University of Amsterdam

UvA-DARE (Digital Academic Repository)

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

Assink, M.; Wibbelink, C.J.M.

DOI
10.31234/osf.io/js4nk

Publication date
2023

Document Version
Final published version

Link to publication

Citation for published version (APA):
Assink, M., & Wibbelink, C. J. M. (2023). *Addressing Dependency in Meta-Analysis: A Companion to Assink and Wibbelink (2016)*. PsyArXiv. https://doi.org/10.31234/osf.io/js4nk

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:03 Oct 2023

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

Mark Assink & Carlijn J. M. Wibbelink

Status: Preprint
Date of release: March 25, 2023

This manuscript is a preprint and has been submitted for publication in a peer-reviewed journal. Citation of this manuscript in APA 7 style is as follows:

Assink, M., & Wibbelink, C. J. M. (2023). *Addressing dependency in meta-analysis: A companion to Assink and Wibbelink (2016)*. PsyArXiv. https://doi.org/10.31234/osf.io/js4nk

Addressing Dependency in Meta-Analysis:

A Companion to Assink and Wibbelink (2016)

Mark Assink[1] and Carlijn J. M. Wibbelink[2]

[1]Research Institute of Child Development and Education, University of Amsterdam

[2]Psychology Research Institute, University of Amsterdam

Author Note

This research note was not externally funded and there are no conflicts of interest to disclose.

This paper can be cited as: Assink, M., & Wibbelink, C. J. M. (2023). *Addressing dependency in meta-analysis: A companion to Assink and Wibbelink (2016)*. PsyArXiv. https://doi.org/10.31234/osf.io/js4nk

Correspondence concerning this article should be addressed to Mark Assink, Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands. E-mail: M.Assink@UvA.nl

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 brief and non-technical description of effect size and standard error interdependency, how the multilevel and the multivariate approach to meta-analysis handle these types of dependency, and the role robust variance estimation methods can have in capturing dependency in effect size data more accurately. Readers are referred to 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

Addressing Dependency in Meta-Analysis:

A Companion to Assink and Wibbelink (2016)

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 interrelatedness – also referred to as effect size (inter)dependency – may occur, for instance because primary researchers used multiple methods for assessing the same construct, examined the same participants over multiple timepoints, or reported results separately for subscales of an instrument that was used to measure a certain construct. In these examples, the assumption that effect sizes are independent from each other is clearly violated. If a meta-analyst synthesizes interrelated effect sizes using conventional methods for meta-analysis, the results can be incorrect and even misleading (Borenstein et al., 2009).

The question how to deal with interrelated effect sizes 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 interrelated effect sizes in meta-analysis. One of these techniques comprises the synthesis of interrelated effect sizes in a three-level meta-analytic model, which we have illustrated in our tutorial that was published in 2016 (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.

Effect Size Interdependency in Multilevel Meta-Analytic Models

In our tutorial we have demonstrated how effect size interdependency can be modeled in multilevel meta-analytic models by using the *metafor* package of Viechtbauer (2010) in the statistical software environment R (Assink & Wibbelink, 2016). 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, 2010; Konstantopoulos, 2011; Raudenbush & Bryk, 1985). When effect sizes cluster together, it can be assumed that the underlying population (or “true”) effects estimated by the effect sizes 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 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. Modeling effect size interdependency is important, because ignoring dependency often leads to an underestimation of standard errors that in turn results in too narrow confidence intervals, and consequently, in an increased likelihood for falsely rejecting null hypotheses (i.e., increased type-1 error rate; Hedges, 2009; Snijders & Bosker, 2012).

Our tutorial specifically illustrates how effect size interdependency can be modeled in a three-level meta-analytic model that allows for the extraction of multiple effect sizes from individual studies (Assink & Wibbelink, 2016). This three-level approach to meta-analysis has been explained and described by multiple methodological scholars (e.g., Cheung, 2014; Van den Noortgate et al., 2013, 2015). In this model, 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 the model, the within- and between-study 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. This three-level structure is a rather straightforward, but powerful way to model effect size dependency that enables the extraction of multiple effect sizes from individual studies that meet the inclusion criteria of a meta-analysis. Extracting multiple relevant effect sizes rather than one effect size increases not only the statistical power, but also the number of research questions that can be addressed in meta-analytic research (Cheung, 2015).

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 & Onghena, 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 interdependency of effect sizes within and between studies (Fernández-Castilla et al., 2020). The three-level model that we illustrated in our tutorial (Assink & Wibbelink, 2016) is a specific form of the multilevel approach to meta-analysis and has been applied in an increasing number of meta-analyses. However, as Fernández-Castilla and colleagues (2020) have pointed out, more complex and sophisticated models can be more appropriate to synthesize effect sizes depending on the nested structure of a particular effect size dataset. For instance, participants (level 1 – sampling level) may be nested within outcomes (level 2) that may be nested within studies (level 3) that may be nested in research groups (level 4). In this example, synthesizing the effect sizes in a four-level rather than a three-level model may better capture the effect size interdependency and may therefore be a better approach.

A different example asking for a more complex approach to modeling effect size interdependency 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 example, 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 (modeled at level 1), within-study variability (modeled at level 2), and both between-study and between-instrument variability (modeled at level 3). We refer the reader to the work of Fernández-Castilla et al. (2019) for a more detailed explanation of cross-classified random effects. Examples like these illustrate that a three-level meta-analytic model as we described in our tutorial (Assink & Wibbelink, 2016) may not sufficiently capture effect size interdependency. Fernández-Castilla et al. (2019) stress the importance of carefully reflecting on the effect size data structure to determine what multilevel structure is most appropriate for the effect size synthesis.

Effect Size and Sampling Error Interdependency

So, three-level and more complex multilevel models can handle dependency in effect size data, but let's focus some more on this dependency. When multiple effect sizes are extracted from individual studies, there are in fact two different forms of dependency that may arise (Nakagawa et al., 2023). First, a single study may examine two or more *non-overlapping* groups of participants and report an effect size for *each separate* participant group. In this case, the same study produces multiple effect sizes, which were all obtained in the same "study context". Therefore, the produced effect sizes are interdependent. This within-study interrelatedness may occur 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. The consequence of this within-study clustering is that one effect size tells us something about the direction and strength of another effect size from the same (within-study) cluster, and thus there is effect size interdependency.

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). This implies that all or some of the sampled participants contribute to more than one outcome, and that 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 interdependency but also sampling error interdependency. From a multivariate perspective, it is important to explicitly model sampling error interdependency (in addition to effect size interdependency), as not modeling the correlated sampling errors inflates the type-1 error rate, and thus the number of falsely rejected null hypotheses. This error rate can be manifested in not only estimates of model coefficients, but also in estimates of their standard errors. Not accounting for dependency in effect size data may therefore introduce bias in effect size estimates as well as the precision of their estimation (Becker, 2000, p. 503).

Returning to the three-level meta-analytic model that we illustrated in our tutorial (Assink & Wibbelink, 2016), the question arises how this model deals with the two forms of dependency described above. The three-level structure deals with dependency in effect sizes by assigning random effects to outcomes (level 2) and to studies (level 3), but the dependency in effect size sampling error is not *explicitly* modeled. This seems to be at odds with the multivariate approach to meta-analysis prescribing that both effect size interdependency and sampling error interdependency should 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 interdependency and standard error interdependency, but state that the latter is accounted 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 correlation in standard errors is accounted for by the random effects of the three-level model that allow the standard error interdependency to “subsume” into the correlation in the underlying true effects that is modeled by the specification of the hierarchical structure with its random effects. As a result, appropriate estimates of mean effect sizes and standard errors are produced in the analyses (Van den Noortgate et al., 2013). Moreover, the simulation studies performed by Van den Noortgate et al. (2013) illustrate that multilevel meta-analytic models can account for *all* dependency in studies including effect size and standard error interdependency. 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 multilevel structure by modeling random effects is sufficient to deal with both effect size interdependency and sampling error interdependency.

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 multilevel models as we have illustrated in our tutorial (Assink & Wibbelink, 2016) assume that heterogeneity in effect sizes is the same for all outcomes that are synthesized (e.g., Van den Noortgate et al., 2015), which is an assumption that can be difficult to hold (Viechtbauer, 2020). Violating this assumption implies that confidence intervals around estimates for more heterogeneous outcomes are too narrow (due to an underestimated standard error), whereas too wide confidence intervals will 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, the multivariate approach to meta-analysis is to be preferred so that both effect size interdependency and standard error interdependency are explicitly modeled. Unlike the multivariate approach, 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, 1015), there are advocates of the multivariate approach suggesting that multilevel modeling may produce invalid estimates of model coefficients whenever the same or partly the same participants contribute to multiple outcomes in studies. They argue that explicitly modeling both effect size interdependency and sampling error interdependency by applying the multivariate approach to meta-analysis yields more appropriate estimates of model coefficients than only modeling the effect size dependency in multilevel meta-analysis.

Specifying Covariances for Modeling Sampling Error Interdependency

Following the multivariate approach to meta-analysis, how can a meta-analyst explicitly model both effect size and standard error interdependency? The answer lies in constructing an appropriate “variance-covariance matrix” (also referred to as “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 pairwise combinations of the variables. The sampling variance refers to the variation in individual variables, whereas a covariance is an unstandardized correlation representing the linear association between two variables and is a measure of how changes in one variable are associated with changes 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 multilevel meta-analytic model resulting in a multivariate meta-analytic model that accounts for effect size interdependency through the specified random effects and for standard error interdependency 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, as the required correlations needed for those computations are seldomly reported by study authors (e.g., Fernández-Castilla et al., 2020). If, for instance, a study reports multiple effect sizes because multiple outcomes were examined in the study to measure a single 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 interdependency in the multivariate approach to meta-analysis. The multilevel approach, however, does not require that the sampling covariance of the effect size estimates is "known" in advance, as the between-study variance acts as a "stand in" for the sampling covariance (Van den Noortgate et al., 2013). This makes multilevel meta-analysis convenient and appealing, as the lack of information on the covariances 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 multivariate meta-analyses are only seldomly performed.

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 opt for constructing an "approximate" or "working" variance-covariance matrix in which the effect size covariances are calculated using an informed estimate (or "guestimate") of one "common" correlation between the effect sizes. The underlying assumption of such a matrix is that there is a *single* correlation between all pairs of effect sizes that come from the same study, which is the same across all studies. Pustejovsky and Tipton (2022, p. 429) refer to this 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 a single 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 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 (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, for examples). Building on the outlined example above, the meta-analyst could compute additional matrices based on lower (e.g., $r = .2$) and higher (e.g., $r = .8$) constant sampling correlations than the correlation that was initially used in constructing the variance-covariance matrix (i.e., $r = .47$). Next, a sensitivity analysis can be performed separately for each matrix to find out 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 the results hold across different estimates of the within-study correlation(s).

Robust Standard Errors and Confidence Intervals

On top of computing a variance-covariance matrix, the Robust Variance Estimation (RVE) method can be used in the effect size synthesis (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, the RVE method tries to "polish up" the standard errors of fixed effect estimates. 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 needing to model the exact nature of the dependency in effect size data. However, although RVE does not require a specification of the covariance structure, the performance of the RVE method improves when a working model such as a multivariate or multilevel model is specified (Hedges et al., 2010). By first specifying a working model and then applying RVE for obtaining standard errors and hypothesis tests, a meta-analyst benefits from the efficiency of a working model for the effect size data dependency structure while retaining the robustness to potential model misspecification (Hong et al., 2018). Specifically, combining a working model with the RVE technique captures the dependency in effect size data more accurately leading 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 to be synthesized 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 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. 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 the authors 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.

Examples in R

We have now briefly covered how dependency in effect size data is mainly dealt with in the multilevel versus the multivariate approaches to meta-analysis and the role RVE (with small sample adjustment) can fulfill in estimating coefficients in a multilevel or multivariate meta-analytic model. In our tutorial, 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 (Assink & Wibbelink, 2016). We used the standardized mean difference (Cohen's *d*) 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 questions that now rise are (1) how the multivariate approach that explicitly models effect size interdependency and standard error dependency can be applied to the dataset we used in our tutorial, and (2) to what extent the results differ between the multilevel and the multivariate approach. For answering the first question, we refer the reader to the excellent examples of Wolfgang Viechtbauer (Viechtbauer, n.d.) who showed in the R statistical environment how our effect size dataset - which is available online as appendix of Assink and Wibbelink (2016) - can be analyzed with the multivariate approach using his *metafor* package and how the RVE method can be applied.

More specific, Viechtbauer (n.d.) demonstrates in his example R code (1) how an approximate variance-covariance matrix can be constructed under the "constant sampling correlation" assumption (Pustejovsky & Tipton, 2022) by estimating one "common" correlation for within-study effect sizes (r = .6 was chosen); (2) how this approximate matrix can then be fed to the *rma.mv* function together with a specification of the hierarchical structure of the effect size data – that is, outcomes (level 2) nested within studies (level 3) – and (3) how the RVE method with a small sample adjustment can be applied to the specified multivariate model to obtain robust standard errors and confidence intervals. In multivariate models, variance components and potential moderating variables can be tested in the same way as we described in our tutorial

(Assink & Wibbelink, 2016). Viechtbauer also illustrates how an approximate variance-covariance matrix can be based on two instead of one "common" correlation 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 $r = .7$ for effect sizes that refer to the *same* type of delinquency and $r = .5$ for effect sizes that refer to *different* types of delinquency. Readers who are interested in applying RVE with specifically CWB as small sample adjustment to a model specified with the *metafor* package can also find example R code online (see Joshi et al., n.d.)

Naturally, when the same effect size data are synthesized in different analytic strategies, the results will be different. When our example dataset (Assink & Wibbelink, 2016) is analyzed using the approaches that we have described here, we see for instance 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 multivariate model assuming a "common" correlation of $r = .6$ for within-study outcomes ($p < .001$, SE = 0.097, 95% CI [0.179; 0.557]), that d equals 0.362 ($p < .001$, SE = 0.093, 95% CI [0.179; 0.545]) in the multivariate model assuming that $r = .7$ for the same types of delinquency and $r = .5$ for different types of delinquency, and finally, that d equals 0.362 ($p = .002$, SE = 0.094, 95% CI [0.161; 0.563] when RVE with small sample adjustment is applied to the latter multivariate model (see Viechtbauer, n.d., for more results). In this particular 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 effect size data where applying different modeling techniques may lead to different conclusions. Further, and not reported here for brevity reasons, different modeling techniques may also lead to differences in results of moderator tests and variance component tests. These differences may be negligible as is the case for the overall effect estimates above, but can also be substantial. 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.

Conclusion

We conclude this short research note with a relevant question: Which modeling approach to synthesizing dependent effect sizes is best, the multivariate or the multilevel approach? Formulating a satisfying and solid answer to this question is however difficult and beyond the scope of this paper. As we have illustrated, the multilevel approach makes assumptions that may be too simplistic for reality, such as the assumption of independency of standard errors and the assumption that heterogeneity in effect sizes is the same for all outcomes that are synthesized. Despite these assumptions, the available simulation studies (e.g., Moeyaert et al., 2017; Van den Noortgate et al., 2013, 2015) reveal that the multilevel approach is robust to misspecification of the correlation structure. However, these studies are limited in the complexity of the models and the conditions that were examined. The multivariate approach on the other hand is fundamentally based on more realistic assumptions, and explicitly models both effect size interdependency and standard error interdependency. A drawback to this approach, however, is that the within-study correlations that are needed for constructing the variance-covariance matrix are often unknown and must be estimated or guessed (or "guestimated"). A meta-analyst may perform sensitivity analyses to deal with this problem, 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 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 (n.d.) 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. Further, several modeling recommendations for estimating *outcome-specific effects* in meta-analyses of interdependent effect sizes 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 specific 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 different 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, 2018; 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.

As a final note, we stress the importance of the availability of tutorials, guidelines, and workflows that support the applied meta-analyst. Because of the technical nature of the literature on dependency in effect size data and different 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 tried to provide a bridge between the technical and often complex literature and the applied meta-analyst, but there is much more to say about modeling effect size dependency than what we have summarized in our tutorial and this research note. Great efforts have been made by scholars to support the practical meta-analyst, but we urge researchers to make the technical literature further accessible to interested applied researchers in different scientific fields.

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. http://dx.doi.org/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.

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. https://doi.org/10.1002/9780470743386.ch24

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. https://doi.org/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. https://doi.org/10.1037/a0032968

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

Cooper H, Hedges L. V., & Valentine J. C. (2010). *The handbook of research synthesis and meta-analysis* (4th edition). 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. https://doi.org/10.1002/jrsm.22

Egger M, Smith G. D., & Altman D. G. (2001) *Systematic reviews in health care: Meta-analysis in context* (2nd edition). 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. https://doi.org/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. https://doi.org/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. https://doi.org/10.3758/s13428-018-1063-2

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

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. (Ed.). (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 dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65. https://doi.org/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. https://doi.org/10.1002/jrsm.1274

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Lawrence Erlbaum Associates.

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. https://doi.org/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. https://doi.org/10.1002/jrsm.1554

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

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*(1), 61–76. https://doi.org/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. https://doi.org/10.1037/bul0000380

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

Merson, B., Pezdek, K., & Saywitz, K. (2016). A meta-analysis of children's self-reports of dietary intake. *Psychology & Health, 32*(2), 186–209. https://doi.org/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. https://doi.org/10.1080/13645579.2016.1252189

Nakagawa1, 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.* EcoEvoRxiv. https://doi.org/10.32942/X2KK5N

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. https://doi.org/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. https://doi.org/10.3758/s13428-018-1156-y

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

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

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. https://doi.org/10.1016/j.cpr.2016.03.006

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods, 4,* 169–187. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1177/0013164403251027

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2017, August 10). *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, July 8). *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. (n.d.). *Some recommended practices.* https://wviechtb.github.io/metafor/reference/misc-recs.html

Viechtbauer, W. (n.d.). *Studies on the association between recidivism and mental health.* https://wviechtb.github.io/metadat/reference/dat.assink2016.html