

Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R

Emily E. Tanner-Smith¹  · Elizabeth Tipton² · Joshua R. Polanin³

Received: 14 August 2015 / Revised: 14 January 2016 / Accepted: 5 February 2016 /
Published online: 26 February 2016
© Springer International Publishing AG 2016

Abstract

Purpose Identifying and understanding causal risk factors for crime over the life-course is a key area of inquiry in developmental criminology. Prospective longitudinal studies provide valuable information about the relationships between risk factors and later criminal offending. Meta-analyses that synthesize findings from these studies can summarize the predictive strength of different risk factors for crime, and offer unique opportunities for examining the developmental variability of risk factors. Complex data structures are common in such meta-analyses, whereby primary studies provide multiple (dependent) effect sizes.

Methods This paper describes a recent innovative method for handling complex meta-analytic data structures arising due to dependent effect sizes: robust variance estimation (RVE). We first present a brief overview of the RVE method, describing the underlying models and estimation procedures and their applicability to meta-analyses of research in developmental criminology. We then present a tutorial on implementing these methods in the R statistical environment, using an example meta-analysis on risk factors for adolescent delinquency.

Results The tutorial demonstrates how to estimate mean effect sizes and meta-regression models using the RVE method in R, with particular emphasis on exploring developmental variation in risk factors for crime and delinquency. The tutorial also illustrates hypothesis testing for meta-regression coefficients, including tests for overall model fit and incremental hypothesis tests.

✉ Emily E. Tanner-Smith
e.tanner-smith@vanderbilt.edu

¹ Vanderbilt University, GPC, Box 0181, Nashville, TN 37203, USA

² Teachers College, Columbia University, 525 W 120th Street, Box 118, New York, NY 10027, USA

³ Development Services Group, Inc., 7315 Wisconsin Ave, Suite 800 East, Bethesda, MD 20814, USA

Conclusions The paper concludes by summarizing the benefits of using the RVE method with complex meta-analytic data structures, highlighting how this method can advance research syntheses in the field of developmental criminology.

Keywords Dependent effect sizes · Protective factors · Risk factors · Robust standard errors

Introduction

Identifying and understanding causal risk factors for crime over the life-course is a key area of inquiry in developmental and life-course criminology [1–4]. Prospective longitudinal studies provide valuable information regarding age-related trajectories of criminal offending, which is crucial not only for informing theories of crime etiology but also for developing interventions aimed at reducing criminal behavior (e.g., [5–7]). Findings from any individual study will rarely provide sufficient evidence for policy or practice recommendations. Even the most well-designed study will be subject to sampling error [8]. Meta-analyses that synthesize findings from multiple studies can provide more compelling evidence than that from a single study, given their greater statistical power to detect effects, generalizability across settings, and ability to systematically appraise conflicting findings across studies. Meta-analyses are therefore uniquely situated to guide the field of developmental and life-course criminology by assisting in the identification of risk factors that are promising targets for intervention and prevention, and identifying effective crime prevention programs. The Campbell Collaboration (www.campbellcollaboration.org), for instance, is an organization committed to producing and disseminating systematic reviews and meta-analyses to identify effective programs in the fields of criminology and other social sciences.

Complex data structures are common in meta-analyses of criminological research, whereby primary studies provide multiple, statistically dependent effect size estimates (i.e., effect size estimates with correlated error terms or correlated effect size parameters). For instance, the guiding example used in this tutorial is a meta-analysis where the effect size of interest is the correlation indexing the relationship between school motivation/attitudes and subsequent delinquency or criminal behavior. The prospective longitudinal studies that were included in this meta-analysis often reported multiple effect sizes (i.e., correlation coefficients) of interest, given the numerous ways in which school motivation/attitudes variables could be operationalized (e.g., academic aspirations, academic self-efficacy) as well as the numerous ways in which crime/delinquency could be operationalized (e.g., property crime, violent crime). In their meta-analysis of correlational research examining the association between school bullying and later aggression/violence, Ttofi et al. [9] also encountered dependent effect sizes, such that primary studies often reported multiple measures of bullying or aggression by informant (e.g., self- vs. peer-rated).

Complex data structures are also common in meta-analyses of intervention effectiveness research. For instance, in a Campbell Collaboration review and meta-analysis examining the effectiveness of drug courts, Mitchell and colleagues

[10] found that evaluation studies reported multiple odds ratio effect sizes indexing the effects of drug courts on multiple indicators of recidivism (e.g., referrals vs. convictions, general vs. specific offense types). In another meta-analysis examining the effectiveness of self-control interventions for reducing delinquency, Piquero, Jennings, and Farrington [2] reported that evaluations often included more than one intervention condition and/or comparison condition, thus yielding a set of statistically dependent effect size estimates. And in their meta-analysis synthesizing research on the effects of “pulling levers” focused deterrence strategies, Braga and Weisburd [11] also encountered dependent effect sizes such that evaluations often reported multiple indicators of crime (e.g., homicides, shots fired, gun assaults, youth gun assaults).

Statistically dependent effect sizes would ideally be synthesized via multivariate meta-analysis methods, which permit the inclusion of dependent effect sizes in a meta-analysis by fully modeling those dependencies [12–15]. Multivariate meta-analysis is rarely used, however, because it assumes knowledge of the underlying covariance structure among effect sizes (i.e., the covariance between effect size estimates within a study), which is seldom reported in primary studies and thus often unknown to the meta-analyst. Historically, most meta-analysts have consequently used data processing and selection techniques for handling dependent effect sizes. This might include, for instance, selecting one effect size per study, either randomly or based on a set of selection criteria, or creating a single synthetic effect size per study by averaging all effect sizes within a given study [14, 16–18]. Although these data processing and selection techniques result in an analytic sample of statistically independent effect sizes for any given analysis, these techniques ultimately result in the loss of potentially valuable information.

One contemporary alternative for handling statistically dependent effect sizes is through robust variance estimation (RVE). This method permits the inclusion of statistically dependent effect sizes within a meta-analysis and therefore permits the meta-analyst to utilize all available information reported in primary studies. This method does not require knowledge of the underlying covariance structure among effect sizes, and thus is practical for use in many scenarios. Recent advances in meta-analytic software have also made RVE easy to implement. Thus, this paper will provide an overview of the RVE method, describing the underlying model and estimation procedures and its applicability to meta-analyses of research in developmental criminology. We then present a tutorial on implementing these methods in the R statistical environment [19], using an example meta-analysis on school motivation risk factors for adolescent crime and delinquency. The goal of this paper is to provide practical guidance to researchers interested in applying this innovative method to complex meta-analytic data structures.

This is the first tutorial to demonstrate the application of RVE in the R statistical environment and the first to discuss the application of RVE for meta-analyses of criminological research. It is also the first to illustrate the use of RVE with small-sample adjustments for both the t - and F -tests, which we introduce below [20, 21]. Although more technical presentations of the RVE method exist (e.g. [22, 23]), as well as a tutorial for implementing RVE in SPSS and Stata [24], this is the first tutorial of which we are aware that provides a user-friendly description of how to implement RVE in the R environment.

Review of Standard Meta-analysis Model

In a standard meta-analysis with independent effect sizes, the fixed-effect model is written as:

$$y_j = \beta_0 + e_j \quad (1)$$

where y_j is an effect size in the j th study ($j=1 \dots m$); β_0 is the average population effect, and e_j is the study level residual $\sim N(0, v_j)$ [25]. Although this fixed-effect model is appropriate under some circumstances, most meta-analyses of criminological research will assume some underlying heterogeneity in effects (e.g., across settings, samples, outcomes) and/or will desire to make inferences to a larger population of studies, in which case a random-effects model will be more appropriate. In a standard meta-analysis with independent effect sizes, the random-effects model is written as:

$$y_j = \beta_0 + u_j + e_j \quad (2)$$

where y_j is an effect size in the j th study ($j=1 \dots m$); β_0 is the average population effect; u_j is the study level random effect $\sim N(0, \tau^2)$, and e_j is the study level residual $\sim N(0, v_j)$ [25]. Given that heterogeneous effect sizes are common in meta-analyses of criminological research, most meta-analysts are also interested in examining the effect of moderator variables that may account for some of the observed variability in effect sizes [26, 27]. This simple random-effects meta-analytic model can thus be extended to include p covariates or effect size moderators $x_1 \dots x_p$:

$$y_j = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + u_j + e_j \quad (3)$$

where β_0 is the intercept, $\beta_1 \dots \beta_p$ are the regression coefficients for the covariates, and $\text{Var}(u_j) = \tau^2_{\text{res}}$ is the residual heterogeneity after adjusting for the covariates. This meta-regression model can thus be used to model the associations between the p covariates or effect size moderators and the effect size y . The weighted least-squares estimate of $\beta = (\beta_1, \dots, \beta_p)$ can be calculated using

$$\mathbf{b} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{W}\mathbf{Y}) \quad (4)$$

where \mathbf{W} is a diagonal weight matrix. The variance of the estimate \mathbf{b} of β can be written as:

$$V(\mathbf{b}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{W}\Sigma\mathbf{W}\mathbf{X}) (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}. \quad (5)$$

In a traditional meta-analysis model where the effect sizes from each study are assumed to be statistically independent, the diagonal elements of Σ are the variances v_j and the off-diagonal covariance estimates are assumed to be zero.

Typically, the weights are defined to be inverse variance, with the diagonals of \mathbf{W} equal to $(1/v_j)$ for each effect size, and the off-diagonals equal to zero; in this case, the variance reduces to:

$$V^M(\mathbf{b}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}. \quad (6)$$

This formula for standard errors is widely used in meta-analyses of criminological research and is appropriate to use when estimated effect sizes are statistically independent. The estimator $V^M(\mathbf{b})$ is inappropriate for use when effect sizes are statistically dependent, however, because it results in standard error estimates that are much too small.

Review of Robust Variance Estimation (RVE)

RVE is a method that can be used to analyze statistically dependent effect sizes in a meta-analysis [22, 23]. Although this method does not require researchers to correctly specify the dependence structure of the data, it is helpful for researchers to understand where and how dependence can arise. For simplicity, we focus here on two broad types of dependence structures: “hierarchical effects” and “correlated effects.” Although we describe these two types of dependence separately, it is common for *both* forms of dependence to arise within the same meta-analysis (see [24]).

In the hierarchical effects case, dependence arises due to multiple studies being nested within a larger cluster such as a research group, lab, region, or country. For instance, a meta-analysis on drug court effectiveness studies might synthesize effect sizes indexing reductions in recidivism among drug court participants across multiple studies, but some drug courts may have conducted multiple independent evaluations at the same drug court site (e.g., by cohort). Thus, in this scenario, the meta-analyst might have effect sizes from multiple independent studies nested within a larger cluster (i.e., drug court site). In this hierarchical effects case, primary study participants (level 1) provide effect sizes in multiple studies (level 2), which are nested within some larger cluster (level 3). This hierarchical effects model assumes independent sampling errors within clusters, which is often a reasonable assumption when there are no overlapping participant samples represented in the effect size estimates.

Perhaps a more common dependence structure that arises in meta-analyses of criminological research is the correlated effects case, where there are dependent sampling errors within clusters. Correlated effects arise when the same participant samples are used to estimate multiple effect sizes. For instance, a meta-analysis on childhood risk factors for juvenile delinquency might synthesize effect sizes indexing the strength of different parenting-related risk factors across multiple studies; individual studies may report multiple measures of parenting-related risk factors, and individual studies may also report longitudinal effects at multiple follow-up points. In this correlated effects case, the meta-analytic model now represents primary study participants (level 1) providing multiple effect estimates (level 2) that are nested within studies (level 3).

The RVE method is agnostic to the type of dependence structure present in the data, so can be used to handle both the hierarchical effects and correlated effects scenarios

described above. Moreover, the same underlying model is used here as in Eq. 2. Assuming a correlated effects model where multiple effect size estimates are nested within studies, the meta-analytic model can now be written as:

$$y_{ij} = \beta_0 + u_j + e_{ij} \tag{7}$$

where for $i = 1 \dots k_j, j = 1 \dots m, y_{ij}$ is the i th effect size in the j th study; β_0 is the average population effect; u_j is the study level random effect such that $\text{Var}(u_j) = \tau^2$ is the between-study variance component; and e_{ij} is the residual for the i th effect size in the j th study. To examine effect size moderators, the RVE model can be extended to include p covariates $x_1 \dots x_p$, where:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \dots + \beta_p x_{pij} + u_j + e_{ij}. \tag{8}$$

With RVE models, the weighted least-squares estimate of $\beta = (\beta_1, \dots, \beta_p)$ can be calculated using the same approach as in the standard meta-analysis model (Eq. 4). The RVE approach differs from the standard meta-analysis approach in the estimate of the variance of \mathbf{b} (Eq. 5). Whereas in standard meta-analysis the model-based estimator (Eq. 6) is employed, in RVE instead this variance is estimated by:

$$V^R(\mathbf{b}) = \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{A}_j \mathbf{e}_j \mathbf{e}'_j \mathbf{A}'_j \mathbf{W}_j \mathbf{X}_j \right) \left(\sum_{j=1}^m \mathbf{X}'_j \mathbf{W}_j \mathbf{X}_j \right)^{-1} \tag{9}$$

where for study $j = 1 \dots m, \mathbf{X}_j$ is the design matrix, \mathbf{W}_j is the weight matrix, \mathbf{A}_j is an adjustment matrix, and \mathbf{e}_j is the estimated residual vector. This adjustment matrix is included to adjust for small-sample bias in the estimator $V^R(\mathbf{b})$. When the adjustment matrix follows the form provided in Tipton [20], the robust variance estimator $V^R(\mathbf{b})$ is unbiased for the true variance $V(\mathbf{b})$ even with a small number of included studies.

The Role of Weights in RVE

Although any weights can be used with the RVE method, Hedges et al. [22] argue that approximately inverse variance weights will often be desired because they are most efficient. Because the RVE estimator is derived for any weights, using weights other than inverse variance weights will not comprise statistical inferences and will only comprise efficiency. Hedges and colleagues [22] therefore proposed two methods for weighting: hierarchical effects weights and correlated effects weights (and these are the weights that are available as default options in the R package used in this tutorial).

In a hierarchical effects model, inverse variance weights can be used:

$$\omega_{ij} = \frac{1}{(v_{ij} + \tau^2 + \omega^2)} \tag{10}$$

where v_{ij} is the within-study sampling variance for effect size i in study j, τ^2 is the estimate of the between-cluster variance component, and ω^2 is the estimate of the

within-cluster between-study variance component. The matrix \mathbf{W}_j thus has the values w_{ij} on the diagonals and zeros on the off-diagonals. Hedges et al. [22] provide method of moments estimators for both τ^2 and ω^2 . Importantly, unlike maximum likelihood estimates, these estimates do not require any distributional assumptions.

In a correlated effects model, Hedges et al. also provide a simple method for estimating inverse variance weights:

$$w_{ij} = \frac{1}{k_j(v_{.j} + \tau^2)} \tag{11}$$

where $v_{.j}$ is the mean of the within-study sampling variances (v_{ij}) for the k_j effect sizes in study j , τ^2 is the estimate of the between-study variance component, and k_j is the number of effect sizes within each study j . Again, Hedges et al. provide a method of moments estimator of τ^2 , which is implemented in the R package used in this tutorial.

Because the choice of weights only affects efficiency in estimates, Tanner-Smith and Tipton [24] recommended that meta-analysts faced with complicated dependence structures (i.e., a mixture of correlated and hierarchical effects) choose weights based on the most prevalent type of dependence. Namely, if most studies provide results from multiple evaluations on different participant samples, but a few also measure multiple outcomes on the same participant samples, then the hierarchical effects weights would be most appropriate. In contrast, if most studies provide the results of a single evaluation with multiple measures on the same participant sample, but a few have multiple cohorts of samples, then the correlated effects weights would be more appropriate.

Adjustment Matrices

In the RVE estimator of the variance (Eq. 9), adjustment matrices \mathbf{A}_j are found on either side of the estimator. In the original formulation of RVE found in Hedges et al. [22], these adjustments were not included, and the estimator was known to under-estimate the true variance when the number of studies is small (see simulations in [22, 23]). Tipton [20] showed that the bias from inclusion of a small number of studies could be reduced if these adjustment matrices were included. In the case in which correlated effects weights are used, these matrices can be written,

$$\mathbf{A}_j = (\mathbf{I} - \mathbf{H}_j)^{-1/2} \tag{12}$$

where $\mathbf{H}_j = \mathbf{X}_j(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'_j\mathbf{W}_j$ is the hat matrix for study j . These adjustments differ from study to study and take into account the leverage and influence of each study in the analysis. The adjustments for the hierarchical effects case, as well as more general adjustments, are found in Tipton [20] and follow a similar form.

Hypothesis Tests for RVE Meta-regression Coefficients

In meta-regression, analysts are typically interested in testing whether effect sizes vary in relation to particular covariates or effect size moderators. Hypotheses regarding a

single coefficient β_k in a meta-regression model (e.g., $H_0: \beta_k=0$) can be tested using the robust variance estimator as:

$$t_{k=} = \frac{\beta_k}{\sqrt{V_k^R}} \quad (13)$$

where b_k is the estimate of β_k and V_k^R is robust estimate of the variance estimate of b_k (i.e., the k th diagonal of the $V^R(\mathbf{b})$ variance-covariance matrix). In large samples, under the null hypothesis this t_k statistic can be shown to be normally distributed (see [22]). In small to moderate samples, Tipton [20] shows that the distribution of t_k is approximately that of a t -distribution, where the degrees of freedom are estimated using a Satterthwaite approximation; this approximation holds as long as the degrees of freedom are greater than or equal to four. When the degrees of freedom are less than four, the approximation does not hold and the type I error can be larger than the stated p value associated with the test.

Similarly, researchers are often interested in multi-parameter tests, including omnibus tests of model fit (e.g., $H_0: \boldsymbol{\beta}=\mathbf{0}$) and incremental tests (e.g., $H_0: \beta_1=\beta_2=0$). Incremental tests are particularly useful in the presence of categorical effect size moderators, whereby multiple dummy or indicator variables might be used to examine the effects of a categorical factor in a meta-regression model (e.g., region; age groups; race composition). In this case, Tipton and Pustejovsky [21] showed that the F test,

$$F_q = \left(\frac{\eta - q + 1}{\eta q} \right) (\mathbf{b}_q)' \left[\mathbf{V}_q^R \right]^{-1} (\mathbf{b}_q) \quad (14)$$

can be used to test the null hypothesis that $\boldsymbol{\beta}_q=\mathbf{0}$, where \mathbf{b}_q is a vector of estimates for the q coefficients being tested, $\mathbf{V}_q^R = V^R(\mathbf{b}_q)$ is the sub-matrix of $V^R(\mathbf{b})$ associated with the q estimates in \mathbf{b}_q , and η is a small-sample correction factor that is empirically estimated. Tipton and Pustejovsky also showed that the distribution of F_q can be approximated by an $F(q, \eta - q + 1)$ distribution, where q is the number of parameters or contrasts tested and η is empirically estimated. Like the robust t test described above, this approximation holds even when the number of studies is small (and, unlike the t test, for all values of the degrees of freedom).

Importantly, for both the robust t test and F test, the degrees of freedom depend on both the number of studies in the meta-analysis (but not the number of effect sizes), and also on features of the covariate(s). For example, for the robust t test, when a covariate is balanced (e.g., a dummy variable for sex with 50 % females and 50 % males), the degrees of freedom are close to $m - p$ (where m is the number of studies and p is the number of coefficients in the model), but grow smaller as the degree of balance in the covariate values decreases. Similarly, the degrees of freedom decrease when a covariate has large skew or an extreme leverage point. This relationship between covariate features and degrees of freedom holds for the F test as well, although the multivariate nature of the F test makes it harder to diagnose. Nevertheless, the most important practical issue for meta-analysts using this method is that because the degrees of freedom vary in relation to covariate features (and not just the number of studies), these small-sample corrected hypothesis tests should always be implemented when

using RVE (and indeed, these are now the default test statistics in the RVE software used in this tutorial).

Comparisons Between RVE and Multilevel Meta-analysis

Multilevel meta-analysis (MLMA) is another method that can be used to include statistically dependent effect sizes in a meta-analysis [28–32]. The MLMA approach decomposes the variance of the effect sizes at each level, and thus explicitly models the dependence structure among effect sizes. In contrast to RVE, MLMA is not agnostic to the dependence structure in the data and is only appropriate in the hierarchical effect case, where primary study participants (level 1) provide one effect size in multiple studies (level 2), which are nested within some larger cluster (level 3). This hierarchical effects model assumes independent sampling errors within clusters, which is a reasonable assumption when there are no overlapping participant samples represented in the effect size estimates. For instance, in a meta-analysis of drug court impact evaluations, an evaluation study might report findings from several consecutive cohorts of drug court participants. In this example, the MLMA approach would be appropriate for handling the dependent effect sizes in this hierarchical structure, and would also be useful for decomposing variance of effect sizes at each level (e.g., between drug court sites, between cohorts within drug court sites). Parameters from these hierarchical effects three-level MLMA models can be estimated using iterative Maximum Likelihood or Restricted Maximum Likelihood algorithms [30, 33].

Because MLMA models assume independent sampling errors within clusters, using MLMA under the correlated effects scenario (i.e., when the same participants provide data on multiple effect sizes) violates the model assumptions of independent sampling errors within level 3 units. Because most developmental criminological meta-analyses with complex data structures will include correlated effects (rather than hierarchical effects), this tutorial will focus solely on the application of the RVE method for correlated effects data (but see [32, 34]).

Benefits and Limitations of RVE

The RVE method permits the inclusion of statistically dependent effect sizes within a meta-analysis, and therefore does not require the analyst to throw away information contained in the effect sizes reported in primary studies (i.e., the type of data loss that would occur if selection criteria were used to select a set of statistically independent effect sizes). Indeed, RVE does not require knowledge of the covariance between the effect sizes within each study, information that is required when applying multivariate meta-analysis methods [35]. Because RVE can be easily implemented using free statistical software (demonstrated later in this tutorial), it is an innovative and attractive method for handling multiple types of complex meta-analytic data structures.

Although RVE is a promising method for handling statistically dependent effect sizes within meta-analyses of criminological research, it also has several limitations that should be noted. The first of these has to do with the treatment of heterogeneity in RVE. Whereas MLMA models heterogeneity at multiple levels of the data, and therefore provides estimates of variance parameters (e.g., τ^2_{L2} and τ^2_{L3}) and corresponding hypothesis tests (i.e., Q , I^2), RVE is simply a method for adjusting standard errors

where the primary focus is on estimating fixed effects (e.g., the mean effect size, meta-regression coefficients). Heterogeneity parameters, although reported in RVE, are incidental to the RVE analysis; they are only needed for the estimation of approximately inverse variance weights, and are often estimated via a simplistic method-of-moments estimator. Thus, RVE models are not intended to provide precise variance parameter estimates, nor test null hypotheses regarding heterogeneity parameters. This is an important issue to highlight; many meta-analysts are interested in estimating precise variance parameter estimates and want to conduct hypothesis tests around these variance parameters, but RVE may not be the most appropriate method for that purpose.

Second, the fact that the degrees of freedom used in hypothesis testing depend on the number of studies and features of the covariate means that in some cases, tests of particular moderators may be surprisingly under-powered. This problem arises when the covariate under study is highly skewed (e.g., most of the values are between 10 and 20, with one value of 100) or imbalanced (e.g., 27 studies are on adolescents and 3 are on adults, with a dummy code for age as a moderator). In these cases, the number of studies could be quite large (e.g., more than 40) and yet the degrees of freedom could be quite small (e.g., less than 10).

Aside from issues of power, this can be particularly problematic when the degrees of freedom fall below 4 for t tests, where the t -distribution approximation no longer holds. In these cases, two approaches are recommended. First, if the degrees of freedom are much smaller than the number of studies, the analyst should carefully examine the covariate values, paying attention to leverage points and imbalances. The suggested approach here is to follow the rules of thumb and guidelines for dealing with “unusual” data (e.g., outliers, leverage, influence points) in regression. For example, if most of the covariate values are between 10 and 20, with one “outlying” value of 100, a strategy may be to remove or winsorize this extreme value; doing so will improve not only the degrees of freedom, but may also help with the interpretability of the findings, since this observation could be exerting large influence on the coefficient estimate as well. Second, if the degrees of freedom are very small, a lower p value should be used; for example, if $p < 0.05$ is used as a threshold elsewhere, for these cases $p < 0.01$ should be used instead (since the type I error is typically higher than stated in these cases).

Finally, RVE provides a statistically appropriate method for the inclusion of all effect sizes in a meta-analysis, but does not address other obstacles commonly encountered in a meta-analysis. For instance, many primary studies fail to report the information needed to estimate a standardized effect size, which often necessitates requesting additional information from primary study authors (and response rates to these requests can be quite low). Such missing effect size data pose a challenge to any meta-analyst, and the use of RVE may not be warranted if standardized effect sizes are not available from a sufficient number of studies. Similarly, the RVE method does not directly address another common obstacle in meta-analyses: how to handle the synthesis of bivariate and partial correlations. Partial correlations are often available from regression coefficients reported in primary studies, and meta-analysts synthesizing correlations must carefully attend to this issue [36]. The RVE method also does not provide a solution to the “apples

and oranges” critique common in meta-analysis. Just as in any other meta-analyses, analysts using RVE should be careful to consider whether the effect sizes included in a meta-analysis are “similar” enough to be synthesized. This is a substantive consideration that must always be guided by content expertise, however, so is not a true limitation of the RVE method itself but rather a potential limitation if meta-analysts carelessly apply the method to their data.

Example Research Questions to Guide the Software Tutorial

To illustrate the application of RVE for handling statistically dependent effect sizes in a correlated effects scenario, this tutorial will be guided by the following research questions: (1) How strongly are school motivation risk factors in childhood and adolescence correlated with subsequent crime and delinquency in adolescence, and (2) How does the crime risk associated with school motivation/attitudes vary across the span of adolescence? This example was selected given its relevance to developmental and life-course criminology and its demonstration of the potential developmental variability in risk factors for crime during adolescence. This example is used solely for pedagogical and demonstration purposes, however, and is not intended to provide a review or synthesis of the most current research evidence. And although these example data use correlation coefficients, the same procedures outlined in this tutorial can be applied to intervention effectiveness meta-analyses that may synthesize standardized mean difference effect sizes, risk ratios, or odds ratios.

Method

Sample

This tutorial uses a meta-analytic database that includes correlations indexing the relationship between school motivation or attitudes (hereafter referred to as school motivation) and a later chargeable delinquent or criminal behavior (hereafter referred to as crime) measured during adolescence and adulthood. This database is a subset from a larger meta-analytic database of studies published through 2002 that used prospective longitudinal panel designs with at least 6 months between measurement waves, which reported the correlation between a risk factor and a subsequent antisocial behavior outcome measured during late adolescence or early adulthood (funded by R01MH051685 and R01MH63288 from the National Institutes of Mental Health to PI Mark Lipsey). To be eligible for inclusion in the larger meta-analytic database, studies could have no more than 50 % attrition between consecutive waves, and must have been conducted in the USA and published in English (see [37]). The subset of studies used in this example includes 113 product-moment correlation coefficients, each of which represent a relationship between a school motivation risk factor measured between ages 11 and 18 and a crime outcome measured between ages 12 and 20. These correlations originated from 17 independent respondent samples from 6 longitudinal studies (see [58-63]). Note that this sample of correlation effect sizes is used solely for

pedagogical purposes for this tutorial, however, and is not intended to provide a current synthesis of the literature (for which an updated, systematic literature review would be needed).

Measures

The effect sizes used in this tutorial are Fisher's z -transformed correlation coefficients representing the longitudinal relationship between a school motivation risk factor measured prior to age 18 and a crime outcome measured at least six months after the initial wave of measurement:

$$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right) \quad (15)$$

where r is the product-moment correlation coefficient between school motivation and crime. The variance of the z -transformed correlation effect size is estimated as:

$$V = \frac{1}{n-3} . \quad (16)$$

Correlation coefficients were coded directly from study reports or estimated from contingency tables or other summary statistics if necessary. As is recommended when estimating correlation effect sizes from primary studies, artifact adjustments were used to correct for the unreliability of any artificial dichotomous variables used in the correlations for underlying continuous variables [26, 38]. We refer readers interested in guidance on how to calculate effect size estimates from primary studies to more comprehensive sources on that topic [25, 26, 38, 39].

Analytic Strategies

On average, studies reported 6.65 correlation effect sizes (SD=3.81, Minimum=2, Maximum=12) measuring the relationship between a school motivation risk factor and a subsequent crime outcome. We therefore used RVE to synthesize these dependent effect size estimates. First, to estimate the overall strength of the correlation between school motivation risk factors and later crime outcomes, we estimated a simple RVE meta-regression model:

$$y_{ij} = \beta_0 + u_j + e_{ij} \quad (17)$$

where y_{ij} is the i th correlation effect size in the j th study; β_0 is the average population effect of the risk-crime correlation; u_j is the study level random effect such that $\text{Var}(u_j) = \tau^2$ is the between-study variance component; and e_{ij} is the residual for the i th effect size in the j th study. Next, to explore variability in the strength of this risk factor over the span of adolescence, we estimated a mixed-effects RVE meta-regression model:

$$y_{ij} = \beta_0 + \beta_1(\text{Age at T}_1)_{ij} + \beta_2(\text{Age at T}_2)_{ij} + \beta_3(\text{Age at T}_1 \times \text{Age at T}_2)_{ij} + u_j + e_{ij} \quad (18)$$

that included the main effects and multiplicative interaction term for the average age of participants at the initial measurement wave (i.e., the age at which the school motivation risk factor was measured) and the age at the follow-up measurement wave (i.e., the age at which the crime outcome was measured). Because the age of participants could vary both within and between studies, we used study-level mean values for all age moderators (see [24] for a discussion on measuring moderators in RVE models). Finally, to examine whether these effects persisted after adjusting for the sex composition of the samples, we estimated a final model:

$$y_{ij} = \beta_0 + \beta_1(\text{Age at } T_1)_{ji} + \beta_2(\text{Age at } T_2)_{ij} + \beta_3(\text{Age at } T_1 \times \text{Age at } T_2)_{ij} \\ + \beta_4(\text{All males}) + \beta_5(\text{Mixed sex}) + u_j + e_{ij} \quad (19)$$

that included two dummy variables indicating whether the participant sample was comprised entirely of males or was of mixed sex composition (with all females being the reference category). Note that this categorization of the gender composition variable is used solely for pedagogical purposes to demonstrate the examination of a categorical effect size moderator. This illustrative example was selected to highlight how RVE can be used to examine developmental changes in risk factors over time, but this meta-regression could easily be extended to include a range of other effect size moderators (provided sufficient sample sizes for parameter estimation). Because the dependence structure in these data reflected correlated effects (i.e., multiple effect sizes available from the same participant samples), we used the recommended approximately inverse variance weights for correlated effects (Eq. 11).

All analyses were conducted in the R statistical environment (version 3.2.2). To illustrate the standard meta-analysis method that (naively) ignores the dependencies in the data, we used the *metafor* package [40]. For the RVE method, the *robumeta* package [41] was used to estimate mean effect sizes and meta-regression models and the *clubSandwich* package [42] was used to estimate the multi-parameter F tests. The package *ggplot2* was used to create the scatter plot [43]. Appendix A provides annotated code using the R Markdown language [44].

Robust Variance Estimation in R

We now provide a tutorial on how to conduct a meta-analysis with RVE in the R statistical environment. Although these methods can also be implemented in SAS and Stata (e.g., see [24, 34]), here we use R because it is freely available from the Comprehensive R Archive Network at <http://cran.r-project.org>. Appendix A includes the R code that reproduces all analyses demonstrated below; Appendix B (in the only supplementary material) Table 3 (and the online supplementary material) provides a text file containing the data used in all analyses.

Prior to synthesizing effect sizes, we examined the distribution of the effect sizes to assess for potential outliers using a violin plot (not shown, but syntax included in Appendix A). The effect size distribution is approximately normal and right skewed, but there do not appear to be any outlying effect sizes.

We now estimate the overall mean effect size, synthesizing the 113 correlations from the 17 independent study samples. For illustrative purposes, we first estimate a naïve standard meta-analysis model that ignores the dependency structure of the data. This naïve meta-analysis model would never be appropriate to use with such a complex meta-analytic data structure; here we present it simply to illustrate how ignoring dependencies can yield inappropriately small standard errors. As shown in Table 1, the naïve standard meta-analysis ($\bar{z}=0.12$, 95 % CI [0.09, 0.14], $\tau^2=0.01$) yields a larger estimate of the average effect size and a narrower confidence interval than the analysis appropriately handling the statistically dependent effect sizes using RVE ($\bar{z}=0.09$, 95 % CI [0.04, 0.15], $\tau^2=0.01$). Thus, these results suggest that school motivation has a small but significant association with crime during adolescence. Transforming this mean Fisher's z -transformed effect size into a Pearson correlation $(e^{2*0.09} - 1) / (e^{2*0.09} + 1)$ yields a mean correlation of 0.09. Although informative for estimating the overall correlation between school motivation and adolescent criminal behavior, this overall mean correlation estimate ignores any potential developmental variability in this risk-crime relationship. Thus, we next estimate a meta-regression model with RVE that models the potential variability in this risk-crime correlation across the span of adolescence.

Table 2 presents the results from the RVE meta-regressions that model the potential developmental variability in the risk of adolescent crime associated with school motivation. The left panel presents results from a model that estimates the main effects of participant ages at time 1 (when school motivation risk was measured) and time 2 (when the crime outcome was measured). Overall, the results provide no evidence that the risk-crime correlation varies depending on the average ages of participants at the different measurements waves. The omnibus F test [21] provides no evidence that the age covariates are jointly significant ($F=1.54$, $df=3.11$, $p=0.34$). This model does not appropriately parameterize the potential developmental variability in risk, however, because it does not include the multiplicative interaction term between the two age variables. Namely, the strength of school motivation as a risk factor at age

Table 1 Mean effect size estimates from standard meta-analysis and RVE meta-analysis

	Standard (Naïve) meta-analysis	RVE meta-analysis
\bar{z}	0.12*	0.09*
SE \bar{z}	0.01	0.03
95 % CI	[0.09, 0.14]	[0.04, 0.15]
τ^2	0.01	0.01
n	113	113
k	17	17
ρ	n/a	0.80

\bar{z} = mean Fisher's z effect size, SE = standard error of mean effect size, τ^2 = between-study variance, n = number of effect sizes, k = number of studies, ρ = assumed correlation between effect sizes

* $p < 0.05$

Table 2 Results from RVE meta-regression modeling developmental variability in risk-crime correlation ($n = 113, k = 17$)

	Age main effects model			Age interaction effects model			Age interaction effects, adjusting for sex		
	<i>b</i>	95 % CI	<i>df</i>	<i>b</i>	95 % CI	<i>df</i>	<i>b</i>	95 % CI	<i>df</i>
Intercept	0.40	* [0.01, 0.80]	6.68	-2.44	[-5.00, 0.12]	5.95	-2.56	[-5.48, 0.36]	5.94
Age _{T1}	0.04	[-0.06, 0.014]	6.10	0.26	* [0.04, 0.48]	7.05	0.28	* [0.05, 0.51]	6.55
Age _{T2}	-0.05	[-0.20, 0.09]	1.92	0.11	[-0.05, 0.28]	6.72	0.12	[-0.07, 0.31]	6.65
Age _{T1} × Age _{T2}	-	-	-	-0.01	* [-0.02, -0.00]	7.15	-0.01	* [-0.03, -0.01]	6.81
All male	-	-	-	-	-	-	.03	[-0.12, 0.18]	5.37
Mixed sex	-	-	-	-	-	-	.01	[-0.11, 0.13]	7.22
Residual r^2	-	0.01	-	-	0.01	-	-	0.01	-
Omnibus <i>F</i>	-	1.54	2, 3.11	-	1.83	3, 3.03	-	1.05	5, 3.08

n = number of effect sizes, k = number of studies, ρ assumed = 0.80

* $p < 0.05$

12 may vary for crime at age 14, 16, 18, etc.; thus, an interaction term is needed to model that potential variability in risk. Furthermore, the small-sample corrected degrees of freedom for the age at time 2 parameter in the main effects model is less than four ($df=1.92$) and thus the p value associated with the hypothesis test is likely to under-estimate the true type I error; because the test is not significant, however, taking this into account does not change the conclusion [20]. Thus, we next estimated a full model that included both the main effects and interaction terms to examine the potential developmental variability in crime risk associated with school motivation.

The middle panel of Table 2 presents results from this full model. In this model, all of the small-sample corrected degrees of freedom are in an acceptable range (i.e., greater than four), indicating that the p values for the associated t tests accurately reflect the type I error. The results indicate that there is a significant interaction between age at time 1 and age at time 2 ($b=-0.01$, 95 % CI $[-0.02, -0.00]$). Although the omnibus F test provides no evidence of joint significance across all covariates ($F=1.83$, $df=3.03$, $p=0.32$), this is likely due to the small number of studies and the imbalance in covariate values. Plotting the results from this interaction in Fig. 1, we see that the magnitude of the correlation between school motivation and criminal behavior generally decreases over the course of adolescence (evidenced by the negative slopes of the regression lines). However, school motivation around age 17 has the strongest correlation with crime during late adolescence and early adulthood (ages 18–20). Although the magnitude of these risk-crime correlations is somewhat small in magnitude, these results nonetheless suggest that school motivation may be an appropriate target for delinquency prevention and intervention programs, and also suggest that the most developmentally appropriate time to target school motivation risk factors may be prior to age 16.

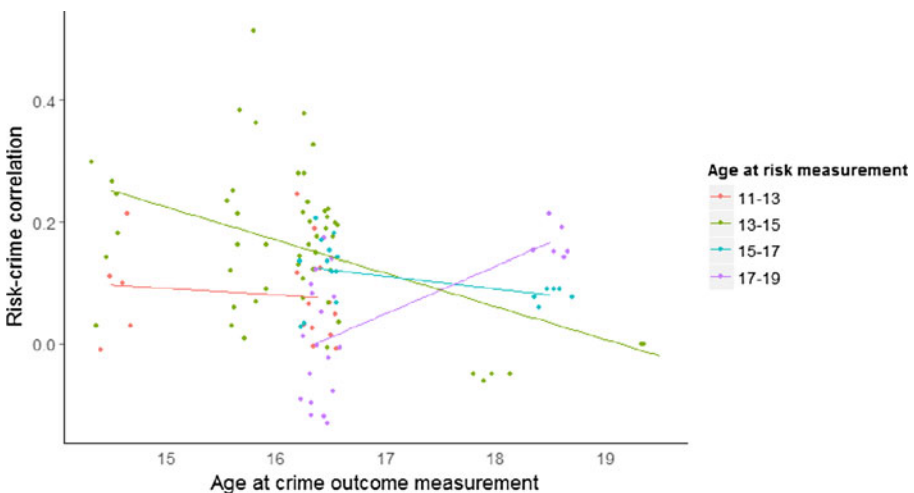


Fig. 1 Developmental variation in the correlation between school motivation risk and subsequent criminal behavior

The right panel of Table 2 includes two additional dummy variables representing the sex composition of the samples in the meta-analysis, contrasting samples that were all male, mixed sex, and all female (the reference category). The results from this model suggest that the developmental variability in the crime risk associated with school motivation is still statistically significant ($b = -0.01$, 95 % CI $[-0.03, -0.01]$) even after adjusting for the sex composition of the primary study samples. Neither of the regression coefficients for the sex composition dummy variables are statistically significant, so it is no surprise that the joint F test regarding the effect of gender composition (i.e., the two dummy variables) is non-significant as well ($F = 0.14$, $df = 6.51$, $p = 0.87$). And similar to the prior model, the omnibus F test provides no evidence of joint significance across all covariates ($F = 1.05$, $df = 3.08$, $p = 0.52$).

Discussion

Meta-analyses in the field of developmental and life-course criminology will often encounter complex meta-analytic data structures with statistically dependent effect size estimates. Although statistically dependent effect sizes would ideally be synthesized via multivariate meta-analysis, that method is often not feasible to implement given that it requires knowledge about the covariance structure of the effect size estimates. Historically, therefore, meta-analysts have been forced to use less optimal methods for handling statistically dependent effect sizes, such as selecting one effect size per study or creating a single synthetic effect size per study, thus ultimately throwing away potentially valuable information from primary studies (e.g., [9–11, 45]).

RVE for meta-regression is a recent innovative technique that can be used to handle statistically dependent effect sizes in a meta-analysis, but does not require information about the effect size covariance structure. Although the RVE method has several limitations (e.g., it does not model variance at all levels of the data, does not provide tests of null hypotheses regarding heterogeneity parameters; see [Benefits and Limitations of RVE](#)), the method is easy to implement and can also be used to handle different types of dependency structures. This software tutorial demonstrated how to use RVE in the R statistical environment, providing example code for future researchers to adapt for their own meta-analyses. Using an example meta-analysis examining the crime risk associated with adolescents' school motivation and attitudes, we illustrated how RVE can be used to address important research synthesis questions regarding the developmental variability in risk factors. Syntheses like these can be useful for identifying risk and protective factors that may be appropriate targets for crime prevention and intervention programs, establishing the most developmentally appropriate time to address these risk factors, and advancing developmental theories of crime etiology.

Although this tutorial illustrated the use of RVE in a meta-analysis with a correlated effects dependency structure, this method is also appropriate for meta-analyses with hierarchical effects dependency structures or a combination of dependence structures (see [24] for a tutorial). Further, although the data used in this tutorial were from a meta-analysis where the effect size metric was

a correlation coefficient, the same procedures can be used in meta-analyses using other effect size metrics such as mean differences, standardized mean differences, odds ratios, risk ratios, risk differences, or phi coefficients. Indeed, the example code in this software tutorial can be easily translated to meta-analyses of program or policy effectiveness, such as those examining the effects of neighborhood watch on crime [46], street-level drug law enforcement [47], or batterer intervention programs [48].

This tutorial demonstrated the application of RVE in the R statistical environment (version 3.2.1) using the *robumeta* [41] and *clubSandwich* [42] packages. The tutorial illustrates the ease with which the RVE method can be implemented in R, with the additional advantage that R is freely available to all researchers. Meta-analysts interested in implementing RVE with other statistical platforms such as SAS, SPSS, or Stata should consult other resources (e.g., [24, 34]). Importantly, analysts should pay attention to the implementation of RVE in these packages, noting if corrections for small numbers of studies have been implemented before proceeding. Currently, the Stata macro (also called “robumeta”) includes small sample adjustments for t tests but not F tests, whereas the SPSS and SAS macros do not include these adjustments.

Complex meta-analytic data structures with statistically dependent effect size will become more common in criminology as the body of primary studies continues to accumulate, and as the focus of meta-analyses shifts away from estimation of mean effect sizes to the examination of variability in effect sizes. Indeed, the RVE method for handling dependent effect sizes is already being adopted by researchers in the fields of ecology [49, 50], education [51], mental health [52], pediatrics [53, 54], psychology [55], psychiatry [56], and substance abuse [57]. The uptake in RVE is not surprising given that this method obviates the loss of information that occurs when meta-analysts artificially create a set of statistically independent effect sizes, either through selection processes or averaging effect sizes within studies.

This tutorial is intended to provide a guide for criminologists interested in applying the RVE method to their complex meta-analytic data structures. This method will be particularly useful for meta-analysts interested in modeling the developmental graduations in risk factors for crime (e.g., what are the age-crime curves for different risk factors), examining the effectiveness of crime prevention programs during key life-course transitions (e.g., are residential mobility programs effective in reducing juvenile delinquency?), and other developmental dimensions of crime across the life-course.

Compliance with Ethical Standards This manuscript does not contain clinical studies or patient data.

Appendix A

R Code for Tutorial Example

Start by installing and loading all required libraries. The code will run using R version 3.2.2 "Fire Safety". Remove “#” to run specific lines of code and produce output.

```
### remove all objects from workspace
#rm(List=Ls())

### set working directory
#setwd("~/My Documents")

### install & load metafor package
#install.packages("metafor")
library(metafor)

## Loading required package: Matrix
## Loading 'metafor' package (version 1.9-8). For an overview
## and introduction to the package please type: help(metafor).

### install & load robumeta package
#install.packages("robumeta")
library(robumeta)

## Loading required package: grid

### install & load clubSandwich package from Github
#install.packages("devtools")
library(devtools)

## WARNING: Rtools is required to build R packages, but is not currently
## installed.
##
## Please download and install Rtools 3.3 from http://cran.r-project.org/bin/windows/Rtools/ and then run find_rtools().

#install_github("jepusto/clubSandwich")
## "clubSandwich" repo does not appear in "user library", but is still
## running
library(clubSandwich)

### install violin plot package
#install.packages("vioplot")
library(vioplot)

## Loading required package: sm
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
```

Load in the dataset and create two new variables.

```
### load example data
SchoolMotivationRisk = read.csv("SchoolMotivationRisk.csv")

### create between-study version of covariates
SchoolMotivationRisk$aget1_m <- group.mean(SchoolMotivationRisk$aget1,
SchoolMotivationRisk$studyid)
SchoolMotivationRisk$aget2_m <- group.mean(SchoolMotivationRisk$aget2,
SchoolMotivationRisk$studyid)
```

Create a "violin plot" to visualize the plot.

```
#vioplot(SchoolMotivationRisk$yi, col="grey", names="School motivation risk")
#title(ylab="Risk-crime correlation")
```

Fit a naive random-effects model that ignores dependencies.

```
res_1<-rma(yi, vi, data=SchoolMotivationRisk)
#print(res_1)
#forest(res_1)
#funnel(res_1)
#predict(res_1, transf=transf.ztor, digits=2)
```

Fit a naive mixed-effects meta-regression model that ignores dependencies.

```
res_2<-rma(yi, vi, mods= ~ aget1_m*aget2_m, data=SchoolMotivationRisk)
#print(res_2)
```

Fit an RVE random-effects model with correlated effects weights.

```
### rho assumed = 0.8; sensitivity analyses could vary rho from 0.1 to 0.9
### small = TRUE option applies the small sample correction to df
res_3<-robu(formula = yi ~ 1, var.eff.size=vi, studynum = studyid,
modelweights = "CORR", rho = 0.8, small=TRUE, data=SchoolMotivationRisk)
#print(res_3)
```

Fit an RVE mixed-effects meta-regression model with age main effects. The Wald_test function tests the linear contrasts from a regression model, using a sandwich estimator for the variance-covariance matrix and a small sample correction for the p-value. Various corrections can be used by specifying test = "". Type "?Wald_test" for more information.

```
res_4<-robu(formula = yi ~ aget1_m + aget2_m, var.eff.size=vi, studynum =
studyid, modelweights = "CORR", rho = 0.8, small=TRUE,
data=SchoolMotivationRisk)
#print(res_4)
#Wald_test(res_4, constraints = 2:3, vcov="CR2")
```

Fit an RVE mixed-effects meta-regression model with age interaction.

```
res_5<-robu(formula = yi ~ aget1_m*aget2_m, var.eff.size=vi, studynum =
studyid, modelweights = "CORR", rho = 0.8, small=TRUE,
data=SchoolMotivationRisk)
#print(res_5)
#Wald_test(res_5, constraints = 2:4, vcov="CR2")
```

Fit an RVE mixed-effects meta-regression model with age interaction, adjusting for sex mix of samples.

```
res_6<-robu(formula = yi ~ aget1_m*aget2_m + sexmix, var.eff.size=vi,
studynum = studyid, modelweights = "CORR", rho = 0.8, small=TRUE,
data=SchoolMotivationRisk)
#print(res_6)
#Wald_test(res_6, constraints = 2:6, vcov="CR2")

### Omnibus F-test for categorical sex mix variable
#Wald_test(res_6, constraints = 4:5, vcov="CR2")
```

Create a graph to illustrate the distribution of the correlations by age. Start by manipulating the dataset.

```
#install.packages("ggplot2")
library(ggplot2)

### graph prediction line at different ages, overlaid with effect sizes
### create interaction variable
SchoolMotivationRisk$aget1_mXaget2_m <-
SchoolMotivationRisk$aget1_m*SchoolMotivationRisk$aget2_m

### create factored age1 variable
SchoolMotivationRisk$age1_cat <- ifelse(SchoolMotivationRisk$aget1 <= 12.9,
1,
ifelse(SchoolMotivationRisk$aget1 <= 14.9, 2,
ifelse(SchoolMotivationRisk$aget1 <= 16.9, 3,
ifelse(SchoolMotivationRisk$aget1 <= 18.9, 4, NA))))

SchoolMotivationRisk$age1_cat <- factor(SchoolMotivationRisk$age1_cat,
levels = c(1,2,3,4),
labels = c("11-13", "13-15", "15-17",
"17-19"))
```

Create a plot using "ggplots2". More information about the various modifiers may be found by typing "?ggplots2".

```
Graph_withoutSE <- ggplot(data = SchoolMotivationRisk,
aes(x = aget2_m, y = yi, color = factor(age1_cat)))
+
geom_point(position = "jitter") +
```

```

geom_smooth(method = lm, se = FALSE) +
scale_color_discrete(name = "Age at risk measurement") +
ylab("Risk-crime correlation") +
xlab("Age at crime outcome measurement") +
theme(axis.ticks.x = element_blank()) +
theme(text = element_text(size = 18)) +
theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(),
      panel.background = element_blank(), axis.line = element_line(colour =
"black"))
#Graph_withoutSE

Graph_withSE <- ggplot(data = SchoolMotivationRisk,
                      aes(x = aget2_m, y = yi, color = factor(age1_cat))) +
  geom_point(position = "jitter") +
  geom_smooth(method = lm) +
  scale_color_discrete(name = "Age at risk measurement") +
  ylab("Risk-crime correlation") +
  xlab("Age at crime outcome measurement") +
  theme(axis.ticks.x = element_blank()) +
  theme(text = element_text(size = 18)) +
  theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank(),
      panel.background = element_blank(), axis.line = element_line(colour =
"black"))
#Graph_withSE

```

Appendix B

Table 3 Data for tutorial example

studyid	yi	vi	sei	aget1	aget2	permale	sexmix
2	-0.060	0.001	0.027	14.5	18	1.000	All male
2	-0.050	0.001	0.027	14.5	18	1.000	All male
3	-0.050	0.001	0.028	14.5	18	0.000	All female
3	-0.050	0.001	0.028	14.5	18	0.000	All female
5	-0.010	0.006	0.080	11.5	13.5	1.000	All male
5	0.030	0.006	0.080	11.5	13.5	1.000	All male
5	0.100	0.006	0.080	11.5	13.5	1.000	All male
6	0.141	0.007	0.085	11.5	13.5	0.000	All female
6	0.110	0.007	0.085	11.5	13.5	0.000	All female
6	0.213	0.007	0.085	11.5	13.5	0.000	All female
7	0.245	0.006	0.080	13.5	15.5	1.000	All male
7	0.299	0.006	0.080	13.5	15.5	1.000	All male
7	0.266	0.006	0.080	13.5	15.5	1.000	All male
8	0.141	0.007	0.085	13.5	15.5	0.000	All female
8	0.182	0.007	0.085	13.5	15.5	0.000	All female
8	0.030	0.007	0.085	13.5	15.5	0.000	All female
24	0.117	0.008	0.089	11	13	0.533	Mixed sex
24	0.048	0.008	0.089	11	14	0.533	Mixed sex
24	0.189	0.008	0.089	11	12	0.533	Mixed sex
24	0.032	0.008	0.089	11	14	0.533	Mixed sex
24	0.246	0.008	0.089	11	14	0.533	Mixed sex
24	0.125	0.008	0.089	11	12	0.533	Mixed sex
24	0.065	0.008	0.089	11	13	0.533	Mixed sex
24	0.068	0.008	0.089	11	13	0.533	Mixed sex
24	-0.003	0.008	0.089	11	12	0.533	Mixed sex
25	0.377	0.007	0.085	14	15	0.550	Mixed sex
25	0.121	0.007	0.085	14	15	0.550	Mixed sex
25	-0.006	0.007	0.085	13	15	0.550	Mixed sex
25	0.130	0.007	0.085	14	15	0.550	Mixed sex
25	-0.007	0.007	0.085	12	15	0.550	Mixed sex
25	0.016	0.007	0.085	12	15	0.550	Mixed sex
25	0.279	0.007	0.085	13	15	0.550	Mixed sex
25	0.036	0.007	0.085	13	15	0.550	Mixed sex
25	0.026	0.007	0.085	12	15	0.550	Mixed sex
26	0.140	0.007	0.086	13	16	0.543	Mixed sex
26	0.176	0.007	0.086	14	15	0.543	Mixed sex
26	0.175	0.007	0.086	13	16	0.543	Mixed sex
26	0.214	0.007	0.086	13	15	0.543	Mixed sex
26	0.144	0.007	0.086	14	16	0.543	Mixed sex
26	0.068	0.007	0.086	13	15	0.543	Mixed sex

Table 3 (continued)

studyid	yi	vi	sei	aget1	aget2	permale	sexmix
26	0.201	0.007	0.086	14	16	0.543	Mixed sex
26	0.163	0.007	0.086	14	15	0.543	Mixed sex
26	0.208	0.007	0.086	14	16	0.543	Mixed sex
26	0.194	0.007	0.086	14	15	0.543	Mixed sex
26	0.107	0.007	0.086	13	15	0.543	Mixed sex
26	0.280	0.007	0.086	13	16	0.543	Mixed sex
27	0.149	0.008	0.089	14	17	0.523	Mixed sex
27	0.198	0.008	0.089	14	17	0.523	Mixed sex
27	0.231	0.008	0.089	14	15	0.523	Mixed sex
27	0.189	0.008	0.089	14	16	0.523	Mixed sex
27	0.216	0.008	0.089	14	16	0.523	Mixed sex
27	0.327	0.008	0.089	14	16	0.523	Mixed sex
27	0.221	0.008	0.089	14	15	0.523	Mixed sex
27	0.122	0.008	0.089	14	17	0.523	Mixed sex
27	0.076	0.008	0.089	14	15	0.523	Mixed sex
28	0.137	0.008	0.088	15	18	0.519	Mixed sex
28	0.182	0.008	0.088	15	18	0.519	Mixed sex
28	0.205	0.008	0.088	16	18	0.519	Mixed sex
28	0.120	0.008	0.088	16	18	0.519	Mixed sex
28	0.174	0.008	0.088	17	18	0.519	Mixed sex
28	0.033	0.008	0.088	16	18	0.519	Mixed sex
28	-0.006	0.008	0.088	17	18	0.519	Mixed sex
28	0.153	0.008	0.088	15	18	0.519	Mixed sex
28	-0.002	0.008	0.088	17	18	0.519	Mixed sex
29	0.119	0.008	0.090	16	18	0.556	Mixed sex
29	0.053	0.008	0.090	17	18	0.556	Mixed sex
29	0.068	0.008	0.090	16	18	0.556	Mixed sex
29	0.141	0.008	0.090	16	18	0.556	Mixed sex
29	0.135	0.008	0.090	16	18	0.556	Mixed sex
29	0.170	0.008	0.090	16	18	0.556	Mixed sex
29	0.014	0.008	0.090	17	18	0.556	Mixed sex
29	0.029	0.008	0.090	16	18	0.556	Mixed sex
29	-0.119	0.008	0.090	17	18	0.556	Mixed sex
29	0.139	0.008	0.090	17	18	0.556	Mixed sex
29	-0.049	0.008	0.090	17	18	0.556	Mixed sex
29	0.083	0.008	0.090	17	18	0.556	Mixed sex
30	0.098	0.014	0.120	17	19	0.534	Mixed sex
30	0.078	0.014	0.120	17	18	0.534	Mixed sex
30	-0.117	0.014	0.120	17	19	0.534	Mixed sex
30	0.122	0.014	0.120	17	18	0.534	Mixed sex
30	-0.097	0.014	0.120	17	20	0.534	Mixed sex

Table 3 (continued)

studyid	yi	vi	sei	aget1	aget2	permale	sexmix
30	-0.131	0.014	0.120	17	20	0.534	Mixed sex
30	-0.078	0.014	0.120	17	18	0.534	Mixed sex
30	-0.090	0.014	0.120	17	20	0.534	Mixed sex
30	-0.022	0.014	0.120	17	19	0.534	Mixed sex
36	0.000	0.002	0.046	13.75	19.5	1.000	All male
36	0.000	0.002	0.046	13.75	19.5	1.000	All male
61	0.152	0.001	0.028	17.68	18.5	1.000	All male
61	0.151	0.001	0.028	17.68	18.5	1.000	All male
61	0.090	0.001	0.028	15.49	18.5	1.000	All male
61	0.141	0.001	0.028	17.68	18.5	1.000	All male
61	0.077	0.001	0.028	15.49	18.5	1.000	All male
61	0.060	0.001	0.028	15.49	18.5	1.000	All male
61	0.090	0.001	0.028	15.49	18.5	1.000	All male
61	0.191	0.001	0.028	17.68	18.5	1.000	All male
61	0.213	0.001	0.028	17.68	18.5	1.000	All male
61	0.090	0.001	0.028	15.49	18.5	1.000	All male
61	0.151	0.001	0.028	17.68	18.5	1.000	All male
61	0.077	0.001	0.028	15.49	18.5	1.000	All male
86	0.010	0.005	0.072	14	16	1.000	All male
86	0.030	0.005	0.072	14	16	1.000	All male
86	0.251	0.004	0.062	13	15	1.000	All male
86	0.090	0.005	0.072	14	16	1.000	All male
86	0.161	0.005	0.072	14	16	1.000	All male
86	0.213	0.005	0.072	14	16	1.000	All male
86	0.363	0.004	0.062	13	15	1.000	All male
87	0.121	0.005	0.073	14	16	0.000	All female
87	0.383	0.004	0.065	13	15	0.000	All female
87	0.060	0.005	0.073	14	16	0.000	All female
87	0.070	0.005	0.073	14	16	0.000	All female
87	0.514	0.004	0.065	13	15	0.000	All female
87	0.234	0.005	0.073	14	16	0.000	All female
87	0.161	0.005	0.073	14	16	0.000	All female

References

*Studies providing effect sizes included in the example database

- Farrington, D. P. (2003). Developmental and life-course criminology: key theoretical and empirical issues—the 2002 Sutherland award address. *Criminology*, *41*, 221.
- Farrington, D. P. (2010). Life-course and developmental theories in criminology. In E. McLaughlin & T. Newburn (Eds.), *The Sage handbook of criminological theory* (pp. 249–269). Thousand Oaks: Sage.
- Le Blanc, M., & Loeber, R. (1998). Developmental criminology updated. *Crime and Justice*, *23*, 115–98.
- Piquero, A. R., Jennings, W. G., & Barnes, J. C. (2012). Violence in criminal careers: a review of the literature from a developmental life-course perspective. *Aggression and Violent Behavior*, *17*, 171–179.
- Farrington, D. P., Ttofi, M. M., & Coid, J. W. (2009). Development of adolescence-limited, late-onset, and persistent offenders from age 8 to age 48. *Aggressive Behavior*, *35*, 150–63.
- Stouthamer-Loeber, M., Loeber, R., Farrington, D. P., Wikstrom, P.-O. H., & Wei, E. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal of Consulting and Clinical Psychology*, *70*, 111–23.
- Thornberry, T. P., & Krohn, M. D. (2001). The development of delinquency: an interactional perspective. In S. O. White (Ed.), *Handbook of youth and justice* (pp. 289–305). New York: Kluwer Academic.
- Ioannidis, J. P. A. (2006). Evolution and translation of research findings: from bench to where? *PLoS Clinical Trials*, *1*, e36. doi:10.1371/journal.pctr.0010036.
- Ttofi, M. M., Farrington, D. P., & Lösel, F. (2012). School bullying as a predictor of violence later in life: a systematic review and meta-analysis of prospective longitudinal studies. *Aggression and Violent Behavior*, *17*, 405–418.
- Mitchell, O., Wilson, D. B., Eggers, A., & MacKenzie, D. L. (2012). Drug courts' effects on criminal offending for juveniles and adults. *Campbell Systematic Reviews*, *4*. doi:10.4073/csr.2012.4
- Braga, A. A., & Weisburd, D. L. (2012). The effects of “pulling levers” focused deterrence strategies on crime. *Campbell Systematic Reviews*, *6*. doi:10.4073/csr.2012.6
- Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499–525). San Diego: Academic Press.
- Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes: random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York: Russell Sage Foundation.
- Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: potential and promise. *Statistics in Medicine*, *30*, 2481–2498.
- Mavridis, D., & Salanti, G. (2011). A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research*, *22*, 133–158.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.
- Marín-Martínez, F., & Sánchez-Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: a cautionary note about procedures. *The Spanish Journal of Psychology*, *2*, 32–38.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *99*, 400–406.
- R Core Team. (2014). *R Foundation for Statistical Computing*. Vienna: R: a language and environment for statistical computing. <http://www.R-project.org/>.
- Tipton, E. (2015). Small simple adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*, 375–393.
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*, 604–634.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65.
- Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods*, *4*, 169–187.
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*, 13–30.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*, 97–111.

26. Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
27. Sánchez-Meca, J., & Marín-Martínez, F. (2010). Meta-analysis. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., Vol. 7, pp. 274–282). Oxford: Elsevier.
28. Beretvas, S. N., & Pastor, D. A. (2003). Using mixed-effects models in reliability generalization studies. *Educational and Psychological Measurement*, *63*, 75–95.
29. Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods*, *19*, 211–229.
30. Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*, 61–76.
31. Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: a comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*, 765–790.
32. Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576–594.
33. Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). London: Sage.
34. Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: a multilevel approach. *Behavior Research Methods*, *47*, 1274–1294.
35. Riley, R. D., Thompson, J. R., & Abrams, K. R. (2008). An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, *9*, 172–186.
36. Aloe, A. M. (2015). Inaccuracy of regression results in replacing bivariate correlations. *Research Synthesis Methods*, *6*, 21–27. doi:10.1002/jrsm.1126.
37. Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2013). Risk factors and crime. In F. T. Cullen & P. Wilcox (Eds.), *The Oxford handbook of criminological theory* (pp. 89–111). New York: Oxford University Press.
38. Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: correcting errors and bias in research findings* (2nd ed.). Newbury Park: Sage.
39. Wilson, D. B. (2013). *Calculating effect sizes [PowerPoint slides and YouTube video]*. Oslo: Campbell Collaboration. Retrieved from: http://www.campbellcollaboration.org/resources/training/The_Introductory_Methods.php.
40. Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48.
41. Fisher, Z., & Tipton, E. (2014). *Robumeta: Robust variance meta-regression*. Available at: <http://cran.r-project.org/web/packages/robumeta/index.html>
42. Pustejovsky, J. E. (2015). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections*. R package version 0.0.0.9000. URL: <https://github.com/jepusto/clubSandwich>
43. Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer.
44. Allaire, J.J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., & Hyndman, R. (2015). markdown: Dynamic documents for R (Version 0.8.1) [Software]. Available from <http://CRAN.R-project.org/package=markdown>
45. Piquero, A. R., Jennings, W. G., & Farrington, D. P. (2010). Self-control interventions for children under age 10 for improving self-control and delinquency and problem behaviors. *Campbell Systematic Reviews*, *2*. doi:10.4073/csr2010.2
46. Bennett, T., Holloway, K., & Farrington, D. P. (2006). Does neighborhood watch reduce crime? A systematic review and meta-analysis. *Journal of Experimental Criminology*, *2*, 437–458.
47. Mazerolle, L., Soole, D. W., & Rombours, S. (2006). Street-level draw law enforcement: a meta-analytical review. *Journal of Experimental Criminology*, *2*, 409–435.
48. Feder, L., & Wilson, D. B. (2005). A meta-analytic review of court-mandated batterer intervention programs: can courts affect abusers' behavior? *Journal of Experimental Criminology*, *1*, 239–262.
49. Pescott, O. L., & Stewart, G. B. (2014). Assessing the impact of human trampling on vegetation: a systematic review and meta-analysis of experimental evidence. *PeerJ*, *2*, e360.
50. Stein, A., Gerstner, K., & Krefl, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, *17*, 866–880.
51. Peng, P., & Fuchs, D. (2014). A meta-analysis of working memory deficits in children with learning difficulties: is there a difference between verbal domain and numerical domain? *Journal of Learning Disabilities*, online first, 1–18. doi:10.1177/0022219414521667.
52. De Vibe, M., Bjørndal, A., Tipton, E., Hammerstrøm, K. T., & Kowalski, K. (2012). Mindfulness based stress reduction (MBSR) for improving health, quality of life, and social functioning in adults. *Campbell Systematic Reviews*, *8*(3)

53. Park, D. E., Johnson, T. S., Nonyane, B. A. S., Chandir, S., Conklin, L., Fleming-Dutra, K. E., & Knoll, M. D. (2014). The differential impact of coadministered vaccines, geographic region, vaccine product and other covariates on pneumococcal conjugate vaccine immunogenicity. *The Pediatric Infectious Disease Journal*, 33(Suppl 2), S130.
54. Parkinson, J. R., Hyde, M. J., Gale, C., Santhakumaran, S., & Modi, N. (2013). Preterm birth and the metabolic syndrome in adult life: a systematic review and meta-analysis. *Pediatrics*, 131, e1240–e1263.
55. Shields, G. S., Bonner, J. C., & Moons, W. G. (2015). Does cortisol influence core executive functions? A meta-analysis of acute cortisol administration effects on working memory, inhibition, and set-shifting. *Psychoneuroendocrinology*, 58, 91–103.
56. Falkenström, F., Markowitz, J. C., Jonker, H., Philips, B., & Holmqvist, R. (2013). Can psychotherapists function as their own controls? Meta-analysis of the “crossed therapist” design in comparative psychotherapy trials. *The Journal of Clinical Psychiatry*, 74, 482.
57. Tanner-Smith, E. E., & Lipsey, M. W. (2015). Brief alcohol interventions for adolescents and young adults: a systematic review and meta-analysis. *Journal of Substance Abuse Treatment*, 51, 1–18.
58. *Ayers, C.D., Williams, J.H., Hawkins, J.D., Peterson, P.L., & Abbott, R. D. (1999). Assessing correlates of onset, escalation, de-escalation, and desistance of delinquent behavior. *Journal of Quantitative Criminology*, 15, 277–306.
59. *Elliott, D. S., & Voss, H. L. (1974). *Delinquency and dropout*. Lexington: D.C. Health and Company.
60. *Espiritu, R. C. (1998). *Are girls different? An examination of developmental gender differences in pathways to delinquency*. Dissertation Abstracts International, 59(06), 3055.
61. *Iovanni, L. M. (1990). *Age, cohort, and period differences in the effect of social control, social learning, and strain variables on self-reported delinquency*. Dissertation Abstracts International, 51(05), 1783.
62. *Loeber, R. (1996). The precursors of young fatherhood and its effect on the delinquency and prosocial careers of teenage males. In R. Loeber, D. H. Huizinga, & T. P. Thornberry (Eds.), *Program of research on the causes and correlates of delinquency: Annual Report 1995–1996* (pp. 306–326).
63. *Wiatrowski, M. D., Hansell, S., Massey, C. R., & Wilson, D. L. (1982). Curriculum tracking and delinquency. *American Sociological Review*, 47, 151–160.