Addressing Dependency in Meta-Analysis: A Companion to Assink and Wibbelink (2016)

Mark Assink [a] ✉ ⓘ and Carlijn J. M. Wibbelink [b] ⓘ

[a]Research Institute of Child Development and Education, University of Amsterdam
[b]Psychology Research Institute, Department of Clinical Psychology, University of Amsterdam

Abstract ∎ This research note elaborates on addressing dependency in effect size data and serves as a companion to our tutorial on fitting three-level meta-analytic models in R (Assink & Wibbelink, 2016). We provide a description of effect size and standard error dependency, explain how both the multilevel and multivariate meta-analytic models handle these types of dependency, and discuss the role of alternative methods in addressing dependency in effect size data, including approximating a variance-covariance matrix and applying a cluster-robust inference method. These alternative methods are illustrated with example R code that builds upon the effect size dataset that we presented and analyzed in our tutorial. We conclude that more simulation studies are needed to provide clearer guidelines for modeling dependency in effect size data and urge statisticians to make the available technical literature further accessible to applied researchers.

Keywords ∎ meta-analysis, three-level meta-analysis, multivariate meta-analysis, robust variance estimation, effect size dependency, sampling error dependency. **Tools** ∎ R.

✉ m.assink@uva.nl

Introduction

Research synthesis in the form of meta-analysis has a dramatic impact on the development of cumulative knowledge in psychology and other disciplines (DeGeest & Schmidt, 2011). As such, meta-analysis has been acknowledged as one of the most important methodological developments in behavioral and other sciences (Cooper et al., 2010; Egger et al., 2001). Most conventional methods for meta-analysis hold the assumption that the synthesis is based on effect sizes that are not related to each other (e.g., Cheung, 2014; Rosenthal, 1984). However, primary studies often report multiple effect sizes that are eligible for inclusion given the scope of a meta-analysis, implying that effect sizes are related to each other. This effect size dependency – also referred to as effect size interrelatedness or interdependency – may occur, for instance because separate experiments were conducted in a primary study, a primary study used multiple methods (e.g., self-report, interviews) for assessing the same construct, or the same participants were examined over multiple timepoints. In these examples, the assumption that effect sizes are independent from each other is clearly violated. If a meta-analyst synthesizes de-

pendent effect sizes using conventional methods for meta-analysis, the results can be incorrect and even misleading (Borenstein et al., 2009). More specific, ignoring dependency tends to underestimate standard errors that in turn results in too narrow confidence intervals, and consequently in an increased likelihood for falsely rejecting null hypotheses (i.e., an inflated type-1 error rate; Hedges, 2009; Snijders & Bosker, 2012).

The question how to deal with dependency in effect size data poses an important challenge that researchers often face when conducting a meta-analysis. In the past few decades an increasing number of scholars have devoted considerable efforts into developing innovative statistical techniques for analyzing dependent effect size data in meta-analysis. One of these techniques comprises the synthesis of dependent effect size data in a three-level meta-analytic model, which we have illustrated in our prior tutorial (Assink & Wibbelink, 2016). Our tutorial provides an introduction to the application of multilevel modeling to meta-analysis, but is rather concise in describing the dependencies that may occur in effect size data and the way in which multilevel meta-analysis deals with dependency in effect size data. This paper serves as a companion to
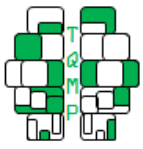
our tutorial and elaborates on this issue. As technical papers as well as excellent overview work on modeling effect size dependency already exist (e.g., Becker, 2000; Cheung, 2014; Fernández-Castilla et al., 2019, 2020; Gleser & Olkin, 2009; Konstantopoulos, 2011; Tipton, 2013; Van den Noortgate et al., 2015), it is not our aim to give a comprehensive review of modeling techniques nor to provide strict modeling guidelines. Instead, this paper elaborates on modeling dependency in meta-analysis that complements our tutorial, and offers practical information that researchers may find useful in their own projects.

Two Types of Dependent Data Structures

Let's first focus on the nature of dependency in effect size data. There are two different forms of dependency that can emerge in effect size data, although both forms can occur in one effect size dataset. First, a single primary study may examine two or more unique or non-overlapping groups of participants and report an effect size for each participant group. For example, Romans et al. (1997) evaluated the effect of Turner's syndrome on executive functioning of participants across three mutually exclusive age groups. In this example, one study produces multiple effect sizes, which were all obtained in the same "study context". Effect sizes produced in the same context are more similar than effect sizes produced in different contexts, for instance because the effect sizes were obtained by the same researchers, who used the same questionnaires, which were filled out under the same conditions. This type of dependency stems from effect sizes being nested within a higher-level cluster, which are commonly referred to as nested effect sizes. The consequence of this clustering is that one effect size tells us something about the direction and strength of another effect size in the same cluster, and thus there is effect size dependency.

Second, studies may assess multiple outcomes in the same or partly the same group of participants and report multiple effect sizes (Nakagawa et al., 2023). For example, Boterhoven De Haan et al. (2020) compared the effectiveness of two trauma-focused treatments for adults with post-traumatic stress disorder (PTSD) from childhood trauma. The authors reported multiple results in their study, as they measured PTSD symptoms with both a self-report questionnaire and a structured clinical interview. In this scenario, overlapping groups of participants contribute to more than one outcome, and the effect size data are therefore "multivariate". In case of multivariate effect size data, there is not only dependency in effect sizes as they share the same study context, but also dependency in the sampling errors (or "sampling variances") of these effect sizes. After all, if the same or partly the same participants contribute to multiple effect size estimates, then these effect sizes have correlated estimation errors (Gleser & Olkin, 2009, p. 284) implying there is not only effect size dependency but also sampling error dependency.

Handling Effect Size Dependency in Advanced Meta-Analytic Models

The question how to deal with dependency in effect size data poses an important challenge that researchers often face when conducting a meta-analysis. Multiple traditional ad-hoc techniques for dealing with dependency have been described in the literature, such as averaging effect sizes within studies, selecting only one effect size per study, and shifting the unit of analysis (Assink & Wibbelink, 2016). The problem with these techniques is not only loss of information resulting in a lower statistical power, but also a limit in research questions that can be addressed in a meta-analysis as informative differences between effect sizes are lost (see Cheung, 2015, for a detailed description of these and other limitations). In the past few decades an increasing number of scholars have devoted considerable efforts into developing innovative statistical techniques for modeling dependency in effect size data in meta-analysis. By stepping away from the traditional univariate approach to meta-analysis that assumes effect size independency, it becomes possible to deal with dependency in effect size data so that all relevant effect sizes – regardless of overlapping samples and/or contexts – can be extracted and analyzed. By synthesizing all relevant effect sizes, a maximum of information is preserved and optimal statistical power can be achieved (Assink & Wibbelink, 2016). Furthermore, by extracting multiple relevant effect sizes rather than one effect size per study the number of research questions that can be addressed in meta-analytic research increases (Cheung, 2015). Two common statistical methods for handling dependency in effect size data are multivariate meta-analysis and multilevel meta-analysis.

Multivariate Meta-Analysis

Analyzing effect sizes in a multivariate model is one of the earliest meta-analytic techniques to address dependency in effect size data due to overlapping samples (Kalaian & Raudenbush, 1996; Raudenbush et al., 1988). A multivariate meta-analysis is an extension of the traditional (univariate) meta-analysis that enables a simultaneous analysis of multiple effect sizes (or outcomes) that come from individual studies. For each of the included studies, there is not one single effect size, but a so-called vector of effect sizes (i.e., a one-column matrix of effect sizes) is created and used in the analyses. When multiple effect sizes are extracted from primary studies, they can differ in strength (and even in direction) across studies, but also within studies. The multivariate model takes both this between-study

and within-study variability in effect sizes into account. The within-study variability is captured by the sampling variance-covariance matrix that is assumed to be known and encompasses the individual variance of each effect size as well as the relations (covariances) between the sampling errors of effect sizes within an individual study. The between-study variability is represented by the estimated population heterogeneity variance-covariance matrix that captures the variability in true effect sizes across studies (Cheung, 2019). Multivariate meta-analysis offers numerous benefits such as incorporating multiple outcomes of individual studies, examining the correlation between outcomes, and testing for differential moderator effects (Cheung, 2015; Van den Noortgate et al., 2013). Moreover, multivariate meta-analysis generally leads to more precise estimates than univariate meta-analysis (Price et al., 2019; Riley, 2009).

Multilevel Meta-Analysis

Multilevel meta-analyses are typically used in contexts where effect sizes are nested in a higher-level cluster (i.e., a multilevel data structure). An effect size dataset is multilevel in nature when effect sizes can be grouped together based on one or more higher-level clustering variables, such as experiments, studies, research groups, or countries (e.g., Hox et al., 2017; Konstantopoulos, 2011; Raudenbush & Bryk, 1985). A meta-analytic model that allows effect sizes to be correlated because they share certain cluster-level characteristics is also known as the hierarchical dependence model (Hedges et al., 2010). When effect sizes are related because they share certain cluster-level characteristics (i.e., the effect sizes "cluster" together), it can be assumed that the underlying population (or "true") effects are more alike within the same level of a clustering variable than across different levels of that clustering variable. This implies that effect sizes belonging to the same cluster are dependent on each other. A meta-analyst can account for this type of effect size dependency by adding cluster-specific random effects to the statistical model, which implies that a random effect is added to each level of the model that corresponds to a grouping or clustering variable. Such random effects meta-analytic models can account for between- and within-cluster heterogeneity in effect sizes, and thus for the within-cluster correlation in the underlying true effects.

Although multilevel meta-analyses are commonly applied in situations with nested effect sizes, a specific application of the multilevel meta-analysis can be used in the context of multivariate effect sizes. Our tutorial illustrates how dependency in effect size data due to overlapping samples can be modeled in a three-level meta-analytic model (Assink & Wibbelink, 2016). This three-level approach to

meta-analysis was introduced by Geeraert et al. (2004) and has been explained and described by multiple methodological scholars (e.g., Cheung, 2014; Van den Noortgate et al., 2013, 2015). In this approach, the effect size data are considered to have a hierarchical or nested structure in which participants (level 1) are nested within outcomes (level 2), which are nested within studies (level 3). By adding a random effect to the (two) higher-order levels of this model, the within- and between-study variability in effect sizes (i.e., the effect size heterogeneity) is modeled, and thus the within-study correlation in the underlying true effects is accounted for. As a result, this three-level model distributes the total variance in effect sizes across three levels: the sampling variance of the individual effect sizes at level 1, the within-study variance in effect sizes at level 2, and the between-study variance in effect sizes at level 3. The variation in effect sizes at level 1 of this model (the sampling variance) is not estimated in the analysis, but approximated using statistical theory and by relying on pooled statistics, which is often the only available information for the meta-analyst. The three-level structure is a rather straightforward, but powerful way to model dependency in effect size data that enables the extraction of multiple effect sizes from individual studies that meet the inclusion criteria of a meta-analysis.

Van den Noortgate and Onghena (2003) showed that applying the multilevel approach to meta-analysis is as effective and accurate in estimating the model coefficients as more traditional random effects techniques. A particularly strong advantage of a multilevel meta-analytic model is its flexibility (Van den Noortgate et al., 2013, 2015). Multiple predictors can easily be added as covariates to the model in attempts to explain within- and/or between-study variance in effect sizes. Moreover, the multilevel model can easily be extended with additional random effects to further model dependency of effect sizes within and between studies (Fernández-Castilla et al., 2020). However, the question arises how the three-level meta-analytic model deals with effect size and sampling error dependency in multivariate effect size data. The three-level meta-analytic model does so by assigning a random effect to both the outcomes at level 2 and the studies at level 3 of the model. As such, the effect size dependency is explicitly modeled, but the dependency in effect size sampling error is not. This seems to be at odds with the multivariate model to meta-analysis prescribing that both effect size dependency and sampling error dependency must be modeled whenever the same or partly the same participants contribute to multiple effect sizes in a primary study. Van den Noortgate et al. (2013) who elaborately describe the three-level meta-analytic model do consider both effect size dependency and standard error dependency, but state that the latter is ac-

counted for in the three-level model by overestimating the study-level variance which subsequently "stands in" for the correlation in standard errors. Put differently, the random effects of the three-level model allow the standard error dependency to "subsume" into the correlation of the underlying true effects through the specification of the model's hierarchical structure. As a result, the estimates of the mean effect sizes and their standard errors that are produced in the analyses are appropriate (Van den Noortgate et al., 2013).

Comparing the Multivariate Meta-Analysis to the Three-Level Meta-Analysis

The multivariate and multilevel meta-analytic models are mathematically closely related. Consequently, they share similarities, yet there are notable differences (see Cheung, 2015, particularly Section 6.4). The multivariate meta-analytic model has three key features: 1) it allows for differences in population means across outcomes, 2) the observed effect sizes are conditionally dependent, and 3) both the degree of variability in true effect sizes and the covariance between true effect sizes can vary (Cheung, 2015, p. 195). The three-level meta-analytic model can be conceptualized as a special case of the multivariate meta-analytic model in which three constraints are modeled (Cheung, 2015): 1) the population effect sizes are equal and, therefore, exchangeable within a cluster, 2) the observed effect sizes are conditionally independent of each other within a cluster, and 3) the variances of the true effect sizes are composed by the level-2 and level-3 variances, while the covariances of the true effect sizes equal the level-3 variance. When certain assumptions are applied to the effect size data, a multivariate model can be approximated with a three-level model. Specifically, these assumptions involve an equal level of heterogeneity and covariance for the true effect sizes, and an equal level of conditional covariance for the observed effect sizes (Cheung, 2015). Although some of these assumptions may be wrong, there are indications that the three-level approach to meta-analysis works reasonably well with multivariate effect size data. Simulation studies performed by Van den Noortgate et al. (2013) illustrate that multilevel meta-analytic models can indeed account for all dependency in multivariate effect size data including effect size and standard error dependency. Moeyaert et al. (2017) underline this finding with results from their simulation study and conclude that multilevel meta-analytic models validly and efficiently account for within-study effect size dependency, and that (explicitly) modeling correlations between standard errors of effect sizes is not needed. According to these studies, applying a three-level structure when analyzing multivariate effect size data is sufficient to deal with both effect size de-

pendency and sampling error dependency.

However, the simulation studies of Van den Noortgate et al. (2013) and Moeyaert et al. (2017) have been reflected upon by others, for instance because only bivariate meta-analytic models without covariates were simulated while in most meta-analyses a substantial number of variables are tested as moderator (Viechtbauer, 2017). Moreover, the assumption that heterogeneity in effect sizes is the same for all outcomes synthesized in a meta-analysis is an assumption that can be difficult to hold (Viechtbauer, 2020). Violating this assumption may imply that confidence intervals around estimates for more heterogeneous outcomes are too narrow (due to an underestimated standard error), whereas too wide confidence intervals may be produced around estimates for less heterogeneous outcomes (due to an overestimated standard error; Viechtbauer, 2017). It has therefore been argued that whenever the same or partly the same participants contribute to multiple effect sizes, a multivariate meta-analytic model is to be preferred so that both effect size dependency and standard error dependency are explicitly modeled. Further, three-level meta-analytic models incorrectly assume that within-study correlations between outcomes and thus sampling covariances of (within-study) effect sizes are zero (Fernández-Castilla et al., 2021; Van den Noortgate et al., 2013, 2015). Although simulation studies have revealed that three-level models are robust to this misspecification of the correlation structure (Moeyaert et al., 2017; Van den Noortgate et al., 2013, 2015), there are advocates of the multivariate meta-analytic method suggesting that the multilevel model may produce invalid estimates of model coefficients whenever the same or partly the same participants contribute to multiple outcomes in studies (e.g., Viechtbauer, 2017; Yang et al., 2023). They argue that explicitly modeling both effect size dependency and sampling error dependency by applying a multivariate meta-analytic model yields more appropriate estimates of model coefficients than a multilevel meta-analytic model.

Constructing a Variance-Covariance Matrix in Multivariate Meta-Analysis

Given the multivariate method for meta-analysis, how can a meta-analyst explicitly model both effect size and standard error dependency? The answer lies in constructing a "variance-covariance matrix" (or in short "covariance matrix") that forms the basis for many multivariate techniques and contains information that is used in estimating the model coefficients and their error terms. Basically, this matrix is a squared table with the same set of variables in the table's rows and columns. The numbers on the table's diagonal that goes from the top-left to the bottom-right represent the variances of the variables, whereas all

off-diagonal numbers represent the covariances of all pair-wise combinations of the variables. The sampling variance refers to the variation in individual variables, whereas a co-variance is an unstandardized correlation representing the linear association between two variables and is a measure of how change in one variable is associated with change in another variable. The covariance between the same two variables equals a variable's variance, and therefore the variances are captured by the diagonal elements of the table. In multivariate meta-analysis, a variance-covariance matrix captures the sampling variances of (within-study) outcomes or observed effect sizes in its diagonal elements and the covariance between all pairwise combinations of two outcomes or observed effect sizes in its off-diagonal elements. So, this matrix indicates how (within-study) outcomes vary and covary, and thus provide information about dependency in outcomes. This matrix can be fed into the meta-analytic model resulting in a multivariate meta-analytic model that accounts for effect size dependency through the specified random effects and for standard error dependency through the information in the variance-covariance matrix.

Unfortunately, constructing a variance-covariance matrix can be difficult. Particularly computing the covariances between the within-study outcomes or observed effect sizes is challenging, as the correlations that are required for those computations are rarely reported by study authors (e.g., Fernández-Castilla et al., 2020). If, for instance, a primary study reports multiple effect sizes because multiple outcomes were examined in a sample to measure one underlying construct, the correlation between those outcomes is required to calculate the covariance between the effect sizes for these outcomes. Consequently, not knowing the within-study correlation poses a problem for modeling standard error dependency in a multivariate meta-analytic model. In contrast, Van den Noortgate et al. (2013) state that the multilevel approach to meta-analysis does not require exact knowledge of the sampling covariances between effect sizes, as the between-study variance acts as a "stand in" for the sampling co-variances. This makes multilevel meta-analysis convenient and appealing, as the lack of information on the covari-ances between effect sizes does not seem to be problematic according to simulation studies (Moeyaert et al., 2017; Van den Noortgate et al., 2013, 2015). In fact, lacking information on covariances is exactly the reason why multi-variate meta-analyses are only rarely performed. Besides adopting a multilevel meta-analytic approach as an alternative method for handling multivariate effect size data, other techniques are available including the construction of an approximated variance-covariance matrix or applying a cluster-robust inference method to a meta-analytic model (Hedges et al., 2010; Pustejovsky & Tipton, 2022; Tipton, 2015).

Approximating a Variance-Covariance Matrix

When the true variance-covariance matrix cannot be computed because information on covariances is not available, a meta-analyst may choose to approximate the multivariate meta-analytic method by constructing an "approximated" or "working" variance-covariance matrix. In such a matrix the effect size covariances are calculated using an informed estimate (or "guestimate") of one "common" correlation between the observed effect sizes. The underlying assumption of this matrix is a single or one common correlation between all (pairs of) effect sizes that were obtained from the same study, and which is the same across all studies. Pustejovsky and Pustejovsky and Tipton (2022, p. 429) refer to this premise as the "constant sampling correlation" assumption. For instance, a meta-analyst that combines both observational and self-report measures of children's eating behavior in one meta-analysis to study the effects of school-based weight interventions may derive from previous empirical research that the correlation between observations and self-reports may be as high as .47 (Merson et al., 2016). This correlation estimate can then be used to construct an approximation of the true variance-covariance matrix. However, when the between-outcomes correlation is estimated as .47, it is implicitly assumed that this correlation holds for the covariances of all pairs of outcomes within and across all studies that are included in the meta-analysis. This may not be realistic and thus difficult to justify, but a variance-covariance matrix can also be constructed using more than one correlation. For instance, when primary studies partly report on within-study outcome correlations, the variance-covariance matrix can be constructed using these reported correlations in addition to an estimated correlation for unreported associations between outcomes. The matrix is then no longer based on the assumption of "constant sampling correlation" but on the assumption of what Pustejovsky and Tipton (2022, p. 430) refer to as the "partially empirical correlations" assumption.

As deciding on the strength of the association(s) between within-study outcomes is challenging, it can be wise to conduct sensitivity analyses with variance-covariance matrices that are based on different estimates of the within-study outcome correlation(s) (see, for instance, Hutchinson et al., 2022; Li et al., 2022; Oliveira et al., 2022). Building on our earlier example, the meta-analyst could compute additional matrices based on lower (e.g., $\rho = .20$) and higher (e.g., $\rho = .80$) constant sampling correlations than the correlation ($\rho = .47$) that was initially used in constructing the variance-covariance matrix. Next, by performing a sensi-

tivity analysis for each of the additional matrices, the meta-analyst can determine to what extent the results are sensitive to alternative decisions on the strength of the correlation(s) between within-study outcomes. Ideally, the initially performed analyses and the sensitivity analyses produce similar results, so that the conclusion is that results hold across different estimates of the within-study correlation(s).
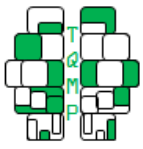
Applying a Cluster-Robust Variance Estimation Method

Another alternative method to handle dependency in effect size data is the Robust Variance Estimation (RVE) method (Hedges et al., 2010; Tipton & Pustejovsky, 2015). Cluster-robust variance estimation methods – also referred to as "sandwich estimators" because of the structure of the underlying formula components – are becoming increasingly popular inferential methods that can be used in making inferences from regression models and do not require precise knowledge of the covariances between (within-study) outcomes and the distribution of standard errors. To put it very simply, RVE tries to "polish up" the standard errors of fixed effect estimates. In this method, the variance components of the model are viewed as auxiliary to the analysis, rather than being central parameters for inference or description (Pustejovsky & Tipton, 2022). For multivariate effect size data, Fisher and Tipton (2015) state that RVE produces valid standard errors, (mean) effect size estimates, confidence intervals, and significance tests in meta-regression without the need to model the exact nature of the dependency in effect size data. Although RVE does not require a specification of the covariance structure, the performance of the RVE method improves when a so-called "working model" is specified that describes the dependency in the effect size data and serves as input for the RVE method (Hedges et al., 2010). Pustejovsky and Tipton (2022) provided a decision tree that meta-analysts can use to select a suitable working model given the characteristics of the effect size dataset that is to be analyzed. These authors describe several working models including the hierarchical effects (HE) model, the correlated effects (CE) model, and the correlated and hierarchical effects (CHE) model that we briefly discuss here (see pp. 427-432 in Pustejovsky & Tipton, 2022, for their decision tree and a full list of all working models they describe).

In short, the HE model addresses dependency in effect size data that is caused by non-overlapping samples which are nested in a higher-level cluster. For instance, when research groups have each produced multiple studies that are included in a meta-analysis, then studies are nested within research groups. This nested structure is accounted for in the HE model which models within-cluster and between-cluster variation in the underlying true ef-

fect sizes and assumes dependency in true effects while considering sampling errors of the observed effect sizes as independent. A different model is the CE model that assumes sampling errors of effect sizes to be dependent because (partly) the same participants contribute to multiple effect sizes (i.e., effect sizes are based on overlapping samples). However, it does not assume within-study variation in the underlying true effect sizes. Finally, the CHE model combines characteristics of the HE and CE models and allows for within-study and between-study variation in the underlying true effect sizes as well as correlated sampling errors of the observed effect sizes. By first specifying a working model that fits the structure of the effect size dataset at hand and applying RVE for obtaining corrected standard errors and hypothesis tests thereafter, a meta-analyst benefits from the efficiency of a working model capturing the effect size dependency while retaining the robustness to potential model misspecification (Hong et al., 2018). Put simply, combining a working model with the RVE technique may address the dependency in effect size data more accurately, which presumably leads to more precise and accurate model coefficients than using RVE alone (Tipton, 2015; Tipton & Pustejovsky, 2015). So, RVE should not be regarded as an alternative technique to a multivariate or multilevel meta-analytic model, but as a complementary technique to a working model that provides a safeguard against model misspecification (Pustejovsky & Tipton, 2022; Tipton & Pustejovsky, 2015).

The originally developed RVE method requires a rather large number of studies in a meta-analysis to obtain accurate results (Hedges et al., 2010). Hedges et al. (2010) roughly suggested that at least 40 studies need to be synthesized for valid results, as synthesizing less studies may lead to underestimated standard errors and inflated type-1 error rates (see also Tipton, 2015). However, depending on the research questions that need to be addressed, it is often not feasible to identify and retrieve so many studies. Therefore, small-sample corrections have been developed for RVE that are based on applying the Satterthwaite correction to the degrees of freedom of model coefficient tests, so that the risk of a type-1 error rate in small sample meta-analyses decreases (Tipton, 2015; Tipton & Pustejovsky, 2015). However, it was found that these methods may still suffer from inflated type-1 error rates or from below-nominal type-1 error rates (Joshi et al., 2022; Tipton & Pustejovsky, 2015);. For specifically multiple-comparison tests, Tipton and Pustejovsky (2015) showed that small sample corrections (and the "HTZ" test in particular) may have low statistical power. In other words, the small sample corrections may not be trustworthy and overly conservative in multiple-comparison tests (Joshi et al., 2022).

In their attempt to overcome these problems, Joshi et al.

(2022) developed an alternative method for correcting RVE when the number of studies is limited. Their technique is based on cluster wild bootstrapping (CWB), which involves re-sampling of entire clusters from the original effect size data (Cameron et al., 2008). This technique does not require a large number of clusters nor that clusters have the same size. Also, effect size sampling errors do not need to be independent and identically distributed (Cameron et al., 2008; MacKinnon, 2009). The simulation studies that Joshi et al. (2022) performed reveal that CWB adequately controls for the type-1 error rate and that CWB has more statistical power compared to other techniques, particularly in multiple-comparison tests. Based on their results, Joshi et al. recommend using CWB for multiple-comparison tests in meta-analyses conducted with RVE although they also stress that more simulation studies are needed to further examine the performance of this technique.

Working Example in R

We have now briefly covered how dependency in effect size data is mainly dealt with in multilevel versus multivariate meta-analytic models, and described what role an approximated variance-covariance matrix and the application of RVE (with small sample adjustment) to a working model can fulfil in estimating coefficients in a meta-analytic model. In our tutorial (Assink & Wibbelink, 2016), we have illustrated a three-level meta-analysis of effect sizes that express the difference in recidivism between delinquent juveniles with versus without a mental health disorder. We used a subset of the dataset synthesized in the original study of Wibbelink et al. (2017) that comprised 100 effect sizes extracted from 17 individual studies (between 1 and 22 effect sizes were extracted per study). The effect sizes were multivariate as the same samples contributed to multiple effect size estimates, due to for example different sources of information (e.g., official records and self-reports) and different types of outcomes (e.g., rearrests and reincarcerations). The standardized mean difference (Cohen's d) was used as common effect size with positive d values indicating more recidivism in juveniles with a mental health disorder compared to juveniles without a mental health disorder. The correlations between outcomes were not reported in the individual studies and therefore applying a multivariate meta-analysis was not feasible. Consequently, an alternative method was needed to include all relevant information without ignoring the dependency between effect sizes. In our tutorial (Assink & Wibbelink, 2016) we demonstrated how the effect size dataset can be analyzed in a three-level meta-analysis. The questions that now rise are (1) how the alternative methods described in this paper can be applied to the dataset we analyzed in our tutorial, including the multivariate meta-analytic

approach with an approximated variance-covariance matrix and the RVE method in combination with a working model, and (2) to what extent the results differ between the three-level meta-analysis, the multivariate meta-analytic approach with an approximated variance-covariance matrix, and the RVE method applied to two working models (i.e., the model based on the multivariate approach and the three-level meta-analytic model).

Results of Alternative Approaches

Wolfgang Viechtbauer (Viechtbauer, 2021) showed in the R statistical environment using his metafor package how our effect size dataset — which is available online as the appendix of Assink and Wibbelink (2016) — can be analyzed with the multivariate approach using an approximated variance-covariance matrix, and how the RVE method can be applied. Below, we present R syntax and output to guide readers in running the analyses and interpreting the results. For brevity, we do not elaborate on important preliminary steps in fitting a meta-analytic model (e.g., outlier detection and testing assumptions), but refer the reader to for instance Hox et al. (2017) and Lipsey and Wilson (2001) and references therein. After installing and loading the `metafor` package, the dataset needs to be imported into the R environment (see Listing 1). Next, by running the syntax in Listing 2, the overall association between juveniles with a mental health disorder and recidivism is estimated based on a multivariate meta-analytic approach with an approximated variance-covariance matrix. We presume a "common" correlation of .60 between all pairs of effect sizes within and across all included primary studies. The `vcalc` function in Listing 2 is part of the metafor package and enables the construction or approximation of the variance-covariance matrix of the sampling errors of dependent effect sizes. The argument `v` refers to the name of the variable that contains all sampling variances of the observed effect sizes. The `cluster=` argument specifies the clustering variable (i.e., studyID) and the `obs` argument specifies the variable that uniquely identifies the observed effect sizes (i.e., effectsizeID). In addition, `data=dataset` is the argument that specifies what object contains the dataset. Finally, the `rho` argument indicates the correlation between the observed effect sizes, which, in this instance, we assume to be .60.

The `rma.mv` function in Listing 2 makes it possible to run a multivariate or multilevel meta-analytic model, in which the `random=` argument specifies the type of parameterization. Similar to our tutorial (Assink & Wibbelink, 2016), we fitted the model with a multilevel — and specifically a three-level – parameterization. However, a multivariate model parameterization would not make any difference as both parameterizations produce

identical results whenever the estimated correlation between underlying true effects (ρ) is positive (Viechtbauer, 2022). The `vcmatrix` argument contains the approximated variance-covariance matrix that is estimated with the `vcalc` function. Running the syntax in Listing 2 produces the output as shown in Output 1. For a detailed explanation of the `rma.mv` function in Listing 2 and the output in Output 1, the reader is referred to Assink and Wibbelink (2016). Based on the output, we can conclude that d equals 0.368 ($p < .001$) in the multivariate meta-analytic approach with an approximated variance-covariance matrix based on a "common" correlation of $\rho = .60$.

Next, two sensitivity analyses are conducted by changing the assumed correlation between the observed effect sizes from $\rho = .60$ into $\rho = .40$ and $\rho = .80$. See the syntax in Listing 3 and output in Output 2. We can conclude that d equals 0.385 ($p < .001$) and 0.354 ($p < .001$) in the sensitivity analyses when assuming a "common" correlation of $\rho = .40$ and $\rho = .80$, respectively.

We now proceed by demonstrating the RVE method to handle dependency in effect size data. As RVE should not be regarded as a substitute for a multivariate or multilevel meta-analytic model, but rather as a complementary technique to a working model, we fitted both the three-level meta-analytic model and the model based on the multivariate approach with an approximated variance-covariance matrix in combination with RVE. First, the clubSandwich package has to be installed and loaded into the R environment; see the syntax in Listing 4. Next, the overall effect is estimated using the RVE method with two working models: the model based on the multivariate meta-analytic approach with an approximated variance-covariance matrix of $\rho = .60$ (see Listing 5) and the three-level meta-analytic model (see Listing 6). The `robust` function in Listings 5 and 6 enables cluster-robust tests and retrieves cluster-robust confidence intervals of the model coefficients of the specified object, which in this case are the objects for the model based on the multivariate approach ("overallmultivariate") and the three-level model ("overallmultilevel").

Tests of individual coefficients and confidence intervals in the `robust` function are by default based on a t-distribution, whereas the omnibus test uses an F-distribution (Viechtbauer, 2023b, p. 294). It is therefore not necessary to add the `tdist=TRUE` argument to the syntax that specifies the working model. Note that this argument should be added to the syntax when RVE is not applied to a working model. Without this argument, test statistics of individual model coefficients are based on a standard normal distribution, and the omnibus test is based on a chi-square distribution which both do not control for the type-1 error rate adequately (Van den Noortgate et al., 2015). In contrast, adding the `tdist=TRUE` argument invokes the t-distribution for test statistics of individual coefficients and an F-distribution for the omnibus test, which slightly mimics the Knapp and Hartung (2003) method. Further, the `dfs="contain"` argument may be added to the syntax, resulting in better approximations of the degrees of freedoms of the t- and F-distributions (Viechtbauer, 2023b, pp. 270, 166). The `cluster=` argument in the `robust` function specifies the clustering variable (i.e., studyID) to construct the sandwich estimator. The `clubSandwich=TRUE` argument implies that the clubSandwich package is invoked to perform cluster-robust tests and to retrieve cluster-robust confidence intervals. With this argument, the recommended CR2 estimator is applied that estimates the variance-covariance matrix using the bias-reduced linearization adjustment (Bell & McCaffry, 2002; Tipton, 2015; Tipton & Pustejovsky, 2015). Moreover, the degrees of freedom of the t-tests are then estimated with a Satterthwaite correction and the F-test is based on the approximate Hotelling's T^2 distribution (Pustejovsky, 2023), which are meant to improve the estimation of coefficients.

Running the syntax in Listing 5 and Listing 6 will produce the outputs as shown in Outputs 3 and 4, respectively. As can be seen in those outputs, the estimates of the overall effect remained unchanged, but the RVE method applies an adjustment to the tests and confidence intervals resulting in different degrees of freedom, larger standard errors and p-values, wider confidence intervals, and smaller t-values.

More examples are available online (Viechtbauer, 2021), including instructions to construct a variance-covariance matrix with two "common" correlations instead of just one, so that associations between within-study outcomes and thus effect size covariances may be captured more realistically. Specifically, Viechtbauer shows how a matrix can be build using $\rho = .70$ for effect sizes that refer to the same type of delinquency and $\rho = .50$ for effect sizes that refer to different types of delinquency. Note that in the multivariate meta-analytic approach, variance components and potential moderating variables can be tested in the same way as we described in our tutorial (Assink & Wibbelink, 2016), and this is illustrated by Viechtbauer (2021). Readers interested in applying RVE with CWB as small sample adjustment to a model specified with the metafor package can also be referred to example R code online (see Joshi et al., 2023). In our working example, we did not apply CWB as it is especially recommended for multiple-contrast hypothesis tests (Joshi et al., 2022), which was not the focus in our example.

Comparing Results of the Different Approaches

Not surprisingly, when the same effect size data are synthesized using different analytic strategies, the results will be

Listing 1 ∎ Importing the Dataset into the R Environment.

```
# Importing data saved in a comma separated values (CSV) file;
# The data file to be imported was named "dataset.csv";
# All data saved in the file "dataset.csv" is read by invoking
# the read.csv function and assigned to a newly created object
# "dataset" by the assignment operator "<-".
dataset <- read.csv("dataset.csv")
```

Listing 2 ∎ Estimating the Overall Effect Based on a Multivariate Meta-Analytic Approach with an Approximated Variance-Covariance Matrix of $\rho = .60$.

```
# Calculating the approximated variance-covariance matrix by
# assuming a correlation between effect sizes within studies of rho=.60.
vcmatrix <- vcalc(v, cluster=studyID, obs=effectsizeID, data=dataset, rho=0.6)
# Estimating the overall effect based on the multivariate
# approach with an approximated variance-covariance matrix.
overallmultivariate <- rma.mv(y, vcmatrix, random=list(~ 1 | effectsizeID, ~ 1 |
    studyID), tdist=TRUE, data=dataset)
# Request a print of the results stored in the object
# "overallmultivariate" in three digits.
summary(overallmultivariate, digits=3)
```

different. When our example dataset (Assink & Wibbelink, 2016) is analyzed using the different methods that we have described, we see that the overall effect (d) equals 0.427 in the three-level model ($p < .001$, $SE = 0.118$, 95% CI [0.195; 0.659]), that d equals 0.368 in the model based on the multivariate approach with one assumed "common" correlation of $\rho = .60$ for within-study outcomes ($p < .001$, $SE = 0.097$, 95% CI [0.176; 0.559]), that d equals 0.385 ($p < .001$, $SE = 0.102$, 95% CI [0.182; 0.587]) and 0.354 ($p < .001$, $SE = 0.093$, 95% CI [0.170; 0.538]) in the sensitivity analyses for the multivariate approach using a "common" correlation of $\rho = .40$ and $\rho = .60$, respectively, and finally, that d equals 0.427 ($p = .003$, $SE = 0.119$, 95% CI [0.175; 0.679]) and 0.368 ($p = .002$, $SE = 0.097$, 95% CI [0.160; 0.575]) when RVE with a small sample adjustment is applied to both working models (i.e., the three-level model and the model based on the multivariate approach with a "common" correlation of $\rho = .60$, respectively). In this example, we do not regard the differences in estimates of the overall effect size and its precision meaningful for clinical practice. However, this may be different in other meta-analyses where applying alternative modeling techniques may lead to different conclusions. Further, and not reported here for brevity, different modeling techniques may also lead to variations in results of moderator tests and variance component tests. This implies that for instance clinical professionals or policy makers can be informed quite differently depending on the choices of a

meta-analyst regarding the modeling strategy.

Discussion

We conclude this research note with a relevant question: Which modeling strategy to synthesize dependent effect size data is optimal? Formulating a satisfying and solid answer to this question is however difficult and beyond the scope of this paper. For multivariate effect size data the multilevel method makes assumptions that may not hold in some contexts, such as the assumption of independency of standard errors and the assumption that heterogeneity in effect sizes is the same for all outcomes in the synthesis. Despite these assumptions, the available simulation studies (e.g., Moeyaert et al., 2017; Van den Noortgate et al., 2015, 2013) reveal that the multilevel method is robust to misspecification of the correlation structure. However, it should be acknowledged that these studies are limited in both the complexity of the tested models and the conditions in which the models were tested.

On the other hand, the assumptions inherent in a multivariate meta-analysis seem more appropriate in a synthesis of multivariate effect size data, as both the effect size dependency and standard error dependency are explicitly modeled. While a three-level meta-analysis can account for dependencies in effect sizes within and between studies, it might not address the standard error dependency as comprehensively as the multivariate meta-analysis. But as discussed in this paper, a major practical drawback to the

Output 1 ■ Output of Listing 2.

```
Multivariate Meta-Analysis Model (k = 100; method: REML)
logLik     Deviance  AIC      BIC      AICc
-72.765    145.530   151.530  159.316  151.783


Variance Components:
           estim    sqrt    nlvls fixed  factor
sigma^2.1  0.155    0.393   100   no     effectsizeID
sigma^2.2  0.081    0.284   17    no     studyID

Test for Heterogeneity:
Q(df = 99) = 745.161, p-val < .001

Model Results:
estimate  se     tval   df pval   ci.lb  ci.ub
0.368     0.097  3.810  99 <.001  0.176  0.559 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Listing 3 ■ Sensitivity Analyses with Approximated Variance-Covariance Matrices with $\rho = .40$ and $\rho = .80$.

```
# Sensitivity analyses by calculating approximated variance-
# covariance matrices with correlations between effect sizes
# within studies of rho=.40 and rho=.80.
vcmatrix0.4 <- vcalc(v, cluster=studyID, obs=effectsizeID, data=dataset, rho=0.4)
vcmatrix0.8 <- vcalc(v, cluster=studyID, obs=effectsizeID, data=dataset, rho=0.8)
# Estimating the overall effects based on the multivariate
# approach with approximated variance-covariance matrices of
# rho=.40 and rho=.80.
overallmultivariate0.4 <- rma.mv(y, vcmatrix0.4, random=list(~ 1 | effectsizeID, ~
    1 | studyID), tdist=TRUE, data=dataset)
overallmultivariate0.8 <- rma.mv(y, vcmatrix0.8, random=list(~ 1 | effectsizeID, ~
    1 | studyID), tdist=TRUE, data=dataset)
# Request a print of the results stored in the objects
# "overallmultivariate0.4" and "overallmultivariate0.8" in three digits.
summary(overallmultivariate0.4, digits=3)
summary(overallmultivariate0.8, digits=3)
```

Listing 4 ■ Installing and Loading the clubSandwich Package.

```
# Installing and loading the clubSandwich package.
install.packages("clubSandwich")
library(clubSandwich)
```

Listing 5 ■ Estimating the Overall Effect Using a Multivariate Working Model with an Approximated Variance-Covariance Matrix Based on rho=0.6 and the RVE Method.

```
# Applying the RVE method to a multivariate working model with an
# approximated variance-covariance matrix of rho=0.60.
overallmultivariateRVE <- robust(overallmultivariate, cluster=studyID,
    clubSandwich=TRUE)
# Request a print of the results stored in the object
# "overallmultivariateRVE" in three digits.
summary(overallmultivariateRVE, digits=3)
```

Output 2 ∎ Output of Listing 3.

```
Multivariate Meta-Analysis Model (k = 100; method: REML)
logLik  Deviance  AIC       BIC       AICc
-72.797 145.593   151.593   159.379   151.846


Variance Components:
          estim   sqrt    nlvls   fixed   factor
sigma^2.1 0.138   0.371   100     no      effectsizeID
sigma^2.2 0.108   0.329   17      no      studyID


Test for Heterogeneity:
Q(df = 99) = 550.476, p-val < .001


Model Results:
estimate  se      tval    df   pval   ci.lb   ci.ub
0.385     0.102   3.768   99   <.001  0.182   0.587  ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multivariate Meta-Analysis Model (k = 100; method: REML)
logLik  Deviance  AIC       BIC       AICc
-73.021 146.042   152.042   159.827   152.294


Variance Components:
          estim   sqrt    nlvls   fixed   factor
sigma^2.1 0.175   0.418   100     no      effectsizeID
sigma^2.2 0.060   0.245   17      no      studyID


Test for Heterogeneity:
Q(df = 99) = 1380.140, p-val < .001


Model Results:
estimate  se      tval    df   pval   ci.lb   ci.ub
0.354     0.093   3.819   99   <.001  0.170   0.538  ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

multivariate meta-analysis is that the within-study correlations required for constructing the variance-covariance matrix are often not reported in primary studies and thus unknown. To deal with this problem, an approximated variance-covariance matrix using an (informed) estimate of the correlation between effect sizes can be constructed. However, the chosen correlation may deviate substantially from the true correlation between effect sizes. Alternatively, the RVE method can be applied to a working model that guards against misspecification of that model. A meta-analyst may perform sensitivity analyses to deal with the problem of approximating a variance-covariance matrix, but when results differ across these analyses, drawing valid conclusions becomes difficult.

From our position, we cannot state which modeling approach is better than the other, as the choice of modeling should be based on the particular scope and aims of a meta-analysis, the structure of the effect size data at hand, the knowledge one has of the correlation structure, and what can be assumed about the effect size distributions for the different (within-study) outcomes that are to

be synthesized. We realize that not providing a straightforward answer to the question above may not be satisfying for pragmatic researchers searching for the "right" modeling approach. However, Viechtbauer (2023b, pp. 166-167) and Pustejovsky and Tipton (2022) have formulated a "general workflow" and a decision-tree for selecting a working model, respectively, that researchers may find useful in deciding on their modeling strategy.

In the current paper and our prior tutorial (Assink & Wibbelink, 2016), we discussed a rather straightforward dependency structure of the effect size data (i.e., effect sizes nested within studies). However, as Fernández-Castilla et al. (2020) have pointed out, more complex and sophisticated models might be more appropriate to synthesize effect sizes depending on the structure of the effect size dataset at hand. For instance, participants may be nested within outcomes that may be nested within studies that may be nested in research groups. In this example, synthesizing the effect sizes in a four-level rather than a three-level model may better capture the effect size dependency and may therefore be the preferred approach. A differ-

Listing 6 ∎ Estimating the Overall Effect Using a Three-Level Working Model and the RVE method.

```
# Building a three-level meta-analytic model.
overallmultilevel <- rma.mv(y, v, random=list(~ 1 | effectsizeID, ~ 1 | studyID),
    data=dataset)
# Applying the RVE method to the three-level working model.
overallmultilevelRVE <- robust(overallmultilevel, cluster=studyID,
    clubSandwich=TRUE)
# Request a print of the results stored in the object
# "overallmultilevelRVE" in three digits.
summary(overallmultilevelRVE, digits=3)
```

Output 3 ∎ Output of Listing 5.

```
Multivariate Meta-Analysis Model (k = 100; method: REML)
logLik   Deviance  AIC       BIC       AICc
-72.765  145.530   151.530   159.316   151.783

Variance Components:
            estim  sqrt   nlvls  fixed  factor
sigma^2.1 0.155  0.393  100    no      effectsizeID
sigma^2.2 0.081  0.284  17     no      studyID

Test for Heterogeneity:
Q(df = 99) = 745.161, p-val < .001

Number of estimates: 100
Number of clusters: 17
Estimates per cluster: 1-22 (mean: 5.88, median: 5)

Model Results:
estimate  ¹se    ¹tval  ¹df    ¹pval  ¹ci.lb  ¹ci.ub
0.368     0.097  3.792  14.53  0.002  0.160   0.575 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1) results based on cluster-robust inference (var-cov estimator: CR2, approx t-test and confidence
    interval, df: Satterthwaite approx)
```

ent example that may require a more complex approach to model dependency in effect size data is when multiple instruments are used across studies to assess a certain outcome, in addition to primary studies reporting on multiple effect sizes (Fernández-Castilla et al., 2019). In this scenario, effect sizes are nested within studies, but at the same time effect sizes are assessed with a specific instrument that may have been used across studies. The type of instrument now serves as a "crossed random effect" that is sometimes referred to as a "cross-classified random effect". In this cross-classified model, effect sizes can vary because of sampling variability, within-study variability, and both between-study and between-instrument variability. We refer the reader to the work of Fernández-Castilla et al. (2019) for a more detailed explanation of cross-classified random effects.

Further, several modeling recommendations for es-

timating outcome-specific effects in meta-analyses with dependent effect size data have been formulated by Fernández-Castilla et al. (2021), and there is one that we like to highlight here. The three-level model that we illustrated in our tutorial (Assink & Wibbelink, 2016) is often used to estimate a mean effect size for specific outcomes next to an overall effect size. For instance, Spruit et al. (2016) synthesized primary studies examining the effects of physical activity interventions on internalizing behaviors in adolescents using the three-level model. Their aim was to obtain one overall intervention effect as well as several outcome-specific effects for types of internalizing behaviors (i.e., depression, anxiety, and "other" internalizing problems). In some of the included studies, multiple effect sizes referred to the same type of internalizing behavior, for instance because the sample was repeatedly measured over time or because the same outcome was assessed with dif-

Output 4 ■ Output of Listing 6.

```
Multivariate Meta-Analysis Model (k = 100; method: REML)
logLik   Deviance  AIC      BIC      AICc
-73.632  147.264   153.264  161.050  153.517

Variance Components:
          estim  sqrt   nlvls  fixed  factor
sigma^2.1 0.112  0.335  100    no     effectsizeID
sigma^2.2 0.188  0.433  17     no     studyID

Test for Heterogeneity:
Q(df = 99) = 808.848, p-val < .001

Number of estimates: 100
Number of clusters: 17
Estimates per cluster: 1-22 (mean: 5.88, median: 5)

Model Results:
estimate  ¹se    ¹tval  ¹df    ¹pval  ¹ci.lb  ¹ci.ub
0.427     0.119  3.597  15.45  0.003  0.175   0.679 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1) results based on cluster-robust inference (var-cov estimator: CR2, approx t-test and confidence
   interval, df: Satterthwaite approx)
```

ferent instruments. In this particular modeling condition, Fernández-Castilla et al. (2021) found that using the three-level meta-analytic model for estimating outcome-specific effects by means of a moderator analysis leads to underestimated standard errors. However, they also found that standard errors are properly estimated when the three-level model is combined with the RVE technique. Based on these results, the authors recommend using the three-level approach with RVE (and the small sample adjustment to RVE; Tipton, 2015) to obtain robust standard errors and appropriate confidence intervals when researchers are interested in outcome-specific effects. Meta-analysts may find this recommendation useful for their work (see Fernández-Castilla et al., 2021, for more detail).

The study of Fernández-Castilla et al. (2021) is only one example of the available work on the performance of different modeling approaches to meta-analysis of dependent outcomes. Other scholars have also compared the performance of different techniques for modeling effect size and sampling error dependency (see, for instance, Fernández-Castilla et al., 2019; Hedges et al., 2010; Moeyaert et al., 2017; Park & Beretvas, 2019; Tipton, 2013, 2015; Van den Noortgate et al., 2013, 2015). Nevertheless, future simulation studies are required to further explore how these techniques perform in more complex models and under a broader range of conditions, so that more solid modeling advice can be offered to meta-analysts.

We conclude with three final notes. First, we highlight that statisticians and methodologists specialized in meta-analysis may not use the same terminology and definitions

in their work. For example, in this paper we have referred to the "multivariate meta-analytic approach with an approximated variance-covariance matrix" as one of the alternatives to a three-level meta-analysis. However, this approach corresponds to what Pustejovsky and Tipton (2022) refer to as the CHE model in which within-study and between-study heterogeneity in true effect sizes as well as correlated sampling errors are modeled by constructing a multilevel meta-analytic model with a common correlation between pairs of effect sizes (p. 429). These differences in terminology further complicate the already challenging task of navigating the technical literature on dependency in effect size data. Second, we stress the importance of the availability of tutorials, guidelines, and workflows that support the applied meta-analyst in choosing a meta-analytical strategy that fits the nature of dependency in an effect size dataset. Because of the technical nature of literature on dependency in effect size data and techniques to handle this dependency, there are not many sources readily employable by non-technical researchers who aim to conduct a meta-analysis. With our tutorial (Assink & Wibbelink, 2016) and the current paper we have tried to provide a bridge between the technical and often complex literature and the applied meta-analyst. However, there is much more to say about modeling dependency in effect size data than what we have summarized in our prior tutorial and the current research note. Great efforts have also been made by other scholars to support the practical meta-analyst, for instance by providing tutorials on conducting meta-analyses with dependent effect size data (e.g.,
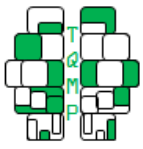
Lu, 2023; Tanner-Smith et al., 2016) and by refining syntax and software (e.g., continuous updates of the metafor package is presented online by Viechtbauer, 2023a). However, we urge researchers to make the technical literature further accessible to interested applied researchers in different scientific fields. Finally, we encourage practical meta-analysts to stay informed on advances in the rapidly evolving meta-analytic techniques and software.

Authors' note

Both authors contributed equally to this manuscript and share first authorship.

References

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. doi: 10.20982/tqmp.12.3.p154.

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & E. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 499–525). Academic Press.

Bell, R. M., & McCaffry, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*(2), 169–181.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Multiple outcomes or time-points within a study. In M. Borenstein, L. V. Hedges, J. P. T. Higgins, & H. R. Rothstein (Eds.), *Introduction to meta-analysis* (pp. 225–238). John Wiley.

Boterhoven De Haan, K. L. B., Lee, C. W., Fassbinder, E., Van Es, S. M., Menninga, S., Meewisse, M. L., ..., & Arntz, A. (2020). Imagery rescripting and eye movement desensitisation and reprocessing as treatment for adults with post-traumatic stress disorder from childhood trauma: Randomised clinical trial. *The British Journal of Psychiatry*, *217*(5), 609–615. doi: 10.1192/bjp.2020.158.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427. doi: 10.1162/rest.90.3.414.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. doi: 10.1037/a0032968.

Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. John Wiley & Sons. doi: 10.1093/acrefore/9780190224851.013.225.

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsy-chology review*, *29*(4), 387–396. doi: 10.1007/s11065-019-09415-6.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2010). *The handbook of research synthesis and meta-analysis* (4th ed.). Russell Sage Foundation.

DeGeest, D., & Schmidt, F. L. (2011). The impact of research synthesis methods on industrial-organizational psychology: The road from pessimism to optimism about cumulative knowledge. *Research Synthesis Methods*, *1*(3-4), 185–197. doi: 10.1002/jrsm.22.

Egger, M., Smith, G. D., & Altman, D. G. (2001). *Systematic reviews in health care: Meta-analysis in context* (2nd ed.). John Wiley & Sons.

Fernández-Castilla, B., Aloe, A. M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Estimating outcome-specific effects in meta-analyses of multiple outcomes: A simulation study. *Behavior Research Methods*, *53*, 702–717. doi: 10.3758/s13428-020-01459-4.

Fernández-Castilla, B., Jamshidi, L., Declerq, L., Beretvas, S. N., Inghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*, 2031–2052. doi: 10.3758/s13428-020-01373-9.

Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behavior Research Methods*, *51*, 1286–1304. doi: 10.3758/s13428-018-1063-2.

Fisher, Z., & Tipton, E. (2015). *An R-package for robust variance estimation in meta-analysis*. arXiv. doi: 10.48550/arXiv.1503.02220.

Geeraert, L., Van den Noortgate, W., Grietens, H., & Onghena, P. (2004). The effects of early prevention programs for families with young children at risk for physical child abuse and neglect: A meta-analysis. *Child Maltreatment*, *9*(3), 277–291. doi: 10.1177/1077559504264265.

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

Hedges, L. V. (2009). Statistical considerations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 37–46). Russell Sage Foundation.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with depen-

dent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. doi: 10.1002/jrsm.5.

Hong, C., Riley, R. D., & Chen, Y. (2018). An improved method for bivariate meta-analysis when within-study correlations are unknown. *Research Synthesis Methods*, *9*(1), 73–88. doi: 10.1002/jrsm.1274.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Hutchinson, B. T., Pammer, K., Bandara, K., & Jack, B. N. (2022). A tale of two theories: A meta-analysis of the attention set and load theories of inattentional blindness. *Psychological Bulletin*, *148*(5-6), 370–396. doi: 10.1037/bul0000371.

Joshi, M., Pustejovsky, J. E., & Beretvas, S. N. (2022). Cluster wild bootstrapping to handle dependent effect sizes in meta-analysis with a small number of studies. *Research Synthesis Methods*, *13*(4), 457–477. doi: 10.1002/jrsm.1554.

Joshi, M., Pustejovsky, J. E., & Cappelli, O. (2023). *Cluster wild bootstrapping for meta-analysis*. https://cran.r-project.org/web/packages/wildmeta/vignettes/cwbmeta.html

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, *1*(3), 227–235. doi: 10.1037/1082-989X.1.3.227.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*, 2693–2710. doi: 10.1002/sim.1482.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. doi: 10.1002/jrsm.35.

Li, L. Y., Dent, A. L., Donberg, R., & Martin, E. A. (2022). Trait and state affective experience among high-risk people in the schizophrenia spectrum: A meta-analytic review. *Psychological Bulletin*, *148*(9-10), 741–763. doi: 10.1037/bul0000380.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage publications.

Lu, M. (2023). Computing within-study covariances, data visualization, and missing data solutions for multivariate meta-analysis with metavcov. *Frontiers in Psychology*, *14*, 11850121–9. doi: 10.3389/fpsyg.2023.1185012.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. In D. A. Belsley & E. J. Kontoghiorghes (Eds.), *Handbook of computational econometrics* (pp. 183–214). John Wiley & Sons.

Merson, B., Pezdek, K., & Saywitz, K. (2016). A meta-analysis of children's self-reports of dietary intake. *Psychology & Health*, *32*(2), 186–209. doi: 10.1080/08870446.2016.1250274.

Moeyaert, M., Ugille, M., Beretvas, S. N., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, *20*(6), 559–572. doi: 10.1080/13645579.2016.1252189.

Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., & Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. *Environmental Evidence*, *12*, 8–8. doi: 10.1186/s13750-023-00301-6.

Oliveira, C. M., Henderson, L., & Hayiou-Thomas, M. E. (2022). *Limited evidence of an association between language, literacy and procedural learning in typical and atypical development: A meta-analysis*. PsyArXiv. doi: 10.31234/osf.io/ndtg6.

Park, S., & Beretvas, S. N. (2019). Synthesizing effects for multiple outcomes per study using robust variance estimation versus the three-level model. *Behavior Research Methods*, *51*, 152–171. doi: 10.3758/s13428-018-1156-y.

Price, M. J., Blake, H. A., Kenyon, S., White, I. R., Jackson, D., Kirkham, J. J., Neilson, J. P., Deeks, J. J., & Riley, R. D. (2019). Empirical comparison of univariate and multivariate meta-analyses in cochrane pregnancy and childbirth reviews with multiple binary outcomes. *Research Synthesis Methods*, *10*(3), 440–451. doi: 10.1002/jrsm.1353.

Pustejovsky, J. E. (2023). *Cluster-robust (sandwich) variance estimators with small-sample corrections*. https://cran.r-project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. doi: 10.1007/s11121-021-01246-3.

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

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical bayes meta-analysis. *Journal of Educational Statistics*, *10*, 75–98. doi: 10.3102/10769986010002075.

Riley, R. D. (2009). Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *172*(4), 789–811. doi: 10.1111/j.1467-985X.2008.00593.x.

Romans, S. M., Roeltgen, D. P., Kushner, H., & Ross, J. L. (1997). Executive function in girls with turner's syndrome. *Developmental Neuropsychology*, *13*, 23–40. doi: 10.1080/87565649709540666.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Sage Publications.

Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and applied multilevel analysis* (2nd ed.). Sage Publications.

Spruit, A., Assink, M., van Vugt, E., van der Put, C., & Stams, G. J. (2016). The effects of physical activity interventions on psychosocial outcomes in adolescents: A meta-analytic review. *Clinical Psychology Review*, *45*, 56–71. doi: 10.1016/j.cpr.2016.03.006.

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

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods*, *4*, 169–187. doi: 10.1002/jrsm.1070.

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

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*(6), 604–634. doi: 10.3102/1076998615606099.

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. doi: 10.3758/s13428-012-0261-6.

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. doi: 10.3758/s13428-014-0527-2.

Van den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*, 765–790. doi: 10.1177/0013164403251027.

Viechtbauer, W. (2017). *Imputing covariance matrices for meta-analysis of dependent effects*. R-sig-meta-analysis – R Special Interest Group for Meta-Analysis. https://stat.ethz.ch/pipermail/r-sig-meta-analysis/2017-August/000103.html

Viechtbauer, W. (2020). *Computing var-covariance matrix with correlations of six non-independent outcomes*. R-sig-meta-analysis – R Special Interest Group for Meta-Analysis. https://stat.ethz.ch/pipermail/r-sig-meta-analysis/2020-July/002258.html

Viechtbauer, W. (2021). *Studies on the association between recidivism and mental health*. Retrieved May 3, 2024, from https://wviechtb.github.io/metadat/reference/dat.assink2016.html

Viechtbauer, W. (2022). *The metafor package: A meta-analysis package for R*. https://www.metafor-project.org/doku.php/analyses:konstantopoulos2011

Viechtbauer, W. (2023a). *The metafor package: A meta-analysis package for R [package news]*. Retrieved March 5, 2024, from https://www.metafor-project.org/doku.php/news:news

Viechtbauer, W. (2023b). *Package 'metafor'* (Version 4.4.0). https://cloud.r-project.org/web/packages/metafor/metafor.pdf

Wibbelink, C. J. M., Hoeve, M., Stams, G. J. J. M., & Oort, F. J. (2017). A meta-analysis of the association between mental disorders and juvenile recidivism. *Aggression and Violent Behavior*, *33*, 78–90. doi: 10.1016/j.avb.2017.01.005.

Yang, Y., Macleod, M., Pan, J., Lagisz, M., & Nakagawa, S. (2023). Advanced methods and implementations for the meta-analyses of animal models: Current practices and future recommendations. *Neuroscience & Biobehavioral Reviews*, *146*, 105016–1. doi: 10.1016/j.neubiorev.2022.105016.

Citation