Meta-regression with Dependent Effect Size Estimates

Larry V Hedges

Northwestern University, USA

Presented at the Joint Cochrane Campbell Colloquium, Keystone Colorado, October 2010

Sources of Dependence in Meta-analysis

Let $T_i = \theta_i + \varepsilon_i$, $i = 1, \ldots, k$

where T_i is the effect size estimate, θ_i is the effect size parameter, and ε_i is the estimation error of the i^{th} effect

Dependence arises in either of two places:

Because the ε_i are correlated

OR

Because the θ_i are correlated

Or both

Dependence Due to Correlated Estimation Errors (correlated ε_i 's)

This form of dependence is familiar in connection with

Multiple estimates from the same sample

- Multiple followups
- Multiple outcome constructs
- Multiple measures of the same outcome construct

Partially overlapping samples (e.g., in multi-armed studies)

- Shared control groups
- Shared treatment groups

I will deal mostly with this form of dependence here

Dependence Due to Correlated Effect Size Parameters (correlated θ_i's)

This form of dependence is familiar in connection with

Higher order clustering of studies

- Studies from the same laboratory
- Studies from the same investigator or team
- Studies reported in the same paper (but using different individuals)
- Groups of coordinated studies

This form of dependency is not often modeled in some areas, but is of great interest in others

Why Do We Care About Dependence?

We care about dependence because it complicates estimation

It complicates computing uncertainty of estimates

It complicates getting efficient estimates

I argue that often (but perhaps not always) the biggest problem posed by dependence is proper estimation of *uncertainty* of estimates

How Do We Deal with Dependence?

- Ignore it without contrivance (often a bad idea, but sometimes acceptable)
- Contrive to ignore it by creating a single synthetic effect size per sample
- Model dependence with full multivariate analysis
- Use new robust methods that estimate empirical standard errors

Full Multivariate Meta-analysis

Full Multivariate Meta-analysis Notation

Assume k studies, each with a vector of effect sizes

Effect Size Parameters	$\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$
Effect Size Estimates	$\mathbf{T}_1, \ldots, \mathbf{T}_m$
Estimation Error Variances	$\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m$
Study Level Covariates	$x_1, \ldots, x_m$
Distributional Assumptions	$\mathbf{T}_j \mid \boldsymbol{\theta}_j \sim \mathrm{N}(\boldsymbol{\theta}_j, \boldsymbol{\Sigma}_j)$

Full multivariate analysis is always possible in theory and is optimal in theory

Key references are:

Hedges & Olkin (1985) *Statistical methods for meta-analysis*

Raudenbush, Becker, & Kalaian (1988) *Psychological Methods*

Kalaian & Raudenbush (1996) *Psychological Methods*

But all of these methods require that the covariance matrix of estimation errors is known

Major Problem

Covariance structure of estimation errors is almost always **unknown**

Sometimes we have *estimates* of this structure, but these are often subject to large sampling errors (of the same order as the effect sizes themselves)

Frequently, we don't even have empirical estimates from the studies in question and we have to impute (guess) based on other sources of information

In large datasets, this is a very tedious process

Robust Variance Estimation

One potential solution comes from work on robust variance estimation in the general linear model

Work in this tradition includes

- Eichler (1967) *5th Berkeley Symposium*
- Huber (1967) *5th Berkeley Symposium*
- White (1980) *Econometrica*
- MacKinnon & White (1984) *J of Econometrics*

Robust Variance Estimates (Univariate Version)

The fundamental ideas are easier to understand as a univariate analysis (estimating the mean effect size) with $k_j = 1$ (one effect per study)

The weighted mean effect size estimate (call it b) is

$$b = \left(\sum_{j=1}^{m} w_j \right)^{-1} \left(\sum_{j=1}^{m} w_j T_j \right) = \sum_{j=1}^{m} \left(\frac{w_j}{w_\bullet} \right) T_j$$

Since the T_i are independent, the variance of b is

$$\sum_{j=1}^{m} \left(\frac{w_j}{w_\bullet} \right)^2 \mathrm{Var}\{T_j\}$$

To get a robust variance estimate we need to get an estimate of this quantity. Do that by substituting $e_j = (T_j - b)^2$ for $\mathrm{Var}\{T_j\}$

.

$$v^R = \sum_{j=1}^{m} \left(\frac{w_j}{w_\bullet} \right)^2 \left(T_j - b \right)^2 = \sum_{j=1}^{m} \left(\frac{w_j}{w_\bullet} \right)^2 e_j^2$$

Robust Variance Estimates (Multivariate Version)

We have k estimates (from $m \le k$ studies) $\mathbf{T} = (T_1, \ldots, T_k)'$

The general model (with covariates) is

$$\mathbf{T} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\mathbf{X}$ is a k x p design matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a p x 1 vector of unknown regression coefficients, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)'$ is a k x 1 vector of residuals.

Usually we would estimate $\boldsymbol{\beta}$ and its variance via

$$\mathbf{b} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{T} \qquad \mathrm{V}\{\mathbf{b}\} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

with $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$ (which might include between-study variance components)

The problem is that although the variances are "known" the covariances of $\boldsymbol{\varepsilon}$ in $\boldsymbol{\Sigma}$ contain unknown correlations

Robust Variance Estimates (Multivariate Version)

Change notation slightly so we have $k = k_1 + \cdots + k_m$ estimates $\mathbf{T} = (T_1, \ldots, T_k)'$ from m studies with $k_j \geq 1$ estimates within the j^{th} study

Partition the vectors $\mathbf{T}$ and $\boldsymbol{\varepsilon}$ into m stacked k_j x 1 vectors $\mathbf{T}_1, \ldots, \mathbf{T}_m$ and $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_m$, and partition the design matrix into m stacked k_j x p matrices so that

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_m \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix} \qquad \mathbf{X}_j = \begin{pmatrix} x_{j11} & \cdots & x_{j1p} \\ \vdots & \ddots & \vdots \\ x_{jn_j1} & \cdots & x_{jn_jp} \end{pmatrix}$$

and the model becomes

$$\begin{pmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix} \qquad \boldsymbol{\varepsilon}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$

and $\boldsymbol{\Sigma}_j$ is an k_j x k_j covariance matrix of effects within studies

Robust Variance Estimates (Multivariate Version)

In this new notation, with (known) weight matrix $\mathbf{W}$ = diag($\mathbf{W}_1$, …, $\mathbf{W}_m$) the weighted least squares estimate of $\boldsymbol{\beta}$ is

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

And its exact covariance matrix is

$$\left(\sum_{j=1}^{m} \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j' \mathbf{W}_j \boldsymbol{\Sigma}_j \mathbf{W}_j \mathbf{X}_j \right) \left(\sum_{j=1}^{m} \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right)^{-1}$$

which still depends on unknown correlations that are part of the $\boldsymbol{\Sigma}_i$' s.

We want to get a robust estimate of this variance (one based only on the observed estimates), but it looks like we have to estimate *every* element of the $\boldsymbol{\Sigma}_j$' s

Robust Variance Estimates (Multivariate Version)

Write the variance of $\sqrt{m}\mathbf{b}$ as

$$\left(\frac{1}{m}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{X}_j\right)^{-1}\left(\frac{1}{m}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{\Sigma}_j\mathbf{W}_j\mathbf{X}_j\right)\left(\frac{1}{m}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{X}_j\right)^{-1}$$

This illustrates that we don't have to estimate every element of each $\mathbf{\Sigma}_j$, only estimate

$$\frac{1}{m}\sum_{j=1}^{m}\mathrm{E}\left\{\mathbf{X}_j'\mathbf{W}_j\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'\mathbf{W}_j\mathbf{X}_j\right\}$$

which is quite feasible if m is large enough. So we look for a theorem that is asymptotic in m (the number of studies)

Note that the correlations do not need to be known or explicitly estimated

Robust Variance Estimates (Multivariate Version)

If we take the $\mathbf{X}_j$ and $\mathbf{W}_j$ to be stochastic (as well as the $\mathbf{T}_j$) estimating the variance of $\mathbf{b}$ also requires estimating

$$\left(\frac{1}{m} \sum_{j=1}^{m} \mathrm{E}\left\{ \mathbf{X}_j' \mathbf{W}_j \mathbf{X}_j \right\} \right)^{-1}$$

There is an argument for having the $\mathbf{X}_i$, at least, be stochastic

The random effects model treats studies as sampled, their covariate values are also sampled, and hence random

Few of the analyses we do take this into account

Robust Variance Estimates (Multivariate Version)

The robust variance estimate of $\mathbf{b}$ is

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

where $\mathbf{e}_j = \mathbf{T}_j - \mathbf{X}_j\mathbf{b}$ is the (k_j x 1) (estimated) residual vector in the j^{th} study

The robust test of the hypothesis H_0: $\beta_a = 0$ uses the test statistic

$$t_a^R = b_a \Big/ \sqrt{v_{aa}^R}$$

where $v_{aa}{}^R$ is the a^{th} diagonal element of $\mathbf{V}^R$ We might use the t-distribution with $m-p$ degrees of freedom for critical values

A robust confidence intervals would use $v_{aa}{}^R$ too

$$b_a - c_{\alpha/2}\sqrt{v_{aa}^R} \le \beta_a \le b_a + c_{\alpha/2}\sqrt{v_{aa}^R}$$

The General Theorem

Under regularity conditions as $m \to \infty$, $\mathbf{V}_m{}^R$ is a consistent estimator of the true covariance matrix, and

$$\sqrt{m}\left(m\mathbf{V}_m{}^R\right)^{-1/2}\left(\mathbf{b}_m - \boldsymbol{\beta}\right)\xrightarrow{L}\mathrm{N}\left(\mathbf{0},\ \mathbf{I}_p\right)$$

where ($\mathbf{V}_m{}^R$) is the robust covariance matrix and $\mathbf{b}_m$ is the estimate of $\boldsymbol{\beta}$ computed from m studies

The regularity conditions include that

- The regressors are uniformly bounded
- The weights are uniformly bounded
- The number of estimates per study is bounded
- The first few moments of $\boldsymbol{\varepsilon}_j$ are uniformly bounded

As well as absolute moment conditions on the ($\boldsymbol{\varepsilon}_j{}'\boldsymbol{\varepsilon}_j$)

This Theorem is Remarkably General

The method can be generally used (if m is not too small)

- For unknown correlation structures
- For different kinds of dependency
- Different kinds of effect sizes
- Different numbers of dependent effects per study or per cluster
- Even if the effect sizes are heterogeneous within studies

Notes on the General Theorem

This theorem is asymptotic in m, the number of studies, not the size of each study

There are no distribution assumptions about the effect sizes

Correlations among the effects within studies need not be specified, but presumably impact the standard errors

The $\mathbf{X}_j$ need not be fixed (random $\mathbf{X}_j$ are more consistent with random effects models)

A Special Case: Estimating the Mean Effect

In the case of estimating the overall average effect, the estimator reduces to a weighted mean

$$T_{\bullet} = \sum_{j=1}^{m} \sum_{i=1}^{k_j} w_{ij} T_{ij} \bigg/ \sum_{j=1}^{m} \sum_{i=1}^{k_j} w_{ij}$$

And the robust variance is

$$v^R = \sum_{j=1}^{m} \sum_{a=1}^{k_j} \sum_{b=1}^{k_j} w_{aj} w_{bj} \left(T_{aj} - T_{\bullet}\right)\left(T_{bj} - T_{\bullet}\right) \bigg/ \left(\sum_{i,j} w_{ij} \right)^2$$

Note that this is similar to, but not identical to, the Knapp-Hartung variance estimate

Note that the exact estimate depends on how weights are assigned

Choosing Weights

This procedure does not lead to obvious choices of weights

Instead, it gives us a variance estimate of an effect based on ***any*** set of weights we might choose

Thus we need to think about weights that might be useful in practice

Choosing Weights

Why do we weight in meta-analysis?

- To obtain more efficient estimates
- To help compute variances (incidentally)

Major gains in efficiency come when optimal weights are unequal

Different effects within studies are often based on the same sample size, which (for some effect size measures) implies that they will have similar variances

Therefore it might make sense to equally weight estimates within studies and allow different weights between studies

Choosing Weights

One weighting scheme we might propose is to give equal weight to all estimates in the same study, but total weight based on the average variance, namely

$$w_j = \frac{1}{\bar{v}_{\bullet j} k_j} = \frac{1}{v_{1j} + \mathrm{L} \ + v_{k_j j}}$$

This is similar to a fixed effects weighting scheme and would not be as efficient as a scheme that incorporated between-study variance or within-study covariances, or both

Many other schemes are possible, with different consequences for efficiency

Choosing Weights

Assume a single between-studies variance component (no differential variances for different effects within studies)

$$T_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + \eta_j + \varepsilon_{ij} \qquad\qquad \eta_j \sim \mathrm{N}(0,\ \tau^2)$$

where ε_{ij} and $\varepsilon_{i'j}$ may be correlated

Use this model (maybe with one effect per study) to estimate τ^2 and then use this between study variance to compute weights

$$w_{ij} = \frac{1}{\left(\hat{\tau}^2 + \bar{v}_{\bullet j}\right)/k_j}$$

Choosing Weights

Impute (guess) a within-study covariance structure for the effect size estimates

Use a hierarchical model to estimate the between-study covariance structure, and use these variance components to compute weights

This should result in

- Better estimates of the variance components
- Better efficiency of the final estimates

Choosing Weights

In principle, differential weighting of correlated estimates within studies *can* matter a lot (e.g., when between-study variance is small and within-study variances and correlations both vary

We need to get a better understanding of whether differential weighing of correlated estimates within studies matters a lot or only a little *in practical situations*

Theory can help but empirical evidence is essential

(Almost) Efficient Weights

We suggest a strategy of assuming a plausible constant value of ρ (the within study correlation of the effects) to estimate weights

It turns out that the estimate of τ^2 is often remarkably insensitive to ρ

For a completely general meta-regression problem we can specify approximate method of moments estimators of τ^2

These can then be used to get almost efficient weights, which can then be used in a meta-regression with robust variance estimates

(Almost) Efficient Weights

In general, if Q_E is the weighted residual sum of squares about the regression line, $w_j = 1/k_j \bar{v}_j$, the method of moments estimate of τ^2 is

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

and $\mathbf{J}_j$ is a k_j x k_j matrix of ones and $\mathbf{V} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$

For Estimating the Mean Effect

Suppose that every effect size estimate in the j^{th} study has weight

$w_j = 1/k_j \bar{v}_j$ then the method of moments estimate of τ^2 is

$$\hat{\tau}^2 = \frac{Q - \left[m - \left(\sum_{j=1}^{m} w_j \Big/ \sum_{j=1}^{m} w_j k_j \right) \right] + \rho \left[1 - \left(\sum_{j=1}^{m} w_j k_j \right)^{-1} \right]}{\sum_{j=1}^{m} w_j k_j - \left(\sum_{j=1}^{m} w_j^2 k_j^2 \right) \Big/ \left(\sum_{j=1}^{m} w_j k_j \right)}$$

Note that when all the $k_j = 1$, this reduces to the usual method of moments estimator of τ^2

Clustering by Groups of Studies

Suppose dependence is induced by clustering of groups of studies

Then the assumption that all correlated effect sizes have the same estimation error variance is implausible

We need a different strategy to estimate almost efficient weights

One idea is to assume two variance components

Between studies within clusters ω^2

Between clusters (between cluster means) τ^2

Clustering by Groups of Studies

Use the weighted residual sum of squares Q_E and another weighted sum of squares

$$Q_1 = \sum_{j=1}^{m} \left(\mathbf{T}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)'\mathbf{J}_j\left(\mathbf{T}_j - \mathbf{X}_j\hat{\boldsymbol{\beta}}\right)$$

Then the method of moments estimators of the variance components are

$$\hat{\omega}^2 = \frac{A_2(Q_1 - C_1) - A_1(Q_E - C_2)}{B_1A_2 - B_2A_1} \qquad \hat{\tau}^2 = \frac{Q_E - C_2}{A_2} - \hat{\omega}^2\frac{B_2}{A_2}$$

Where

$$A_1 = \sum_{j=1}^{m} k_j^2 - tr\left(\mathbf{V}\sum_{j=1}^{m} k_j\mathbf{X}_j'\mathbf{J}_j\mathbf{W}_j\mathbf{X}_j\right) - tr\left(\mathbf{V}\sum_{j=1}^{m} k_j\mathbf{X}_j'\mathbf{W}_j\mathbf{J}_j\mathbf{X}_j\right) + tr\left(\mathbf{V}\left[\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{J}_j\mathbf{X}_j\right]\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{J}_j\mathbf{W}_j\mathbf{X}_j\right)$$

$$B_1 = \sum_{j=1}^{m} k_j - tr\left(\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{J}_j\mathbf{W}_j\mathbf{X}_j\right) - tr\left(\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{J}_j\mathbf{X}_j\right) + tr\left(\mathbf{V}\left[\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{J}_j\mathbf{X}_j\right]\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j^2\mathbf{X}_j\right) \quad C_2 = \sum_{j=1}^{m} k_j - p$$

$$C_1 = tr\left(\mathbf{W}^{-1}\right) - tr\left(\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{J}_j\mathbf{X}_j\right) \quad A_2 = tr\left(\mathbf{W}\right) - tr\left(\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j\mathbf{J}_j\mathbf{W}_j\mathbf{X}_j\right) \quad B_2 = tr\left(\mathbf{W}\right) - tr\left(\mathbf{V}\sum_{j=1}^{m}\mathbf{X}_j'\mathbf{W}_j^2\mathbf{X}_j\right)$$

How Small Can m Safely Be?

We need evidence about rates of convergence to the asymptotic distribution and what it depends on

Our simulations suggest that convergence is relatively rapid for standardized mean differences and equally correlated estimates (confidence intervals have close to nominal content when m = 10 and the k_j = 2 to 5)

Presumably it depends on k/m and the correlation structure

Conclusion

The robust variance estimate looks promising for cases where there are a large number of studies, some or all of which have multiple (correlated) effect size estimates

It is extremely easy to implement (compared to exact multivariate methods)

These estimates are not optimally efficient, but may be close to optimal with a little work estimating approximate variance components

The small sample properties look pretty good so far, but more study will help us better understand the small m (small number of studies) properties

Thank You!

Regularity Conditions Can Fail

These regularity conditions are often met, but can fail to be true

The conditions on regressors are met with bounded fixed $\mathbf{X}_j$' s

The moment conditions on the $\boldsymbol{\varepsilon}_j$'s fail

- always for one instrument IV estimators,
- *d*-statistics with less than 4 degrees of freedom, and
- (formally, but not actually) for odds and risk ratios