 0.2700 & 0.1824 \\ 0.5226 & 0.7737 & 1.0000 & 0.8121 & 0.6717 & 0.5555 & 0.4511 & 0.3490 & 0.2357 \\ 0.4244 & 0.6283 & 0.8121 & 1.0000 & 0.8270 & 0.6839 & 0.5555 & 0.4298 & 0.2903 \\ 0.3510 & 0.5197 & 0.6717 & 0.8270 & 1.0000 & 0.8270 & 0.6717 & 0.5197 & 0.3510 \\ 0.2903 & 0.4298 & 0.5555 & 0.6839 & 0.8270 & 1.0000 & 0.8121 & 0.6283 & 0.4244 \\ 0.2357 & 0.3490 & 0.4511 & 0.5555 & 0.6717 & 0.8121 & 1.0000 & 0.7737 & 0.5226 \\ 0.1824 & 0.2700 & 0.3490 & 0.4298 & 0.5197 & 0.6283 & 0.7737 & 1.0000 & 0.6754 \\ 0.1232 & 0.1824 & 0.2357 & 0.2903 & 0.3510 & 0.4244 & 0.5226 & 0.6754 & 1.0000 \end{bmatrix}$$

 $k = 9, \rho = .20$ 

$$\mathbf{R}_3 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 \\ 1.0000 & 0.4640 & 0.2361 & 0.1227 & 0.0641 & 0.0335 & 0.0174 & 0.0089 & 0.0041 \\ 0.4640 & 1.0000 & 0.5089 & 0.2645 & 0.1382 & 0.0722 & 0.0375 & 0.0191 & 0.0089 \\ 0.2361 & 0.5089 & 1.0000 & 0.5197 & 0.2715 & 0.1418 & 0.0737 & 0.0375 & 0.0174 \\ 0.1227 & 0.2645 & 0.5197 & 1.0000 & 0.5224 & 0.2729 & 0.1418 & 0.0722 & 0.0335 \\ 0.0641 & 0.1382 & 0.2715 & 0.5224 & 1.0000 & 0.5224 & 0.2715 & 0.1382 & 0.0641 \\ 0.0335 & 0.0722 & 0.1418 & 0.2729 & 0.5224 & 1.0000 & 0.5197 & 0.2645 & 0.1227 \\ 0.0174 & 0.0375 & 0.0737 & 0.1418 & 0.2715 & 0.5197 & 1.0000 & 0.5089 & 0.2361 \\ 0.0089 & 0.0191 & 0.0375 & 0.0722 & 0.1382 & 0.2645 & 0.5089 & 1.0000 & 0.4640 \\ 0.0041 & 0.0089 & 0.0174 & 0.0335 & 0.0641 & 0.1227 & 0.2361 & 0.4640 & 1.0000 \end{bmatrix}$$

 $k = 6, \rho = .90$ 

$$\mathbf{R}_4 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ 1.0000 & 0.9276 & 0.8944 & 0.8681 & 0.8371 & 0.7765 \\ 0.9276 & 1.0000 & 0.9642 & 0.9359 & 0.9024 & 0.8371 \\ 0.8944 & 0.9642 & 1.0000 & 0.9706 & 0.9359 & 0.8681 \\ 0.8681 & 0.9359 & 0.9706 & 1.0000 & 0.9642 & 0.8944 \\ 0.8371 & 0.9024 & 0.9359 & 0.9642 & 1.0000 & 0.9276 \\ 0.7765 & 0.8371 & 0.8681 & 0.8944 & 0.9276 & 1.0000 \end{bmatrix}$$

$$k = 6, \rho = .50$$

$$\mathbf{R}_5 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ 1.0000 & 0.6547 & 0.4849 & 0.3689 & 0.2732 & 0.1789 \\ 0.6547 & 1.0000 & 0.7407 & 0.5634 & 0.4173 & 0.2732 \\ 0.4849 & 0.7407 & 1.0000 & 0.7606 & 0.5634 & 0.3689 \\ 0.3689 & 0.5634 & 0.7606 & 1.0000 & 0.7407 & 0.4849 \\ 0.2732 & 0.4173 & 0.5634 & 0.7407 & 1.0000 & 0.6547 \\ 0.1789 & 0.2732 & 0.3689 & 0.4849 & 0.6547 & 1.0000 \end{bmatrix}$$

$$k = 6, \rho = .20$$

$$\mathbf{R}_6 = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ 1.0000 & 0.3905 & 0.1633 & 0.0690 & 0.0288 & 0.0113 \\ 0.3905 & 1.0000 & 0.4183 & 0.1766 & 0.0739 & 0.0288 \\ 0.1633 & 0.4183 & 1.0000 & 0.4222 & 0.1766 & 0.0690 \\ 0.0690 & 0.1766 & 0.4222 & 1.0000 & 0.4183 & 0.1633 \\ 0.0288 & 0.0739 & 0.1766 & 0.4183 & 1.0000 & 0.3905 \\ 0.0113 & 0.0288 & 0.0690 & 0.1633 & 0.3905 & 1.0000 \end{bmatrix}$$

$$k = 3, \rho = .90$$

$$\mathbf{R}_7 = \begin{bmatrix} b_1 & b_2 & b_3 \\ 1.0000 & 0.9249 & 0.8554 \\ 0.9249 & 1.0000 & 0.9249 \\ 0.8554 & 0.9249 & 1.0000 \end{bmatrix}$$

$$k = 3, \rho = .50$$

$$\mathbf{R}_8 = \begin{bmatrix} b_1 & b_2 & b_3 \\ 1.0000 & 0.5851 & 0.3423 \\ 0.5851 & 1.0000 & 0.5851 \\ 0.3423 & 0.5851 & 1.0000 \end{bmatrix}$$

$$k = 3, \rho = .20$$

$$\mathbf{R}_9 = \begin{bmatrix} b_1 & b_2 & b_3 \\ 1.0000 & 0.2659 & 0.0707 \\ 0.2659 & 1.0000 & 0.2659 \\ 0.0707 & 0.2659 & 1.0000 \end{bmatrix}$$

A sample-specific random effect  $\tau^2$  was generated for each sample's set of effect size estimates  $\tau^2 \sim N(0, p \times v)$ , where  $v$  is the within-study variance and  $p$  is an arbitrary proportion. For this study,  $p$  took the value of .00, .50, and 1.00, which correspond to  $I^2$  values of .00, .33, and .50, respectively.

*Focal slope approach.* This approach involved generating  $n$  primary study units from a  $k$ -variate normal distribution, each having a mean of 0 and variance of  $1 - \rho$ , for each of  $m$  studies included in each meta-analysis, where  $\rho$  is correlation between any two primary units. A population covariance matrix for potential model covariates,  $\Sigma$ , was specified as

$$\Sigma = \mathbf{R} \times (1 - \rho) =$$

$$\begin{bmatrix} Y & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 & X_8 & X_9 & X_{10} \\ 1.00 & .80 & .60 & .50 & .40 & .10 & .00 & .20 & .10 & .10 & .00 \\ .80 & 1.00 & .40 & .30 & .20 & .10 & .30 & .20 & .20 & .10 & .10 \\ .60 & .40 & 1.00 & .30 & .30 & .20 & .10 & .10 & .20 & .10 & .10 \\ .50 & .30 & .30 & 1.00 & .30 & .40 & .20 & .10 & .20 & .20 & .00 \\ .40 & .20 & .30 & .30 & 1.00 & .50 & .40 & .30 & .20 & .20 & .10 \\ .10 & .10 & .20 & .40 & .50 & 1.00 & .10 & .20 & .10 & .20 & .10 \\ .00 & .30 & .10 & .20 & .40 & .10 & 1.00 & .40 & .30 & .20 & .20 \\ .20 & .20 & .10 & .10 & .30 & .20 & .40 & 1.00 & .20 & .10 & .20 \\ .10 & .20 & .20 & .20 & .10 & .30 & .20 & .10 & 1.00 & .20 & .00 \\ .10 & .10 & .10 & .20 & .20 & .20 & .20 & .10 & .20 & 1.00 & .10 \\ .00 & .10 & .10 & .00 & .10 & .10 & .20 & .20 & .00 & .10 & 1.00 \end{bmatrix} \times (1 - \rho)$$

For each sample, a random, normally distributed data vector,  $\zeta \sim N(0, \rho)$  was added to the  $k$ -variate data generated based on  $\Sigma$ . Data for each study then had an intra-class correlation equal to  $\rho$ .

The focal predictor for these simulations was  $X_1$ . Notice that each additional covariate,  $X_2$  to  $X_{10}$ , has a small to moderate positive correlation with either or both the focal predictor  $X_1$  or the outcome  $Y$  yielding unique parameter values for  $X_1$  under different model specifications. For example, the slope parameter in

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

differ from

$$Y_i = \beta_o + \beta_1 X_{1i} + \beta_5 X_{5i} + \beta_9 X_{9i} + e_i$$

Each sample contained  $k$  diversely specified models. Each model included  $Y$  and  $X_1$  from  $\mathbf{R}$ . Covariates were randomly selecting from  $\mathbf{R}$  thereby ensuring diverse model specification that affects the slope parameter of the focal predictor.

A sample-specific random effect  $\tau^2$  was generated for each sample's set of effect size estimates  $\tau^2 \sim N(0, p \times v)$ , where  $v$  is the within-study variance and  $p$  is an arbitrary proportion. For this study,  $p$  took the value of .00, .50, and 1.00, which correspond to  $I^2$  values of .00, .33, and .50, respectively.

For each study a vector of  $k$  slope estimates was generated using ordinary least squares

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

where  $\mathbf{X}$  is a matrix of covariates sampled from the population covariance matrix,  $\mathbf{Y}$  is the model outcome sampled from the population covariance matrix. For the focal slope approach,  $k$  estimates of  $\beta_1$  were extracted from  $\mathbf{b}$  and meta-regression indicator variables were generated based on the elements of  $\mathbf{X}$ . For the model-based approach, all elements of  $\mathbf{b}$  were extracted and modeled with meta-regression. And, the variance of slope was estimated with

$$\text{Var}(\mathbf{b}) = (\mathbf{X}^{-1}\mathbf{X})^{-1}\hat{\sigma}^2,$$

where  $\hat{\sigma}^2$ , is the estimated residual variance (mean square error in this case).

In both the focal slope approach and the model-based approach slope estimates were analyzed using meta-regression with robust standard errors from (60). And, the fixed effects weights from (68) were used throughout these simulations.

**Analyses.** Parameter recovery was the focus of these simulations. A 95% robust confidence interval, from (65), was specified for each meta-regression parameter in both the focal slope approach and the model-based approach. The performance of the robust standard errors was evaluated based on their probability of recovering the parameter value from the data on which they were used. In the focal slope approach, the parameter of interest was the unconditional slope estimate of the focal predictor. In the model-based approach,  $k$  parameters were estimated. These parameter values were equivalent to estimates based on the empirical population covariance matrix values,  $\Sigma$ .

Proportion of bias was computed for each set of conditions for both the focal slope approach and the model-based approach. This was computed as

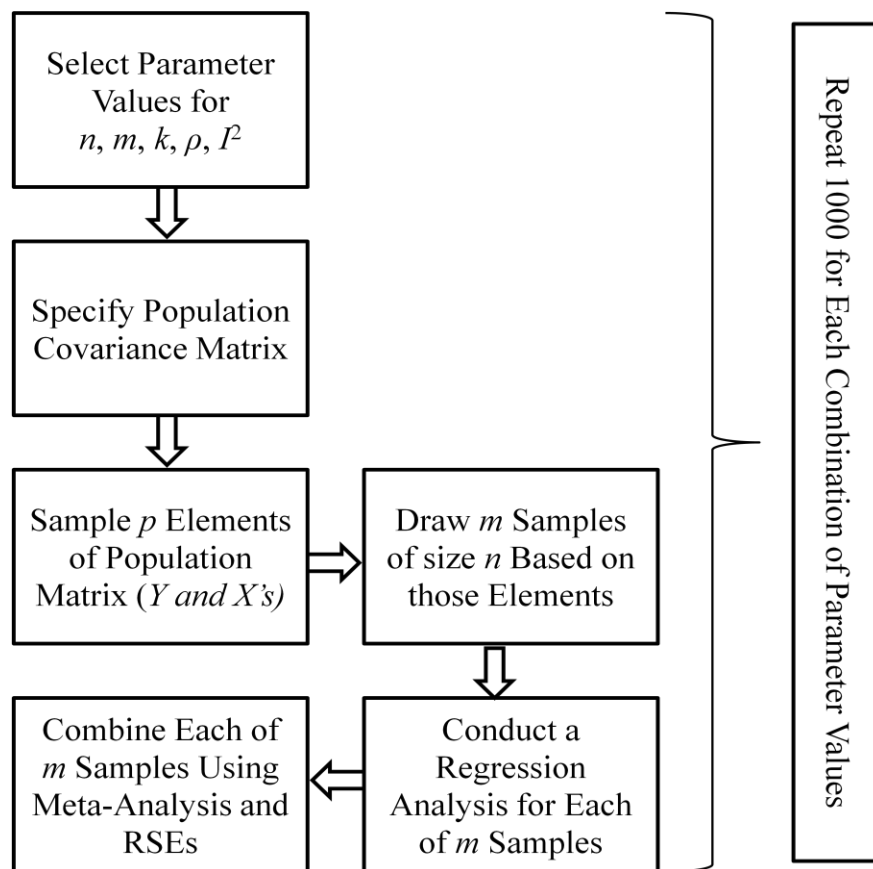
$$\text{PBIAS} = \frac{\sum_{i=1}^N (\theta_i - \hat{\theta}_i)}{\sum_{i=1}^N \theta_i}, \quad (76)$$

where  $\hat{\theta}_i$  is the meta-regression parameter estimate for the  $i$ th replicate and  $\theta_i$  is the parameter value for the  $i$ th replicate. This provides an overall estimate of the proportion of bias (PBIAS) for each of the meta-regression point estimates across all 1000 replications in each of the 243 simulation conditions. That is, the average tendency for simulated data to be larger or smaller than the specified parameter values (Gupta et al., 1999). Optimal values of PBIAS are zero (i.e. no bias), positive values indicate systematic overestimation, and negative values indicate systematic underestimation (Gupta et al., 1999). This does not inform the performance of the standard errors in these simulations but it does signal potential problems with point estimation, which is

especially important for the focal slope approach where truly heterogeneous parameter estimates are combined.

The following diagram summarizes the general data generation process for both meta-analytic approaches:

Figure 6. Simulation Data Generation and Analysis Steps



## CHAPTER FOUR

### RESULTS

This chapter presents the results of Phase I and Phase II of the analyses outlined in Chapter 3. The first half of this chapter reports the results of two applications of using robust standard errors to combine multiple regression estimates (Phase I). The second half of this chapter reports, through a series of simulations, on the performance of the robust variance estimator for combining regression estimates with meta-analysis.

#### **Phase I results**

The following section presents the results from the two applications from Phase I. The model-based approach results are presented first, followed by the focal slope approach.

**Model-based example.** Twelve independent samples, included in Link and Mulligan (1986), synthesized using meta-analysis and robust standard errors. A common educational production function model was specified for each sample. The following table presents some simple descriptive statistics, that were reported, for these samples.



Table 3. Link and Mulligan (1986) Sample Characteristics

<i>Sample</i>	<i>n</i>
Grade 3 – White	1,531
Grade 3 – African American	268
Grade 3 – Latino	103
Grade 4 – White	1,387
Grade 4 – African American	236
Grade 4 – Latino	106
Grade 5 – White	1,513
Grade 5 – African American	266
Grade 5 – Latino	106
Grade 6 – White	1,981
Grade 6 – African American	222
Grade 6 – Latino	109

The following steps were taken to combine these 12 samples. First, information from each sample regarding the slope estimate, the variance of the slope estimate, the number of slope estimates, the mean of the slope estimate variances, and the identity of each slope was entered into a data file. Second, weights for each of the slope estimates were computed using (68), where each effect size in each study receives the same weight. The following linear model was then used to estimate the conditional mean effect size estimates

$$\mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} ,$$

where  $\mathbf{b}$  is the vector of effect size estimates (raw slope estimates),  $\mathbf{X}$  is a identity matrix containing dummy indicators for each of the model predictors (number of books in the home, gender, mother's education, compensatory education status, placement in large classroom with low achievement, teacher experience, weekly hours spent in mathematics, and weekly hours spent in mathematics squared) and  $\boldsymbol{\beta}$  is a vector of unknown meta-

regression slope estimates for each of the predictors included in  $\mathbf{X}$ . Robust standard errors for each of the meta-regression slope estimates were estimated using (58).

Table 4 presents the combined results for Link and Mulligan (1986). The mean slope estimates for each predictor were not statistically significant with the exception of the mathematics pretest ( $\bar{b} = .7729$ ,  $CI = 0.5795, .9662$ ). Teacher experience, the school input in this example, was not statistically significant ( $\bar{b} = .7846$ ,  $CI = -1.7590, 3.3281$ ). The between-studies variance component,  $\tau^2$ , was invariant to different values of  $\rho$  within the first four decimal places.

Table 4. Meta-Regression Analysis of Link and Mulligan (1986).

	<i>Mean Slope Estimate</i>	<i>Robust SE</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>
Books in Home	0.0298	0.0201	-0.0096	0.0691
Gender (Male)	-1.0621	0.7401	-2.5128	0.3886
Mother's Education	1.0795	1.3374	-1.5419	3.7008
Compensatory Education	-4.5454	1.6786	-7.8355	-1.2553
Math Pretest	0.7729	0.0987	0.5795	0.9662
Low Class Achievement	-1.4331	1.0378	-3.4673	0.6010
Teacher Experience (6+ Years)	0.7846	1.2977	-1.7590	3.3281
Weekly Hours in Math	1.4346	1.5994	-1.7003	4.5694
$\tau^2 = 12.73$				

**Focal slope example.** The following table presents some descriptive statistics about the studies included in this example. Even among this small subset of studies, substantial differences among these studies remain. Note also, that two studies are at the school- and district- (or county-) level. The inferences derived from studies at different ecological levels may not be compatible. For example, the relationship between socio-economic status and student achievement at the student level may be irreconcilably

different from the relationship between average student achievement and average student socioeconomic status at the school or district level (i.e. they may be fundamentally different parameters). More importantly, the performance of this robust variance estimator has not been explored with fewer than 10 studies and it may not provide accurate estimates in situations like this. Because of the many limitations to this kind of analysis, this example is only provided for illustrative purposes.

Table 5 presents some descriptive statistics of the data used in this analysis.

Table 5. Focal Slope Meta-Analysis Study Characteristics.

	<i>n</i>	<i>Level of Analysis</i>	<i>k</i>	<i>Controls</i>
Dugan (1976)	47	School	2	Fam, Sch
Gyimah-Brempong, & Gyapong (1991)	175	District	6	Fam,Sch,
Register & Grimes (1991)	2360	Student	1	Fam, Sch, Stu,
Ribich & Murphy (1975)	9527	Student	3	Fam, Stu,
	194-			
Ritzen & Winkler (1977)	217	Student	4	Fam, Sch, Tch
Grimes & Register (1990)	1620	Student	4	Fam, Sch, Stu, Tch

*Note:* Fam = Family; Sch = School; Stu = Student; Tch = Teacher

The following linear model was used to estimate the conditional mean effect size estimate:

$$\mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} ,$$

where  $\mathbf{b}$  is the vector of effect size estimates (raw slope estimates),  $\mathbf{X}$  is a design matrix containing dummy indicators for each of the model predictor categories used in each of regression models, and  $\boldsymbol{\beta}$  is a vector of unknown meta-regression slope estimates for each of the predictors included in  $\mathbf{X}$ . Robust standard errors for each of the meta-regression slope estimates were estimated using (58).

Table 6 presents the conditional mean slope estimates from the meta-regression analysis of focal slope data.

Table 6. Meta-Regression Analysis of Link and Mulligan (1986).

	<i>Mean Estimate</i>	<i>Robust SE</i>	<i>Lower 95% CI</i>	<i>Upper 95% CI</i>
Intercept	-0.0122	0.0140	-0.0396	0.0153
Family	-0.0026	0.0127	-0.0275	0.0224
Student	0.0146	0.0255	-0.0353	0.0645
Teacher	0.0034	0.0047	-0.0058	0.0127
School	0.0013	0.0000	0.0013	0.0013

$\tau^2 < 0.0001$

These results indicate that the estimated unconditional relationship between PPE and student achievement is statistically zero. Including school-related covariates may have positive effect on the half-standardized PPE slope estimates, holding constant the effects of the other meta-regression model covariates. The between-studies variance component,  $\tau^2$ , was invariant to different values of  $\rho$  within the first four decimal places.

## Phase II Results

The following section details the simulation study results from Phase II of these analyses. The first part of this section presents the results of the model-based approach simulations and the second part presents the results of the focal slope approach simulations.

**Model-based approach.** This section presents the results of the simulations for the model-based approach. The simulations have been grouped into smaller pieces in this section.

Table 7 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$  without any between-sample heterogeneity and where  $k$  is set equal to 3. The robust variance estimator provided confidence intervals that recovered the parameter value almost exactly 95% of the time in each condition presented. PBIAS was nearly zero for each of these conditions, too.

Table 7. Model-Based Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 3$

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0081	0.9540	0.0040	0.9523	0.0043	0.9497
12	60	-0.0065	0.9537	-0.0003	0.9580	0.0001	0.9593
12	90	-0.0051	0.9583	-0.0026	0.9580	-0.0022	0.9567
20	30	0.0122	0.9423	0.0013	0.9550	0.0026	0.9480
20	60	0.0041	0.9457	0.0022	0.9493	0.0004	0.9503
20	90	0.0043	0.9520	0.0027	0.9487	0.0013	0.9543
40	30	-0.0066	0.9540	0.0032	0.9457	-0.0010	0.9400
40	60	-0.0009	0.9550	0.0008	0.9587	0.0010	0.9490
40	90	0.0008	0.9497	-0.0024	0.9553	0.0001	0.9557

Table 8 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$  without any between-sample heterogeneity and where  $k$  is set equal to 6. Coverage probabilities in each of these conditions were close to the nominal .95. When the number of samples,  $m$ , was large, the robust standard errors were most efficient but largely unaffected by values of the other parameter values. PBIAS in each of these conditions were near zero.

Table 8. Model-Based Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 6$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	-0.0049	0.9837	-0.0046	0.9857	0.0028	0.9812
12	60	-0.0029	0.9880	-0.0026	0.9840	0.0061	0.9832
12	90	0.0022	0.9855	-0.0007	0.9882	-0.0061	0.9907
20	30	0.0044	0.9670	0.0018	0.9710	-0.0044	0.9680
20	60	0.0036	0.9715	-0.0031	0.9703	-0.0018	0.9703
20	90	0.0003	0.9692	0.0023	0.9733	-0.0004	0.9753
40	30	0.0016	0.9677	0.0069	0.9628	-0.0012	0.9515
40	60	-0.0016	0.9583	0.0014	0.9573	0.0032	0.9570
40	90	-0.0017	0.9625	-0.0020	0.9630	-0.0004	0.9630

Table 9 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$  without any between-sample heterogeneity and where  $k$  is set equal to 9. Coverage probabilities in each of these conditions were at least .95. When  $m$  was small, again, the performance of the robust variance estimator was least efficient. PBIAS was consistently near zero in each of these conditions.

Table 9. Model-Based Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 9$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0053	0.9989	0.0015	0.9990	0.0006	0.9994
12	60	0.0017	0.9988	-0.0037	0.9991	0.0007	0.9987
12	90	-0.0004	0.9993	-0.0030	0.9997	0.0002	0.9991
20	30	0.0026	0.9872	0.0076	0.9819	-0.0002	0.9778
20	60	-0.0003	0.9829	0.0000	0.9883	0.0003	0.9848
20	90	-0.0003	0.9857	-0.0005	0.9869	0.0005	0.9856
40	30	0.0021	0.9677	-0.0007	0.9582	0.0000	0.9700
40	60	-0.0017	0.9697	0.0003	0.9683	0.0000	0.9726
40	90	-0.0015	0.9673	-0.0004	0.9648	-0.0003	0.9694

Table 10 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 3. In the presence of moderate between-sample heterogeneity, the performance of the robust variance estimator is comparable to the conditions using the same parameter values and no between-sample heterogeneity. Coverage probabilities were close to .95 and PBIAS was always near zero.

Table 10. Model-Based Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 3$

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	-0.0035	0.9527	-0.0018	0.9627	-0.0063	0.9583
12	60	-0.0047	0.9637	-0.0013	0.9560	-0.0007	0.9480
12	90	-0.0012	0.9590	-0.0028	0.9540	-0.0023	0.9633
20	30	-0.0032	0.9650	0.0014	0.9470	0.0003	0.9600
20	60	0.0035	0.9507	0.0018	0.9527	-0.0025	0.9617
20	90	-0.0057	0.9460	0.0013	0.9547	-0.0022	0.9520
40	30	-0.0025	0.9537	0.0016	0.9567	-0.0011	0.9420
40	60	-0.0002	0.9610	-0.0003	0.9507	-0.0007	0.9533
40	90	0.0008	0.9463	0.0005	0.9513	0.0001	0.9523

Table 11 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 6. All coverage probabilities were still near or above .95 and PBIAS was always near zero. These results differ minimally from the conditions with zero between-sample heterogeneity.

Table 11. Model-Based Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 6$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0059	0.9852	0.0077	0.9838	-0.0061	0.9845
12	60	-0.0005	0.9843	0.0037	0.9820	0.0015	0.9835
12	90	-0.0019	0.9853	-0.0021	0.9875	0.0047	0.9795
20	30	-0.0043	0.9738	-0.0006	0.9745	-0.0028	0.9728
20	60	0.0007	0.9712	-0.0030	0.9705	-0.0011	0.9750
20	90	0.0001	0.9728	-0.0016	0.9748	-0.0003	0.9737
40	30	-0.0008	0.9625	-0.0002	0.9625	0.0051	0.9687
40	60	-0.0011	0.9595	-0.0021	0.9582	-0.0002	0.9593
40	90	0.0010	0.9603	0.0002	0.9588	0.0005	0.9547

Table 12 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 9. These results were also comparable to the conditions where between-sample heterogeneity was set to zero. The robust standard errors are accurate in each condition but more efficient with larger numbers of samples. PBIAS was always near zero.

Table 12. Model-Based Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 9$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0026	0.9997	0.0008	0.9991	0.0006	0.9997
12	60	0.0007	0.9990	-0.0019	0.9992	-0.0003	0.9989
12	90	-0.0027	0.9989	0.0048	0.9993	-0.0008	0.9999
20	30	-0.0026	0.9862	0.0030	0.9870	0.0009	0.9846
20	60	-0.0021	0.9862	-0.0013	0.9851	0.0007	0.9813
20	90	-0.0015	0.9879	0.0007	0.9872	-0.0002	0.9876
40	30	0.0004	0.9699	-0.0013	0.9663	-0.0003	0.9673
40	60	0.0021	0.9642	0.0005	0.9683	-0.0004	0.9714
40	90	0.0004	0.9681	0.0020	0.9656	-0.0001	0.9684



Table 13 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 3. Not unlike the two sets of conditions previously presented where  $k = 3$  and between-sample heterogeneity was zero or moderate, these results continue to show accurate and efficient estimation. The coverage probabilities in each of these conditions are near .95 and PBIAS is consistently near zero.

Table 13. Model-Based Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 3$

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0021	0.9567	0.0060	0.9587	0.0013	0.9560
12	60	-0.0080	0.9527	-0.0001	0.9597	-0.0008	0.9557
12	90	0.0015	0.9577	-0.0060	0.9583	-0.0024	0.9490
20	30	0.0010	0.9503	0.0057	0.9607	0.0022	0.9503
20	60	0.0024	0.9577	-0.0011	0.9503	-0.0009	0.9517
20	90	0.0006	0.9517	0.0020	0.9540	-0.0007	0.9527
40	30	-0.0044	0.9527	-0.0012	0.9540	-0.0035	0.9573
40	60	-0.0003	0.9503	-0.0020	0.9580	-0.0006	0.9517
40	90	-0.0004	0.9550	-0.0003	0.9520	-0.0016	0.9483

Table 14 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 6. Coverage probabilities were always at least .95 and the estimator was again most efficient when the number of studies included in the meta-analysis was larger. PBIAS was consistently near zero.

Table 14. Model-Based Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 6$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0068	0.9843	-0.0011	0.9875	0.0116	0.9843
12	60	0.0050	0.9862	-0.0059	0.9862	0.0043	0.9877
12	90	-0.0034	0.9890	-0.0007	0.9855	0.0022	0.9862
20	30	0.0009	0.9695	0.0068	0.9685	-0.0018	0.9713
20	60	-0.0006	0.9760	0.0025	0.9697	-0.0030	0.9670
20	90	0.0009	0.9717	0.0025	0.9720	0.0008	0.9728
40	30	-0.0001	0.9630	0.0039	0.9540	0.0019	0.9602
40	60	-0.0012	0.9553	0.0012	0.9667	0.0007	0.9603
40	90	0.0001	0.9567	0.0018	0.9602	-0.0047	0.9603

Table 15 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 9. The robust variance estimator was least efficient under these conditions, as might be expected. However, the standard errors were accurate in each condition and were most efficient when the number of studies included in the meta-analysis was large. PBIAS was continued to be near zero in each of these conditions.

Table 15. Model-Based Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 9$ 

$m$	$n$	$\rho = .2$		$\rho = .5$		$\rho = .9$	
		PBIAS	Coverage	PBIAS	Coverage	PBIAS	Coverage
12	30	0.0024	0.9997	0.0002	0.9996	0.0010	0.9994
12	60	0.0029	0.9993	-0.0005	0.9991	-0.0001	0.9992
12	90	-0.0003	0.9988	0.0003	0.9997	-0.0003	0.9993
20	30	0.0054	0.9858	0.0002	0.9878	-0.0006	0.9842
20	60	-0.0004	0.9857	-0.0017	0.9876	-0.0003	0.9868
20	90	0.0012	0.9878	-0.0007	0.9880	-0.0001	0.9864
40	30	-0.0010	0.9699	-0.0005	0.9713	0.0000	0.9723
40	60	-0.0017	0.9710	-0.0011	0.9737	0.0002	0.9717
40	90	0.0004	0.9688	-0.0010	0.9637	-0.0001	0.9677

**Summary.** In all, the robust variance estimator performed well under the different conditions specified for these simulations. Under specified variation in  $n$ ,  $m$ ,  $k$ ,  $\rho$ ,  $I^2$  the nominal probability content was usually close to 95 percent. The values of  $n$ ,  $\rho$ , and  $I^2$  had little effect on the robust standard errors for the meta-regression model coefficients in these analyses. Coverage was most dependent on  $m$  and  $k$ , not unexpectedly. Standard errors were smallest when the ratio of  $m$  to  $k$  is smallest (i.e.  $m = 12$ ,  $k = 9$ ). PBIAS was not systematically positive or negative for these simulations. And, it was consistently near zero.

None of these conditions led to a dramatic increase in the probability of type I errors. When the intervals departed at all markedly from nominal probability content, it was in the conservative direction. That is, robust confidence intervals were never excessively small in these simulations but sometimes larger than they were expected to be.

These results are promising for this type of meta-analysis. While Becker and Wu's GLS method requires the slope variance-covariance matrix to accurately produce meta-regression standard errors, the robust standard errors in these simulations performed well without any information about the slope variance-covariance matrix.

**Focal slope approach.** This section presents the results of the simulations for the focal slope approach. The results are grouped into smaller pieces throughout this section, grouping results into separate tables according to values  $\tau^2/I^2$  and  $k$ .

In running these analyses, an immediate and obvious result was that the point estimates were systematically biased with the estimand ( $\theta$ ) consistently being about 2% larger than the estimate ( $\hat{\theta}$ ). However, this result does not imply that the standard errors were systematically flawed, too. To facilitate a complete investigation of the performance of the robust variance estimator in these conditions, the point estimates were adjusted by adding the PBIAS value as to each of the point estimates in each of the simulation conditions. This systematically minimizes the bias and allows for a closer look at the coverage probabilities of the robust confidence intervals. The unadjusted and adjusted coverage probabilities are presented in each of the tables below.

Table 16 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with no between-sample heterogeneity, and where  $k$  is set equal to 3. The adjusted coverage probabilities across the conditions presented in this table are close to or above the specified .95. Adjusted coverage is closest to .95 when the number of studies is large and coverage appears to be relatively unaffected by values of  $n$  and  $\rho$ .

Table 16. Focal Slope Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 3$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0194	0.9940	0.9960	0.0197	0.9970	0.9970	0.0240	0.9930	0.9910
12	60	0.0220	0.9850	0.9920	0.0222	0.9960	0.9930	0.0208	0.9900	0.9910
12	90	0.0182	0.9840	0.9920	0.0209	0.9940	0.9960	0.0227	0.9860	0.9980
20	30	0.0222	0.9600	0.9680	0.0238	0.9580	0.9740	0.0218	0.9470	0.9640
20	60	0.0217	0.9340	0.9690	0.0203	0.9460	0.9750	0.0233	0.9340	0.9570
20	90	0.0218	0.9310	0.9580	0.0198	0.9400	0.9720	0.0205	0.9270	0.9660
40	30	0.0203	0.9250	0.9490	0.0220	0.9210	0.9450	0.0237	0.9100	0.9470
40	60	0.0216	0.8880	0.9580	0.0210	0.8740	0.9410	0.0221	0.8780	0.9450
40	90	0.0212	0.8440	0.9530	0.0212	0.8400	0.9620	0.0209	0.8440	0.9540

Table 17 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with no between-sample heterogeneity, and where  $k$  is set equal to 6. Adjusted coverage probabilities are still close to or above the specified .95 and the estimator's efficiency again appears to be driven by the number of study samples included in the analysis more than anything else.

Table 17. Focal Slope Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 6$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0220	0.9930	0.9970	0.0227	0.9950	0.9990	0.0232	0.9960	0.9940
12	60	0.0175	0.9940	0.9980	0.0207	0.9970	0.9980	0.0197	0.9940	0.9990
12	90	0.0218	0.9940	0.9970	0.0208	0.9950	1.0000	0.0203	0.9950	0.9990
20	30	0.0219	0.9600	0.9740	0.0239	0.9490	0.9670	0.0215	0.9510	0.9650
20	60	0.0224	0.9370	0.9690	0.0209	0.9420	0.9760	0.0219	0.9490	0.9760
20	90	0.0215	0.9000	0.9750	0.0204	0.9100	0.9820	0.0202	0.9250	0.9650
40	30	0.0234	0.9120	0.9530	0.0226	0.9200	0.9480	0.0220	0.9230	0.9580
40	60	0.0215	0.8640	0.9620	0.0210	0.8620	0.9550	0.0219	0.8560	0.9450
40	90	0.0202	0.8210	0.9570	0.0220	0.7850	0.9570	0.0214	0.7920	0.9650

Table 18 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with no between-sample heterogeneity, and where  $k$  is set equal to 9. Adjusted coverage probabilities were close to or above .95 in all conditions.

Table 18. Focal Slope Approach:  $\tau^2 = 0 \times v$ ;  $I^2 = 0$ ;  $k = 9$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0231	0.9960	0.9980	0.0245	0.9940	0.9990	0.0210	0.9970	0.9990
12	60	0.0199	0.9950	0.9980	0.0204	0.9970	1.0000	0.0227	0.9930	1.0000
12	90	0.0208	0.9950	0.9990	0.0201	0.9970	0.9990	0.0210	0.9880	0.9990
20	30	0.0234	0.9650	0.9820	0.0235	0.9560	0.9650	0.0210	0.9710	0.9740
20	60	0.0199	0.9500	0.9690	0.0219	0.9530	0.9820	0.0227	0.9340	0.9770
20	90	0.0201	0.9340	0.9770	0.0198	0.9170	0.9760	0.0216	0.9200	0.9750
40	30	0.0224	0.8990	0.9550	0.0244	0.9020	0.9590	0.0227	0.8980	0.9500
40	60	0.0212	0.8650	0.9500	0.0213	0.8480	0.9670	0.0210	0.8580	0.9530
40	90	0.0204	0.7890	0.9560	0.0212	0.7810	0.9440	0.0205	0.7760	0.9580

Table 19 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 3. The adjusted coverage probabilities in these conditions are also close to .95 or larger, indicating at least accurate though sometimes inefficient standard errors.

Table 19. Focal Slope Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 3$

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0185	0.9910	0.9910	0.0220	0.9930	0.9960	0.0139	0.9970	0.9950
12	60	0.0246	0.9940	0.9960	0.0211	0.9950	0.9990	0.0188	0.9960	0.9960
12	90	0.0195	0.9910	0.9940	0.0234	0.9960	0.9960	0.0216	0.9930	0.9960
20	30	0.0168	0.9620	0.9710	0.0234	0.9690	0.9700	0.0222	0.9600	0.9580
20	60	0.0217	0.9540	0.9630	0.0201	0.9540	0.9730	0.0214	0.9610	0.9730
20	90	0.0221	0.9500	0.9630	0.0209	0.9500	0.9770	0.0212	0.9560	0.9710
40	30	0.0194	0.9370	0.9500	0.0198	0.9350	0.9520	0.0212	0.9420	0.9660
40	60	0.0204	0.9310	0.9570	0.0223	0.9260	0.9590	0.0202	0.9310	0.9600
40	90	0.0197	0.9060	0.9540	0.0191	0.9270	0.9550	0.0189	0.9150	0.9600

Table 20 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 6. Adjusted coverage probabilities were close to or above .95 in all conditions. Notice that in one condition, when  $\rho = .9$ ,  $n = 30$ , and  $m = 12$ , the robust confidence intervals captured the population value 100% of the time.

Table 20. Focal Slope Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 6$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0129	0.9970	0.9990	0.0179	0.9990	0.9970	0.0191	1.0000	1.0000
12	60	0.0227	0.9950	0.9970	0.0181	0.9960	0.9970	0.0186	0.9960	0.9970
12	90	0.0213	0.9950	0.9980	0.0176	0.9960	0.9970	0.0175	0.9980	0.9980
20	30	0.0217	0.9690	0.9780	0.0171	0.9750	0.9770	0.0191	0.9830	0.9790
20	60	0.0206	0.9690	0.9660	0.0215	0.9620	0.9810	0.0217	0.9760	0.9780
20	90	0.0209	0.9690	0.9780	0.0208	0.9580	0.9760	0.0197	0.9670	0.9700
40	30	0.0211	0.9350	0.9500	0.0187	0.9560	0.9620	0.0196	0.9490	0.9610
40	60	0.0195	0.9210	0.9450	0.0194	0.9350	0.9620	0.0214	0.9220	0.9540
40	90	0.0190	0.9070	0.9550	0.0208	0.9150	0.9660	0.0215	0.9190	0.9610

Table 21 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with moderate between-sample heterogeneity, and where  $k$  is set equal to 9. Adjusted coverage probabilities were close to or above .95 in all conditions.

Table 21. Focal Slope Approach:  $\tau^2 = .5 \times v$ ;  $I^2 = .33$ ;  $k = 9$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0234	0.9950	0.9970	0.0202	0.9950	0.9950	0.0239	0.9960	0.9970
12	60	0.0179	0.9970	0.9970	0.0186	0.9950	0.9980	0.0202	0.9950	0.9960
12	90	0.0179	0.9950	0.9980	0.0192	0.9940	0.9960	0.0191	0.9960	0.9970
20	30	0.0218	0.9700	0.9720	0.0222	0.9730	0.9830	0.0207	0.9720	0.9780
20	60	0.0202	0.9660	0.9680	0.0200	0.9780	0.9790	0.0199	0.9720	0.9870
20	90	0.0190	0.9600	0.9700	0.0191	0.9670	0.9700	0.0207	0.9600	0.9630
40	30	0.0218	0.9540	0.9620	0.0208	0.9530	0.9660	0.0191	0.9580	0.9620
40	60	0.0180	0.9410	0.9660	0.0215	0.9330	0.9660	0.0202	0.9410	0.9660
40	90	0.0211	0.9230	0.9650	0.0189	0.9320	0.9610	0.0200	0.9220	0.9600



Table 22 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 3. Adjusted coverage probabilities were close to or above .95 in all conditions. The coverage probabilities increased slightly as values of  $\tau^2$  increased.

Table 22. Focal Slope Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 3$

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0199	0.9970	0.9960	0.0326	0.9940	0.9940	0.0226	0.9950	0.9960
12	60	0.0173	0.9970	0.9960	0.0215	0.9900	0.9910	0.0178	0.9950	0.9960
12	90	0.0199	0.9910	0.9930	0.0204	0.9940	0.9950	0.0216	0.9950	0.9980
20	30	0.0221	0.9670	0.9760	0.0208	0.9710	0.9730	0.0211	0.9600	0.9650
20	60	0.0220	0.9560	0.9680	0.0212	0.9560	0.9560	0.0200	0.9620	0.9720
20	90	0.0198	0.9520	0.9620	0.0187	0.9730	0.9870	0.0200	0.9510	0.9600
40	30	0.0169	0.9500	0.9550	0.0220	0.9460	0.9530	0.0177	0.9550	0.9560
40	60	0.0219	0.9290	0.9570	0.0208	0.9310	0.9580	0.0210	0.9370	0.9570
40	90	0.0200	0.9290	0.9640	0.0194	0.9370	0.9590	0.0188	0.9350	0.9600

Table 23 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 6. Adjusted coverage probabilities were close to or above .95 in all conditions.

Table 23. Focal Slope Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 6$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0221	0.9970	0.9980	0.0264	1.0000	1.0000	0.0175	0.9980	0.9990
12	60	0.0220	0.9960	0.9970	0.0210	0.9970	0.9960	0.0197	0.9970	0.9960
12	90	0.0240	0.9940	0.9970	0.0243	0.9960	0.9930	0.0205	0.9960	0.9960
20	30	0.0273	0.9770	0.9760	0.0174	0.9720	0.9740	0.0208	0.9700	0.9690
20	60	0.0178	0.9760	0.9730	0.0181	0.9740	0.9720	0.0196	0.9750	0.9830
20	90	0.0187	0.9600	0.9720	0.0221	0.9610	0.9670	0.0177	0.9660	0.9750
40	30	0.0239	0.9470	0.9580	0.0187	0.9570	0.9530	0.0189	0.9380	0.9550
40	60	0.0198	0.9550	0.9630	0.0156	0.9470	0.9570	0.0191	0.9500	0.9580
40	90	0.0214	0.9280	0.9510	0.0189	0.9400	0.9550	0.0201	0.9350	0.9510

Table 24 presents the results of the simulations across all values of  $m$ ,  $n$ , and  $\rho$ , with large between-sample heterogeneity, and where  $k$  is set equal to 9. Adjusted coverage probabilities were close to or above .95 in all conditions and seem to increase in this scenario at least partly because of the value of  $\rho$ .

Table 24. Focal Slope Approach:  $\tau^2 = 1 \times v$ ;  $I^2 = .5$ ;  $k = 9$ 

		$\rho = .2$			$\rho = .5$			$\rho = .9$		
$m$	$n$	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage	PBIAS	Coverage	Adj. Coverage
12	30	0.0230	0.9960	0.9970	0.0147	0.9960	0.9960	0.0248	1.0000	1.0000
12	60	0.0165	0.9960	0.9990	0.0205	0.9950	0.9960	0.0214	0.9960	0.9990
12	90	0.0203	0.9960	0.9990	0.0194	0.9970	0.9980	0.0218	0.9960	0.9990
20	30	0.0197	0.9710	0.9710	0.0178	0.9780	0.9730	0.0091	0.9840	0.9830
20	60	0.0159	0.9690	0.9730	0.0229	0.9690	0.9730	0.0222	0.9710	0.9830
20	90	0.0195	0.9640	0.9750	0.0196	0.9660	0.9810	0.0232	0.9610	0.9740
40	30	0.0168	0.9640	0.9620	0.0160	0.9710	0.9680	0.0212	0.9490	0.9560
40	60	0.0223	0.9540	0.9660	0.0212	0.9470	0.9520	0.0202	0.9580	0.9690
40	90	0.0205	0.9350	0.9500	0.0207	0.9320	0.9620	0.0203	0.9430	0.9680

**Summary.** The results of these simulations indicate that the focal slope approach is a biased method for combining regression slope estimates with meta-analysis. PBIS is systematically positive, indicated a general over estimation of the unconditional relationship between the focal predictor and the outcome. And, the bias in these conditions was substantially greater than the bias in the model-based simulation conditions. This stems from combining slope estimates from diversely specified models, and therefore estimates of diverse parameters. This bias is further highlighted by the coverage probabilities of the intervals for the truly fixed effects scenario (i.e.  $I^2 = 0$ ). Coverage probabilities are lowest when  $m$  is large (e.g.  $m = 40$ ). In these scenarios, there may be a substantial increase in the probability of a type I error.

If considered as a missing data problem (e.g. Wu, 2006; Wu & Pigott, 2008), the limitations of the dummy variable approach used here become more apparent. Jones (1996) showed that using dummy indicators for missing predictors in regression analysis produces biased point estimates. Allison (2001) reinforces the limitations to this approach and while it was popularized in the 1980s as a viable approach to handling missing data, it is no longer an acceptable solution.

The degree of model specification diversity likely, then, affected the bias in the point estimates. It is possible that estimates from more similar models had less bias than estimates from less similar models, but because this diversity parameter was not controlled in these studies, this relationship was not investigated.

In the presence of non-zero between-study variation (i.e.  $I^2 = .33$  and  $I^2 = .50$ ) the robust confidence intervals were most likely to cover the parameter value. While this is perhaps an interesting finding, robust variance estimation in the context of between-studies heterogeneity is no panacea for biased point estimation.

However, in looking at the adjusted coverage probabilities, there is no evidence that the estimated robust standard errors were inaccurate. If a simple adjustment could be made to point estimates using this focal slope approach to meta-regression, robust standard errors may still be useful.

## CHAPTER FIVE

### DISCUSSION

#### **Conclusions**

This study investigated the use of robust variance estimation for combining regression estimates with meta-analysis. Two approaches were explored: a focal slope approach and a model-based approach. The results of this study all indicate that the robust variance estimator proposed by Hedges, Tipton, and Johnson (2010) is an effective solution for addressing dependencies induced through multiple regression models.

In the focal slope approach, multiple dependent slope estimates were extracted from diversely specified models within the same study sample. These estimates were combined using meta-analysis and modeled using meta-regression and robust standard errors. The application presented in this paper, for illustration only, presented a number of practical issues to conducting this type of synthesis. First, when a pool of studies uses radically diverse models, the meta-analysts may have to crudely categorize the model predictors into “families” of predictors. A natural problem that goes along with this type of categorization is loss of interpretability. Second, half-standardizing the slope estimates, as was done in this application, may dramatically limit the pool of available studies. This is a problem with retaining the raw scales of the slope estimates, which has advantages in some circumstances. In this application, half-standardizing the PPE coefficients still allowed us to talk about the relationship between dollars spent and

standardized student achievement. The results of the simulations for the focal slope approach highlighted the parameter estimation bias inherent to this focal slope method. Adjusting for this bias allowed for a closer look at the performance of the robust confidence intervals across the specified scenarios. The estimator appeared to perform well across the simulation conditions. Coverage probabilities were close to or above the specified .95 in all scenarios.

In the model-based approach, commonly specified models were combined used meta-regression and robust standard errors. The application in this paper examined 12 samples from Link and Mulligan (1986). The authors specified the same model for each of their independent samples. These slope estimates were combined and a mean model with robust standard errors was produced. This implementation of this strategy was straight forward, but, not without potential difficulties in the social sciences. If the dataset used for the two applications in this paper is at all representative of the social sciences, finding a set of commonly specified models across a pool of studies is rare. This may be attributable the social science's poor record of replication research. However, one potential interesting application of this approach could be in legislatively-driven analytical models. For example, some states are building teacher evaluation systems that require specific predictors in their models. Another area where this approach might be popular is in medical research, where for example, specific genetic models are repeatedly studied across different individuals.

The results of the simulations for this approach were promising. The robust confidence intervals for each set of parameter combinations were close to nominal probability content in nearly all of the specified conditions. And, when they departed from nominal coverage, it was in the conservative direction. This was most notable when the ratio of studies  $m$  to the number of parameter estimates  $k$  was small.

The robust variance estimator outlined by Hedges, Tipton, and Johnson (2010) provides a flexible method for combining multivariate results. The results of this study indicate that using robust variance estimation is a viable alternative to the GLS methods outlined by Wu and Becker (2007), which require the slope covariance matrix.

### **Limitations**

There are several limitations both to the methods used in this paper as well as to the inferences that can be drawn from this research. The simulations conducted for this paper had the luxury of equivalent measurement scales across samples. In the social sciences, homogeneous measurements of regression model predictors and regression model outcomes across a pool of samples is rare. The methods discussed and presented in this research provide no solution to that problem. Standardizing regression coefficients, while sometimes controversial (e.g. Greenland, 1986), may be a viable solution, especially for the context of combining commonly specified regression models (i.e. the model-based approach in this paper; Kim, 2011).

These methods provide no solution to poor model specification and estimation in primary studies. If model assumptions are violated in the primary studies, those violations

will adversely affect any subsequent meta-analytic estimates. See Freedman (2006) for a vibrant discussion of using robust standard errors in the context of poorly constructed regression models.

Another problem that persists is diverse model specification. As shown in results from the simulations using the focal slope approach, biased point estimation will continue to be a problem for slope estimates extracted from diversely specified models and combined using meta-analysis, or a meta-analytic derivative such as Stanley and Jarrell's (1989) meta-regression analysis (MRA). For this reason, focal slope meta-analytic methods such as these are not recommended.

The primary study regression methods used throughout this study were ordinary least squares. It is possible, and likely, that researchers will use alternative estimation procedures (e.g. multistage modeling; hierarchical linear modeling; etc.). Estimates from models using different estimation procedures may bias the meta-analytic estimates. This may be an appropriate area for future research.

These simulations covered 243 different conditions using a wide range of the values for the parameters investigated in this research. However, other conditions may exist in which the performance of the robust variance estimator performs less favorably (e.g. with fewer than 12 study samples) or more favorably (e.g. using a different set of weights). This is a natural and unavoidable limitation of simulation studies.



**Future directions**

I have identified several areas for future research on this topic. First, and perhaps most important, an understanding of optimal robust variance weights is needed. In both approaches, the fixed effects weights from Hedges, Tipton, and Johnson (2010) were used. The estimator will work for any set of weights, but its efficiency may be maximized with an alternative weighting scheme. This warrants extended investigation. Second, larger applications of the model-based approach would help identify potential difficulties with implementing that type of meta-analysis in different fields where data and reporting may differ significantly. Third, these methods may be extrapolated to other multivariate methods such as structural equation modeling and factor analysis. Fourth, extrapolating the model-based approach to situations where multiple estimation techniques are used across primary study samples would be an additionally informative set of investigations.

APPENDIX A:  
SIMULATION CODE

```

#-----
#robust.se() is the robust variance estimator provided by Hedges, Tipton, and
#Johnson (2010)
#This function must be loaded prior to using MAREgSim() or MASlopeSim()
#-----
robust.se <- function(data, X.full, rho) {
  p = ncol(X.full) - 2
  N = max(data$study)
  sumXWX = 0
  sumXWy = 0
  sumXWJWX = 0
  sumXWVWX = 0
  sumXW.sig.m.v.WX = 0
  for (i in (1: N)) {
    tab = data[data$study == i, ]
    W = diag(tab$weights, tab$k[1])
    tab2 = X.full[X.full$study == i, ]
    tab3 = cbind(tab2[-c(1)])
    X = data.matrix(tab3)
    dimnames(X) = NULL
    y = cbind(tab$effect.size)
    one = cbind(rep(1, tab$k[1]))
    J = one %*% t(one)
    sigma = (tab$s %*% t(tab$s))
    vee = diag(tab$s^2, tab$k[1])
    SigmV = sigma - vee
    sumXWX = sumXWX + t(X) %*% W %*% X
    sumXWy = sumXWy + t(X) %*% W %*% y
    sumXWJWX = sumXWJWX + t(X) %*% W %*% J %*% W %*% X
    sumXWVWX = sumXWVWX + t(X) %*% W %*% vee %*% W %*% X
    sumXW.sig.m.v.WX = sumXW.sig.m.v.WX + t(X) %*% W %*% SigmV %*% W %*% X
  }
  b = solve(sumXWX) %*% sumXWy
  X = data.matrix(X.full[-c(1)])
  dimnames(X) = NULL
  data$pred = X %*% b
  data$e = data$effect.size - data$pred
  W = diag(data$weights)
  sumW = sum(data$weights)
  Qe = t(data$e) %*% W %*% data$e
  pval = 1 - pchisq(Qe, N * p)
  denom = sumW - sum(diag(solve(sumXWX) %*% sumXWJWX))
  term1 = (Qe - N + sum(diag(solve(sumXWX) %*% sumXWVWX))) / denom
  term2 = (sum(diag(solve(sumXWX) %*% sumXW.sig.m.v.WX))) / denom
  tau.sql = term1 + rho * term2
  tau.sq = ifelse(tau.sql < 0, 0, tau.sql)
  data$r.weights = 1 / (data$k * (data$mean.v + tau.sq))
  sumXWX.r = 0
  sumXWy.r = 0
  for (i in (1: N)) {
    tab = data[data$study == i, ]
    W = diag(tab$r.weights, tab$k[1])
    tab2 = X.full[X.full$study == i, ]
    tab3 = cbind(tab2[-c(1)])
    X = data.matrix(tab3)
    dimnames(X) = NULL
    y = cbind(tab$effect.size)
    sumXWX.r = sumXWX.r + t(X) %*% W %*% X
  }
}

```

```

sumXWy.r = sumXWy.r + t(X) %*% W %*% y
}
b.r = solve(sumXWX.r) %*% sumXWy.r
X = data.matrix(X.full[-c(1)])
dimnames(X) = NULL
data$pred.r = X %*% b.r
data$e.r = cbind(data$effect.size) - data$pred.r
sumXWeeWX.r = 0
for (i in (1:N)) {
  tab = data[data$study == i, ]
  sigma.hat.r = tab$e.r %*% t(tab$e.r)
  W = diag(tab$r.weights, tab$k[1])
  tab2 = X.full[X.full$study == i, ]
  tab3 = cbind(tab2[-c(1)])
  X = data.matrix(tab3)
  dimnames(X) = NULL
  sumXWeeWX.r = sumXWeeWX.r + t(X) %*% W %*% sigma.hat.r %*% W %*% X
}
VR.r = solve(sumXWX.r) %*% sumXWeeWX.r %*% solve(sumXWX.r)
SE = c(rep(0, p+1))
for (i in (1:(p+1))) {
  SE[i] = sqrt(VR.r[i, i]) * sqrt(N / (N - (p + 1)))
}
labels = c(colnames(X.full[2:length(X.full)]))
output = data.frame(labels, as.numeric(b.r), as.numeric(SE))
names(output) = c("beta", "estimate", "RSE")
return(list("Tau Square Estimate" = tau.sq, "Robust Standard Errors" =
  output, "Qe" = Qe, "P-Value" = pval))
}

```

```

#-----
#MARegSim() simulates meta-analyses of parallel regression models using
#robust standard errors.
#-----
MARegSim <- function(x){
  MAs = x$MAs
  m = x$m
  k = x$k
  n = x$n
  rho = x$rho
  tau2 = x$tau2
  library(MASS)
  library(Matrix)
  StudyRegSim <- function(n = n, m = m, k = k){
    rho = round(rho, 2)
    rho[k == 4 & rho == .1] = .140
    rho[k == 4 & rho == .2] = .257
    rho[k == 4 & rho == .3] = .356
    rho[k == 4 & rho == .4] = .441
    rho[k == 4 & rho == .5] = .505
    rho[k == 4 & rho == .6] = .560
    rho[k == 4 & rho == .7] = .607
    rho[k == 4 & rho == .8] = .645
    rho[k == 4 & rho == .9] = .679
    rho[k == 7 & rho == .1] = .240
    rho[k == 7 & rho == .2] = .3595
    rho[k == 7 & rho == .3] = .436
    rho[k == 7 & rho == .4] = .475
    rho[k == 7 & rho == .5] = .503
    rho[k == 7 & rho == .6] = .5211
    rho[k == 7 & rho == .7] = .534
    rho[k == 7 & rho == .8] = .5433
    rho[k == 7 & rho == .9] = .5499
    rho[k == 10 & rho == .1] = .3007
    rho[k == 10 & rho == .2] = .411
    rho[k == 10 & rho == .3] = .462
    rho[k == 10 & rho == .4] = .4871
    rho[k == 10 & rho == .5] = .5019
    rho[k == 10 & rho == .6] = .510925
    rho[k == 10 & rho == .7] = .5169
    rho[k == 10 & rho == .8] = .52092
    rho[k == 10 & rho == .9] = .5238
    cov_mat = matrix(rho, k, k)
    mu = triu(cov_mat, 2)
    cov_mat1 = tril(cov_mat, -2)
    cov_mat[as.matrix(mu) != 0] = 0
    cov_mat[as.matrix(cov_mat1) != 0] = 0
    cov_mat[1, 2:ncol(cov_mat)] = .4
    cov_mat[2:ncol(cov_mat), 1] = .4
    diag(cov_mat) = 1
    cov_mat = nearPD(cov_mat)$mat
    colnames(cov_mat) = c("y", paste("x", 2:ncol(cov_mat) - 1, sep = ""))
    rownames(cov_mat) = c("y", paste("x", 2:ncol(cov_mat) - 1, sep = ""))
    POP = data.frame(mvrnorm(1000, mu = rep(0, ncol(cov_mat)), Sigma = cov_mat,
                             empirical = T))
    POP.cor = cor(POP$x1, POP$y)
    attach(POP)
    POP.formula = as.formula(paste("y ~ ", paste(names(POP)[2:ncol(POP)],

```

```

                                collapse = "+"))
POP.fit = lm(POP.formula)
POP.params = data.frame(summary(POP.fit)$coefficients)
POP.params$Param.ID = names(POP)
POP.params = POP.params[, c("Param.ID", "Estimate")]
names(POP.params) = c("Param.ID", "param")
detach(POP)
#-----
-
#regSim() generates primary study data, which is sampled from the population
#covariance matrix (cov_mat) using a multiavriate normal distribution.
#-----
-
regSim <- function(n){
  if(n == "rand") {n = ceiling(runif(1, 30, 1000))}
  else{n = n}
  #sample elements (from x2 to xp) from the population covariance matrix
  samp = 3:ncol(cov_mat)
  #extract sampled elements from the population coviance matrix
  pop = cov_mat[c(1:2, samp), c(1:2, samp)]
  p = ncol(pop)
  #populate a primary study with n = n, each variable having mean = 0, sd =
  #sqrt(diag(cov_mat))
  pop = data.frame(mvrnorm(n, mu = rep(0, p), Sigma = pop))
  attach(pop)
  pred.names = names(pop)
  #construct the regression model for each primary study
  formula = as.formula(paste("y ~ ", paste(pred.names[2:ncol(pop)],
                                collapse = "+")))
  #fit the linear regression model (y ~ x1 + x2 + ... + xp)
  fit = lm(formula)
  detach(pop)
  #extract model coefficients and standard errors
  out = data.frame(summary(fit)$coefficients[, 1:2])
  names(out) = c("effect.size", "s")
  out = out[, c("effect.size", "s")]
  out$var.eff.size = out$s^2
  out$s.size = n
  out$Var.ID = factor(rownames(out))
  return(out)
}
#loop regSim() over m = number of primary studies to be generated
res = list()
for(i in 1:m){
  res[[i]] = regSim(n)
  res[[i]]$study = i
}
#stack simulation results into a data frame
res = data.frame(do.call("rbind",res))
#reshape predictors into dummy variables (i.e. from Var.ID which was a
#factor variable indicating what coefficient it represents
res = data.frame(model.matrix(~ effect.size + s + var.eff.size +
                                study + s.size + Var.ID, res))

p = ncol(res) - 6
colnames(res) = c("Intercept", "effect.size", "s", "var.eff.size",
                  "study", "s.size", paste("x", 1:p, sep = ""))
rownames(res) = 1:nrow(res)
#compute within-study mean variance

```

```

mean.v = aggregate(res$var.eff.size, by = list(c(res$study)), mean)
names(mean.v) = c("study", "mean.v")
res = merge(res, mean.v, by = "study")
#count the number of ESEs per study
k1 = aggregate(res$effect.size, by = list(c(res$study)), length)
names(k1) = c("study", "k")
res = merge(res, k1, by = "study")
v = sum((res$effect.size - mean(res$effect.size))^2) / (m * (k - 2))
#compute fixed effects weights per Hedges, Tipton, and Johanson (2010)
res$weights = 1 / (res$k * res$mean.v)
s2 = sum(res$weights * (m * (k - 1))) / (
  sum(res$weights)^2 - sum(res$weights^2))
re = data.frame(study = 1:m, re = rnorm(m, 0, sqrt(tau2 * s2)))
res = merge(res, re, by = "study")
res$effect.size = res$effect.size + res$re
res = res[, - ncol(res)]
#subset data into appropriate subsets for robust.se()
sub1 = res[, c("study", "k", "mean.v", "weights", "effect.size",
  "var.eff.size", "s")]
p = ncol(res) - 9
sub2 = res[, c("study", "Intercept", paste("x", 1:p, sep = ""))]
#compute robust standard errors for ESEs
rse = robust.se(sub1, sub2, .8)
RSE = data.frame(rse[2])
RSE$tau2 = unlist(rse[1])
RSE$Qe = unlist(rse[3])
res = RSE
names(res) = c("Var.ID", "Estimate", "RSE", "tau2", "Qe")
res1 = merge(res, POP.params, by.x = "Var.ID", by.y = "Param.ID")
#compute robust confidence intervals
res1$lb = res1$Estimate - 1.96 * res1$RSE
res1$ub = res1$Estimate + 1.96 * res1$RSE
#compute probability content
res1$coverage = 0
res1$coverage[res1$param >= res1$lb &
  res1$param <= res1$ub] = 1
cor.b = summary(POP.fit, correlation = T)$correlation
cor.b = cor.b[-1, -1]
diag(cor.b) = rep(NA, ncol(cor.b))
cov.b = summary(POP.fit)$cov.unscaled
return(res1)
}
#loop StudyRegSim() over MAs: number of meta-analyses
res = list()
for(i in 1:MA){
  res[[i]] = StudyRegSim(m = m, k = k, n = n)
  res[[i]]$MA = i
}
#stack meta-analyses into a dataframe
res = data.frame(do.call("rbind", res))
#return the stacked dataframe of meta-analytic estimates
return(res)
}

```

```

#-----
#MASslopeSim() simulates meta-analyses of focal slope estimates derived
#from unparallel regression models using robust standard errors.
#-----
MASslopeSim <- function(x){
  library(MASS)
  library(reshape)
  MAs = x$MAs
  n = x$n
  m = x$m
  rho = x$rho
  k = x$k
  tau2 = x$tau2
  StudyRegSim <- function(n = n, rho = rho, m = m){
    #this covariance matrix induces a new parameter estimate with each additional
    #covariate
    cov_mat = matrix(c(1.0, 0.8, 0.6, 0.5, 0.4, 0.1, 0.0, 0.2, 0.1, 0.1, 0.0,
                        0.8, 1.0, 0.4, 0.3, 0.2, 0.1, 0.3, 0.2, 0.2, 0.1, 0.1,
                        0.6, 0.4, 1.0, 0.3, 0.3, 0.2, 0.1, 0.1, 0.2, 0.1, 0.1,
                        0.5, 0.3, 0.3, 1.0, 0.3, 0.4, 0.2, 0.1, 0.2, 0.2, 0.0,
                        0.4, 0.2, 0.3, 0.3, 1.0, 0.5, 0.4, 0.3, 0.2, 0.2, 0.1,
                        0.1, 0.1, 0.2, 0.4, 0.5, 1.0, 0.1, 0.2, 0.1, 0.2, 0.1,
                        0.0, 0.3, 0.1, 0.2, 0.4, 0.1, 1.0, 0.4, 0.3, 0.2, 0.2,
                        0.2, 0.2, 0.1, 0.1, 0.3, 0.2, 0.4, 1.0, 0.2, 0.1, 0.2,
                        0.1, 0.2, 0.2, 0.2, 0.1, 0.3, 0.2, 0.1, 1.0, 0.2, 0.0,
                        0.1, 0.1, 0.1, 0.2, 0.2, 0.2, 0.2, 0.1, 0.2, 1.0, 0.1,
                        0.0, 0.1, 0.1, 0.0, 0.1, 0.1, 0.1, 0.2, 0.0, 0.1, 1.0),
                      nrow = 11, ncol = 11, byrow = T)
    if(n == "rand") {n = ceiling(runif(1, 30, 1000))}
    colnames(cov_mat) = c("y", paste("x", 1:(ncol(cov_mat) - 1), sep = ""))
    rownames(cov_mat) = c("y", paste("x", 1:(ncol(cov_mat) - 1), sep = ""))
    POP = data.frame(mvrnorm(1000, mu = rep(0, ncol(cov_mat)), Sigma = cov_mat,
                                empirical = T))
    #set unconditional parameter value (r_x1.y = .8)
    POP.cor = summary(lm(POP$y ~ POP$x1))$coefficients[2, 1]
    regSim <- function(n = n, m = m, k = k, rho = rho){
      #set variances for each variable to 1 - rho
      cov_mat = cov_mat * (1 - rho)
      diag(cov_mat) = 1 - rho
      res = list()
      fit = list()
      for(j in 1:m){
        pop = data.frame(mvrnorm(n, mu = rep(0, ncol(cov_mat)), Sigma =
                                cov_mat, empirical = F))
        names(pop) = c("y", paste("x", 2:ncol(cov_mat) - 1, sep = ""))
        pop = pop + rnorm(1, 0, sqrt(rho))
        for(i in 1:k){
          #sample elements from the population covariance matrix
          pop.samp = pop[, c(1:2, unique(ceiling(runif(runif(1, 2,
                                ncol(cov_mat))), 2, ncol(cov_mat)))))]
          attach(pop.samp)
          pred.names = names(pop.samp)
          formula = as.formula(paste("y ~ ",
                                paste(pred.names[2:ncol(pop.samp)], collapse = "+")))
          pop.fit = lm(formula)
          detach(pop.samp)
          fit[[i]] = data.frame(summary(pop.fit)$coefficients[, 1:2])
          fit[[i]]$Estimate = fit[[i]]$Estimate
        }
      }
    }
  }
}

```



```

        fit[[i]]$Var.ID = factor(rownames(fit[[i]]))
        fit[[i]]$model = i
    }
    res[[j]] = data.frame(do.call("rbind", fit))
    res[[j]]$study = j
}
res = data.frame(do.call("rbind", res))
names(res) = c("effect.size", "s", "Var.ID", "model", "study")
res = res[, c("effect.size", "s", "Var.ID", "model", "study")]
res$var.eff.size = res$s^2
res$s.size = n
rownames(res) = 1:nrow(res)
return(res)
}
res = regSim(n = n, m = m, k = k, rho = rho)
res = data.frame(model.matrix( ~ effect.size + s + var.eff.size + study +
                             s.size + model + Var.ID, res))

p = ncol(res) - 7
colnames(res) = c("Intercept", "effect.size", "s", "var.eff.size",
                  "study", "s.size", "model", paste("x", 1:p, sep = ""))
mean.v = aggregate(res$var.eff.size[res$x1 == 1],
                    by = list(c(res$study[res$x1 == 1])), mean)
rownames(res) = 1:nrow(res)
names(mean.v) = c("study", "mean.v")
res = merge(res, mean.v, by = "study")
#identify covariates used in each model in each study
res$study.mod = paste(res$study, res$model)
#dummy code meta-regression model specification covariates
covs = res[, c("study.mod", paste("x", 2:p, sep = ""))]
covs.agg = aggregate(covs[, 2:ncol(covs)],
                     by = list(c(covs$study.mod)), sum)
names(covs.agg) = names(covs)
names(covs.agg)[2:ncol(covs.agg)] =
    toupper(names(covs.agg)[2:ncol(covs.agg)])
covs = covs.agg
covs[, 2:ncol(covs)][covs[, 2:ncol(covs)] != 0] = 1
res = merge(res, covs, by = "study.mod")
res = res[res$x1 == 1, ]
k1 = aggregate(res$effect.size, by = list(c(res$study)), length)
names(k1) = c("study", "k")
res = merge(res, k1, by = "study")
#compute fixed effects weights per Hedges, Tipton, and Johnson (2010)
res$weights = 1 / (res$k * res$mean.v)
s2 = sum(res$weights * (m * k)) / (
    sum(res$weights)^2 - sum(res$weights^2))
re = data.frame(study = 1:m, re = rnorm(m, 0, sqrt(tau2 * s2)))
res = merge(res, re, by = "study")
res$effect.size = res$effect.size + res$re
res = res[, - ncol(res)]
#subset data for robust.se()
sub1 = res[, c("study", "k", "mean.v", "weights", "effect.size",
              "var.eff.size", "s")]
sub2 = res[, c("study", "Intercept", paste("X", 2:p, sep = ""))]
#compute robust standard errors for ESEs
rse = robust.se(sub1, sub2, .8)
RSE = data.frame(rse[2])
RSE$tau2 = unlist(rse[1])
res1 = RSE

```

```

names(res1) = c("Var.ID", "Estimate", "RSE", "tau2")
res1$param = POP.cor
res1 = res1[res1$Var.ID == "Intercept", ]
#compute robust confidence intervals
res1$lb = res1$Estimate - 1.96 * res1$RSE
res1$ub = res1$Estimate + 1.96 * res1$RSE
#compute probability content
res1$coverage = 0
res1$coverage[res1$param >= res1$lb &
               res1$param <= res1$ub] = 1
return(res1)
}
#loop StudyRegSim() over MAs: number of meta-analyses
res = list()
for(i in 1:MA){
  res[[i]] = StudyRegSim(m = m, n = n, rho = rho)
  res[[i]]$MA = i
}
#stack meta-analyses into a dataframe
res = data.frame(do.call("rbind",res))
#return the stacked dataframe of meta-analytic estimates
return(res)
}

```

## REFERENCES

- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, 95(4), 648-680.
- Aloe, A. M. (2009). *A partial effect size for the synthesis of multiple regression models* (Ph.D.). The Florida State University, United States -- Florida.
- Allison, P. D. (2001). *Missing Data*. SAGE.
- Berlin, J. A., Longnecker, M. P., & Greenland, S. (1993). Meta-analysis of epidemiologic dose-response data. *Epidemiology*, 4(3), 218-228.
- Bini, L. M., Coelho, A. S. G., & Diniz-Filho, J. A. F. (2001). Is the relationship between population density and body size consistent across independent studies? A meta-analytical approach. *Revista Brasileira de Biologia*, 61(1).
- Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods*, 14(3), 225–238.
- Borenstein, M. (2009). Effect sizes for continuous data. In Harris M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221-235). New York, NY: Russell Sage Foundation.
- Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review*, 85(2), 238-243.
- Colegrave, A. D., & Giles, M. J. (2008). School cost functions: A meta-regression analysis. *Economics of Education Review*, 27(6), 688-696.
- Cooper, H. M. (2009). *Research synthesis and meta-analysis: A step-by-step approach*. SAGE.
- Cooper, H. M., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 3-16). New York, NY: Russell Sage Foundation.

- Cooper, H. M., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological methods*, 14(2), 165.
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.
- Corrao, G., Bagnardi, V., Zambon, A., & Arico, S. (1999). Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. *Addiction*, 94(10), 1551-1573.
- Crouch, G. I. (1995). A meta-analysis of tourism demand. *Annals of Tourism Research*, 22(1), 103–118.
- Crouch, G. I. (1996). Demand elasticities in international marketing:: A meta-analytical application to tourism. *Journal of Business Research*, 36(2), 117–136.
- Doucouliaos, C. (2005). Publication Bias in the Economic Freedom and Economic Growth Literature. *Journal of Economic Surveys*, 19(3), 367-387.
- Doucouliaos, H., & Laroche, P. (2009). Unions and profits: A meta-regression analysis. *Industrial Relations: A Journal of Economy and Society*, 48(1), 146-184.
- Fleiss, J., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In Harris M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237–253). New York: Russell Sage Foundation.
- Froot, K. A. (1989). Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis*, 24(03), 333-355.
- Freedman, D. A. (2006). On the so-called “Huber sandwich estimator” and “robust standard errors.” *The American Statistician*, 60(4), 299–302.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher*, 5(10), 3-8.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Greenland, S. (1987). Quantitative methods in the review of epidemiologic literature. *Epidemiologic reviews*, 9, 1-30.

- Greenland, S. (1993). A meta-analysis of coffee, myocardial infarction, and coronary death. *Epidemiology*, 4(4), 366-374.
- Greenland, S., & Longnecker, M. P. (1992). Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American Journal of Epidemiology*, 135(11), 1301 -1309.
- Greenland, S., Schlesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology*, 123(2), 203-208.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2), 135–143.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45 -62.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-28.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9(4), 426–445.
- Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Money does matter somewhere: A reply to Hanushek. *Educational Researcher*, 23(4), 9 -10.
- 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.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: correcting error and bias in research findings*. SAGE.
- Ishak, K. J., Platt, R. W., Joseph, L., & Hanley, J. A. (2008). Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Statistics in medicine*, 27(5), 670–686.
- Jarrell, S. B., & Stanley, T. D. (1990). A meta-analysis of the union-nonunion wage gap. *Industrial and Labor Relations Review*, 44(1), 54-67.
- Jarrell, S. B., & Stanley, T. D. (2004). Declining bias and gender wage discrimination? A meta-regression analysis. *Journal of Human Resources*, 39(3), 828-838.

- Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 222–230.
- Kim, R. S. (2011). *Standardized Regression Coefficients as Indices of Effect Sizes in Meta-Analysis* (Ph.D.). The Florida State University, United States -- Florida.
- Kock, A., & Gemünden, H. G. (2009). *A guideline to meta-analysis*.
- Link, C. R., & Mulligan, J. G. (1986). The merits of a longer school day. *Economics of Education Review*, 5(4), 373–381.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Longnecker, M. P. (1994). Alcoholic beverage consumption in relation to risk of breast cancer: Meta-analysis and review. *Cancer Causes & Control*, 5(1), 73-82.
- Lux, S., Crook, T. R., & Woehr, D. J. (2011). Mixing business with politics: A meta-analysis of the antecedents and outcomes of corporate political activity. *Journal of Management*, 37(1), 223.
- Maynard, R., & Crawford, D. (1976). School performance. *Rural income maintenance experiment: Final report*.
- Orsini, N., Bellocco, R., & Greenland, S. (2006). Generalized least squares for trend estimation of summarized dose-response data. *Stata Journal*, 6(1), 40.
- Paul, P. A., Lipps, P. E., & Madden, L. V. (2006). Meta-analysis of regression coefficients for the relationship between fusarium head blight and deoxynivalenol content of wheat. *Phytopathology*, 96(9), 951-961.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65(4), 373.
- Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *The Journal of Applied Psychology*, 90(1), 175-181.
- Raju, N. S., Pappas, S., & Williams, C. P. (1989). An empirical Monte Carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology*, 74(6), 901.

- Raju, Nambury S., Fralicx, R., & Steinhaus, S. D. (1986). Covariance and regression slope models for studying validity generalization. *Applied Psychological Measurement*, 10(2), 195 -211.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods*. SAGE.
- Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants. *Nature*, 421(6918), 57–60.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Thousand Oaks, CA: Sage.
- 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.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5 -11.
- Stanley, T. D., & Doucouliagos, H. (2010). Picture this: A simple graph that reveals much ado about research. *SSRN eLibrary*, 24(1), 170-191.
- Stanley, T. D., & Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2), 161-170.
- Stanley, T. D., & Jarrell, S. B. (1998). Gender wage discrimination bias? A meta-regression analysis. *The Journal of Human Resources*, 33(4), 947-973.
- Szymanski, D. M., Troy, L. C., & Bharadwaj, S. G. (1995). Order of entry and business performance: An empirical synthesis and reexamination. *The Journal of Marketing*, 59(4), 17-33.
- Tipton, L. E. (2011). *Methods for Generalizing from Experiments* (Ph.D., Statistics). Northwestern University, United States -- Illinois.
- Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys*, 19(3), 479-511.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838.

## VITA

Ryan Williams was born and raised in Arizona. In 2002 he graduated from Douglas High School in Douglas, Arizona. He completed his undergraduate studies at Indiana University in Bloomington, Indiana, majoring in psychology and sociology. He completed his studies in May, 2006.

Ryan enrolled at Loyola University Chicago in August 2006. Ryan's graduate studies were in part funded by a National Science Foundation grant. During his graduate career he presented papers at various conferences including the American Psychological Association, the American Educational Research Association, the Society for Research Synthesis Methodology and the Campbell Collaboration. From 2008-2011 he was the managing editor of the Campbell Collaboration methods group. He served as an ad-hoc reviewer for the Campbell Collaboration and Research Synthesis Methods from 2008-2012. He has conducted a wide range of applied research at Rehabilitation Institute of Chicago and at American Institutes for Research. He completed his master's degree in research methodology in 2009 and his PhD in research methodology in 2012.

Ryan is an Assistant Professor of quantitative methods in the College of Education, Health, and Human Sciences at The University of Memphis. He lives with his wife, Adrienne Lessard, in Memphis, Tennessee.



