Meta-Analysis: Let's start at the very beginning

Elizabeth Tipton

Northwestern University

Overview

1. Overview: Combining effect sizes

2. 2-levels of nesting: Independent effect sizes model
3. 3-levels of nesting: Hierarchical model
4. 3-levels + correlation: Multivariate model

5. Robust variance estimation

Lab: Estimating average effect sizes and heterogeneity

Combining effect sizes

An overview

Data that you need

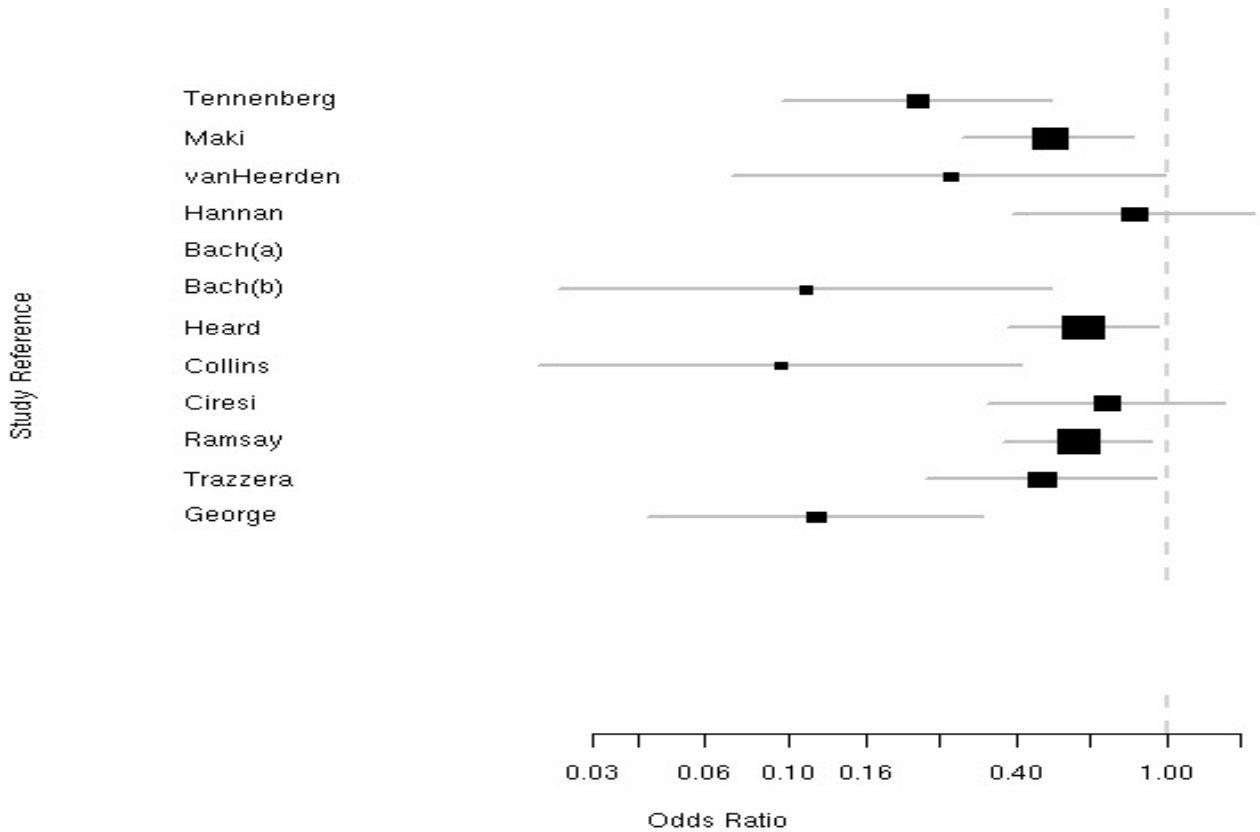
In order to combine effect sizes, whether descriptively or inferentially, at *a minimum* you need two types of data:

- Effect size estimates
- Estimates of the precision of the effect size estimates (e.g., variance, standard errors)

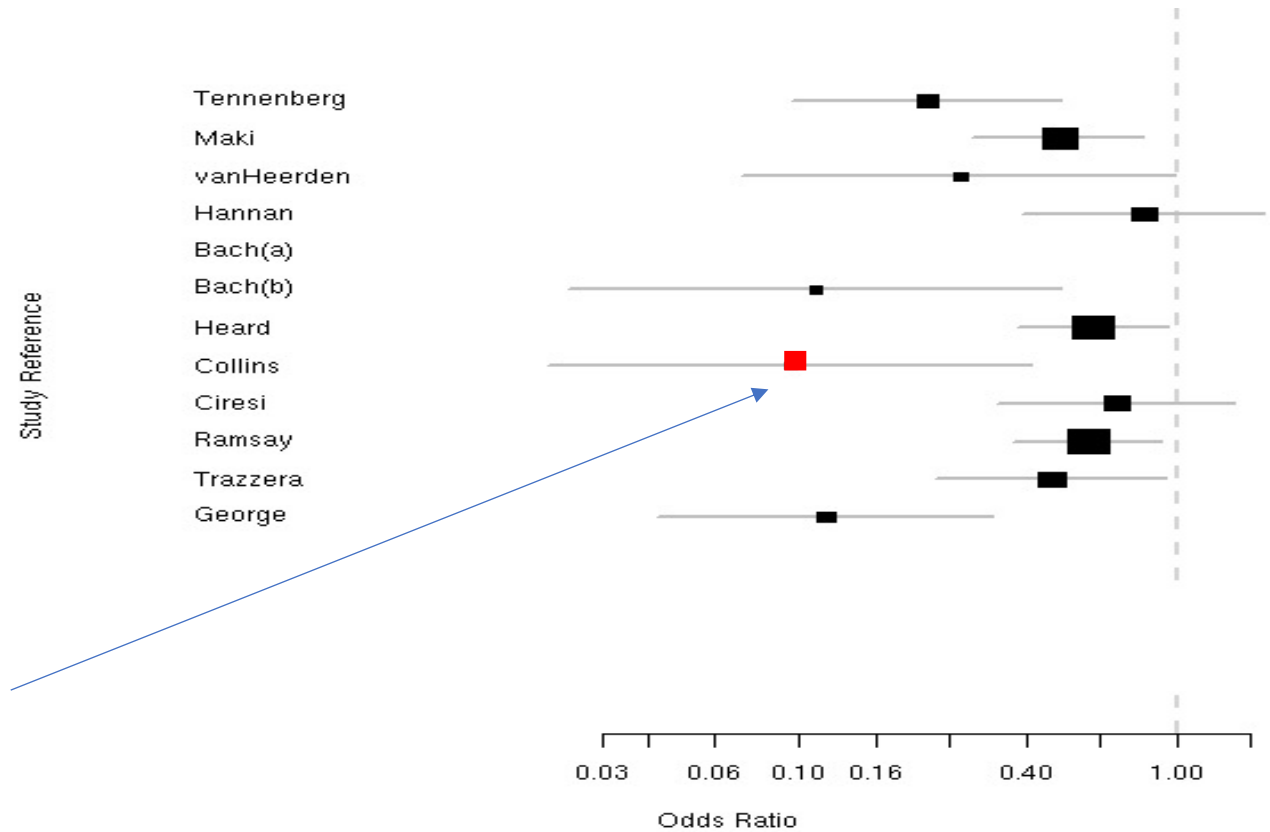
The goal of meta-analysis is to pool these estimates across studies to estimate:

- An **average** effect size (a summary measure)
- The amount of **heterogeneity** in effect sizes
- **Moderators** of effect sizes

Remember: Forest plots

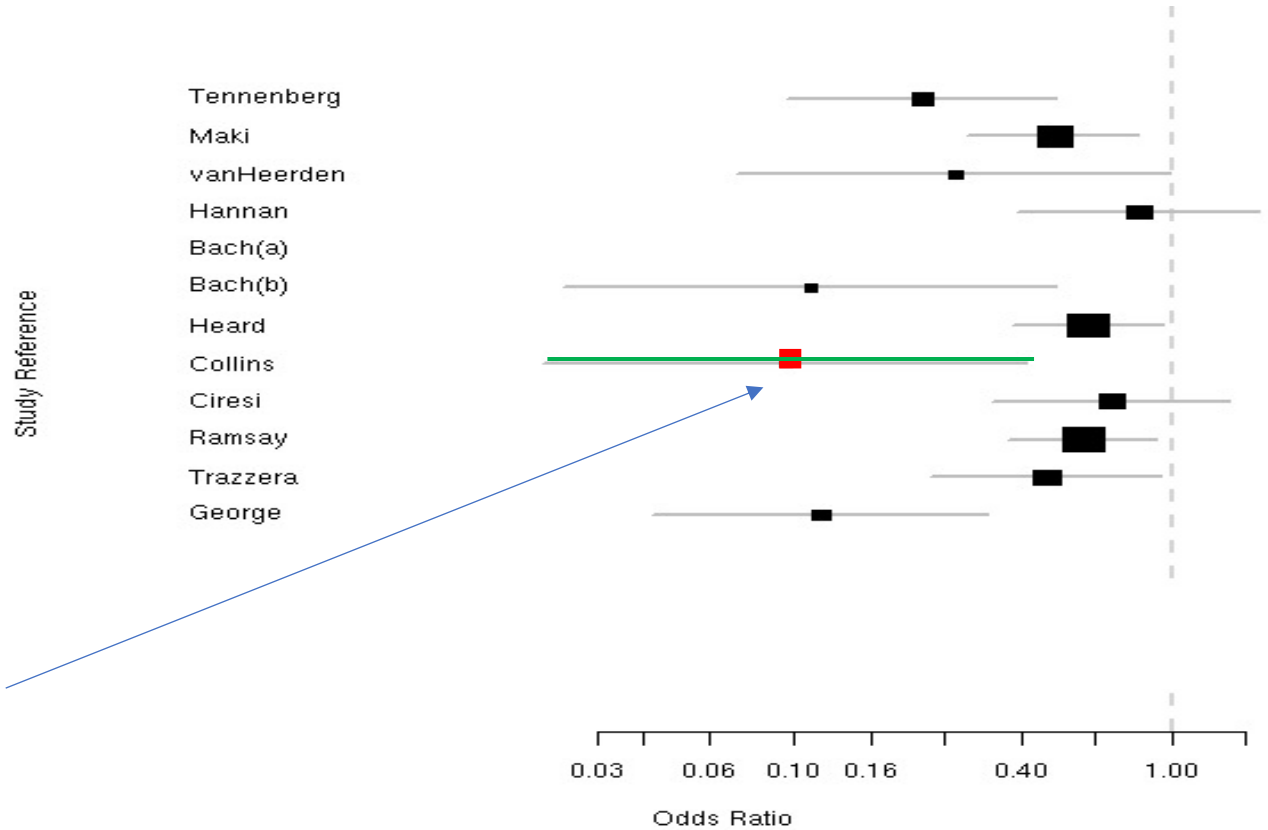


Remember: Forest plots



Estimate

Remember: Forest plots



Estimate

Precision

Know thy data

In order to combine estimates across studies – to meta-analyze – you need to **choose an appropriate model**.

Doing so requires knowledge of the **data structure**:

- Are there **multiple effect sizes** in some studies?
- Are the effect sizes **measured** on independent samples? Are they measured on the same people?
- Did the primary papers provide any information on **how correlated** the outcomes might be?

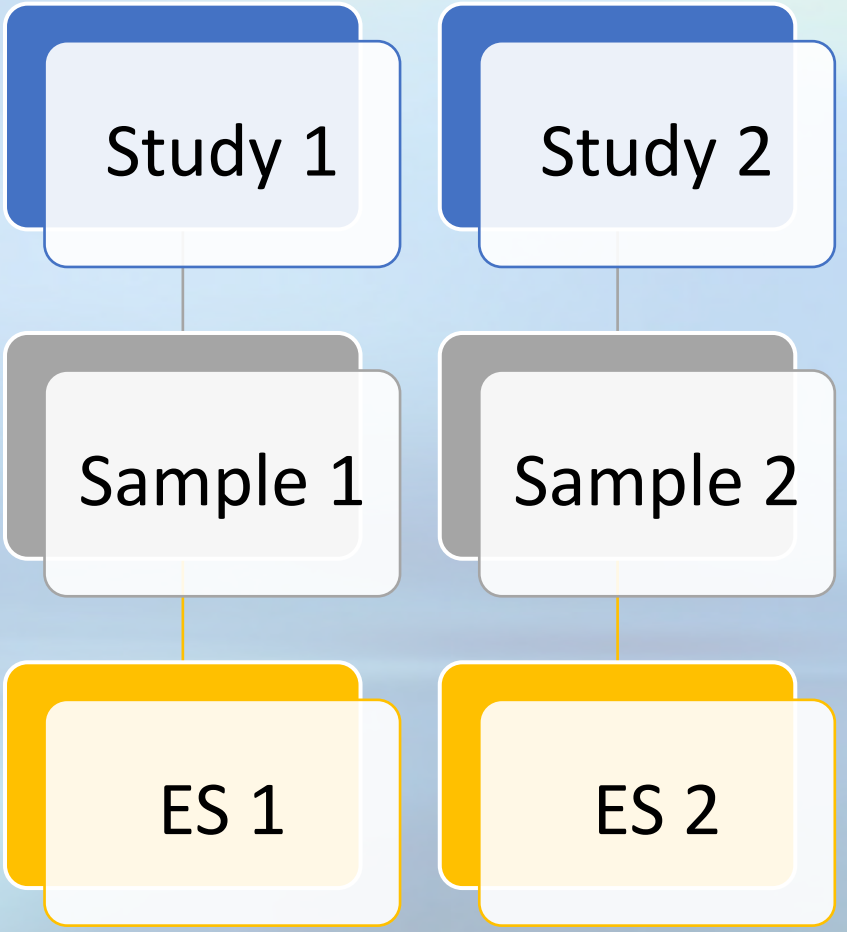
Independent effect sizes model

This is the simplest model. It requires each study to contribute a single ES.

Examples:

- Every study contributes the results of an experiment, with a focus only on writing outcomes.
- Some studies include multiple measures, but these are averaged in the meta-analysis.

Example of Independent Effect Size Model Data Structure



Multilevel effect sizes model

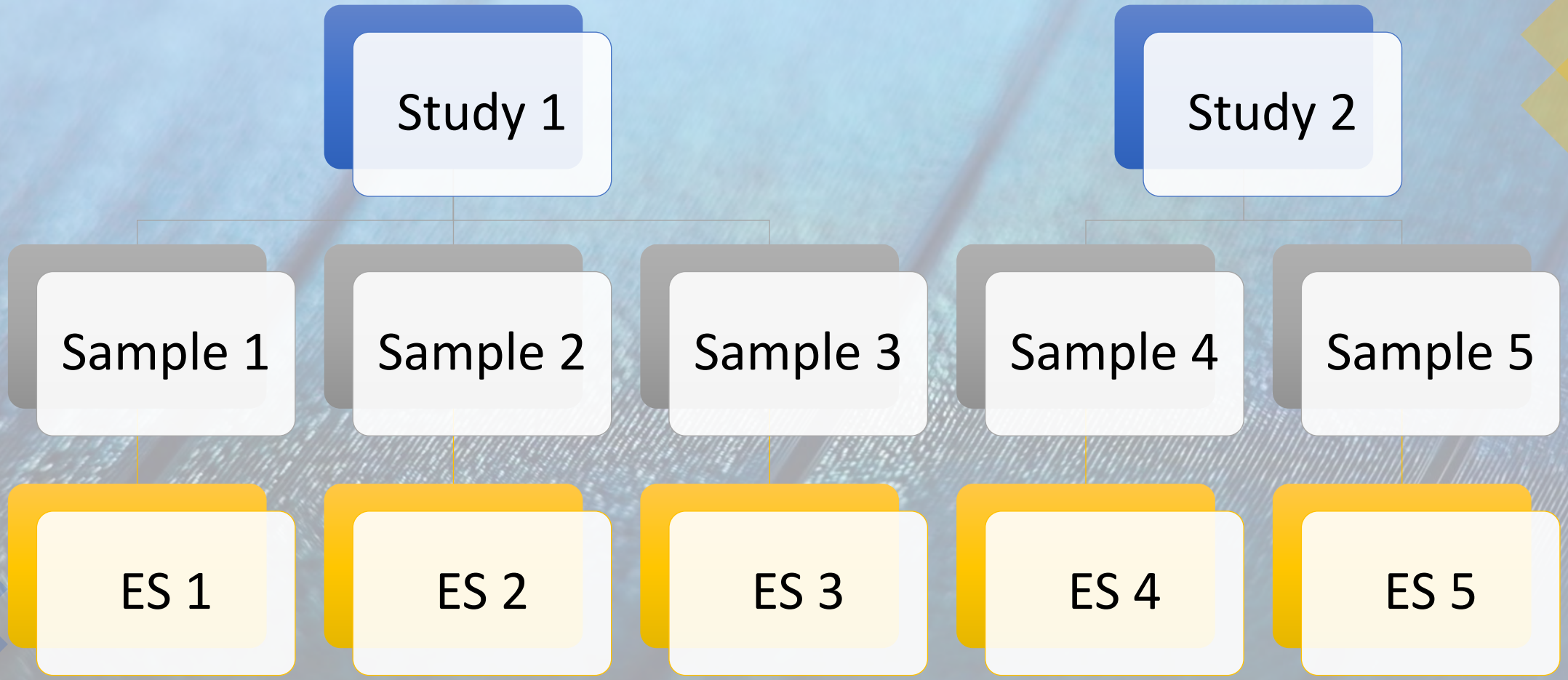
Each study contributes 1 or more effect size estimates.

Within each study, the effect sizes are independent.

Examples:

- The same lab publishes several papers (on different samples of participants) using similar protocols
- A single paper reports the results of multiple experiments using similar protocols or samples (but not the same samples)

Example of Multilevel Effect Size Model Data Structure



Multivariate effect model

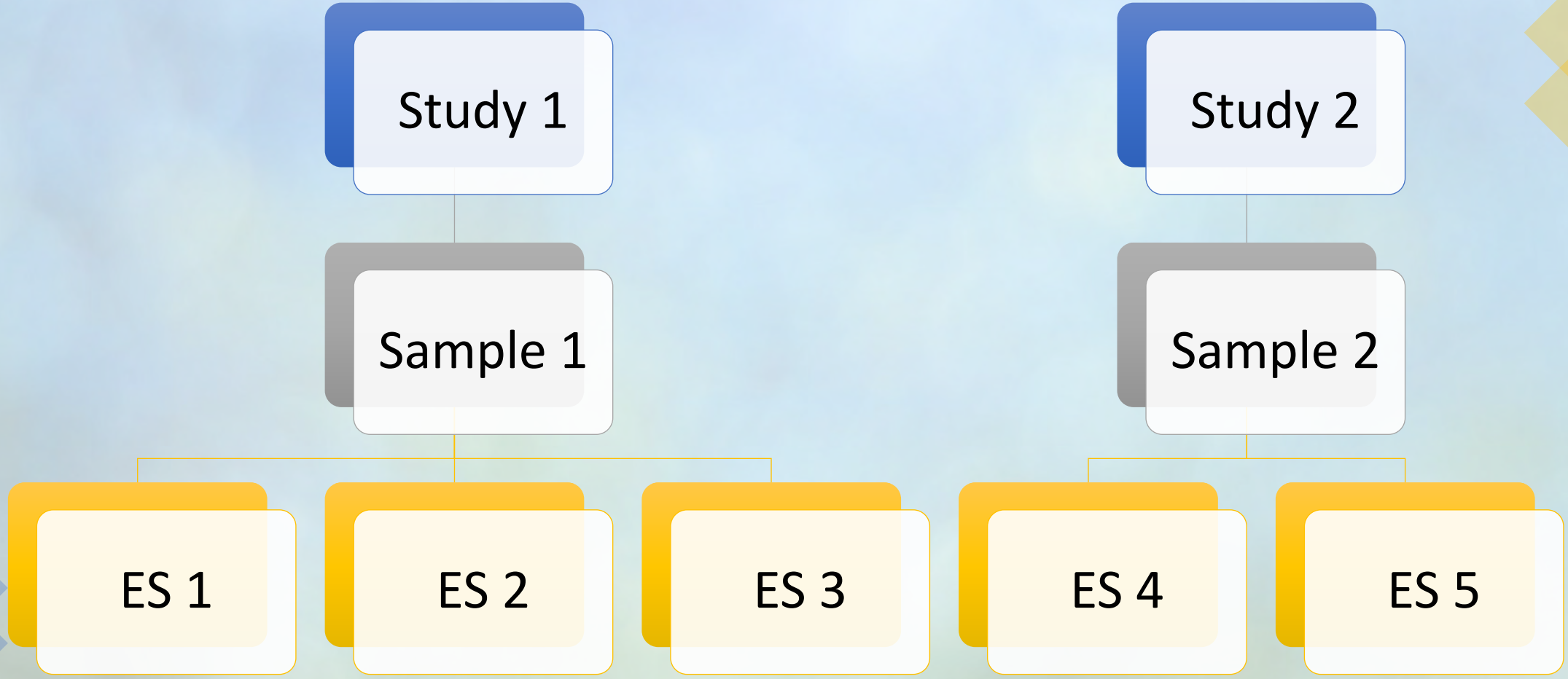
Each study contributes 1 or more effect size estimates.

Within each study, the effect sizes are **correlated.**

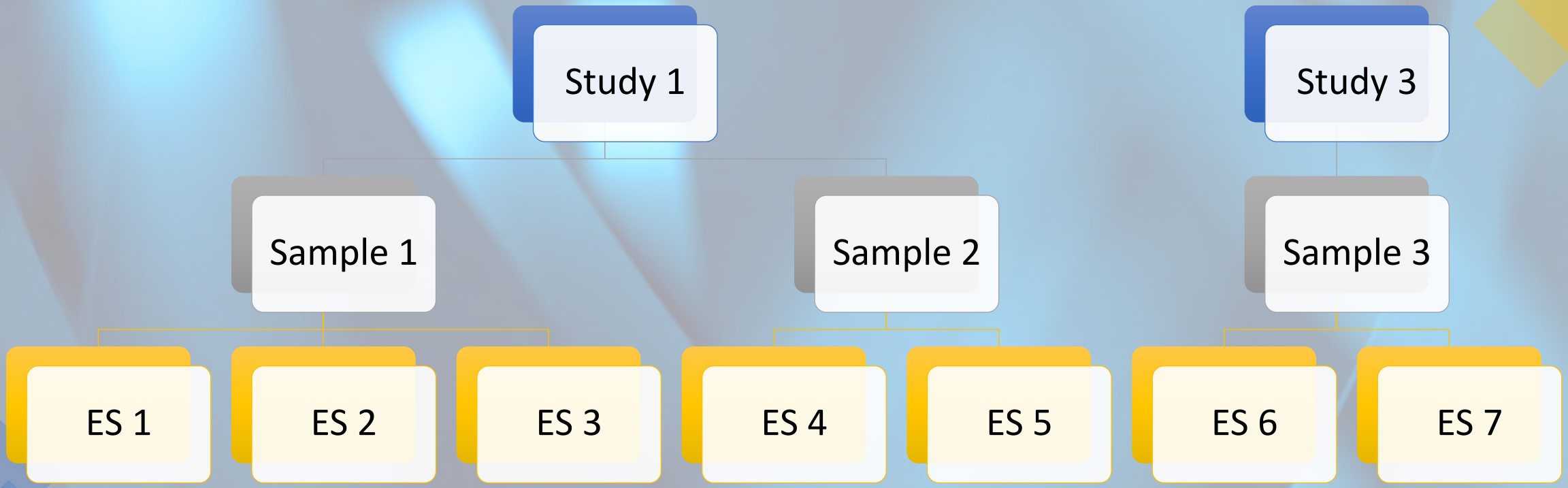
Examples:

- An experiment assigns students to conditions, then measures outcomes using 2 different measures (e.g., accuracy, response time).
- An experiment has 2 treatment arms and 1 control arm.

Example of Multivariate Effect Model Data Structure #1



Example of Multivariate Model Data Structure #2



Choosing a model

It is imperative that you **choose a model that properly accounts for your data structure.**

Reasons **not** to choose a model:

- It’s easy!
- It’s in my software and I don’t know how to use other software.
- I don’t know how to do anything else.
- Everyone else is doing it.
- Hedges or Cooper or Lipsey or someone else once said 30 years ago that this is ok.

Independent effect size model

Combining effect sizes

General model

Assume there are $i = 1, \ldots, k$ studies:

$\theta_1, \theta_2, \ldots, \theta_k$ effect size parameters (e.g. δ, ρ, ω)

$T_1, T_2, \ldots, T_k$ estimates of effect size (e.g. d, g, r, o)

$v_1, v_2, \ldots, v_k$ variances, where

$$v_i = var(Ti|\theta_i) = SE(Ti|\theta_i)^2$$

Then we assume the model

$$T_i \mid \theta_i \sim N(\theta_i, v_i)$$

and that the variances v_i are known.

General model

Assume there are $i = 1, \ldots, k$ studies:

$\theta_1, \theta_2, \ldots, \theta_k$ effect size parameters (e.g. δ, ρ, ω)

$T_1, T_2, \ldots, T_k$ estimates of effect size (e.g. d, g, r, o)

$v_1, v_2, \ldots, v_k$ variances, where

$$v_i = var(T_i|\theta_i) = SE(T_i|\theta_i)^2$$

This is the effect size reported in study i

This is the standard error reported in study i

Then we assume the model

$$T_i \mid \theta_i \sim N(\theta_i, v_i)$$

and that the variances v_i are known.

Assume:
-Unbiased
-Variance known
-Estimate is normally distributed

Fixed effects

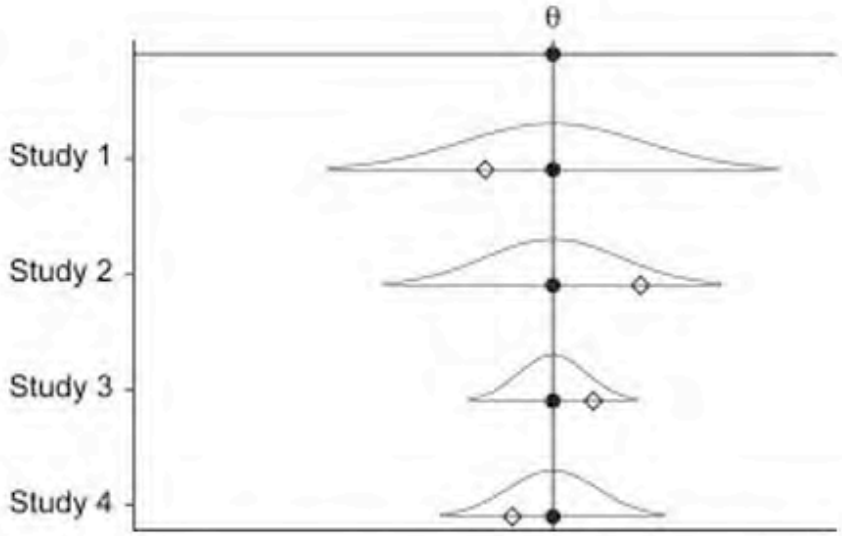


Fig. 1. Schematic representation of fixed-effects meta-analysis. Shown is a hypothetical meta-analysis of four studies, represented by four normal distributions. The fixed-effects model assumes that the true (unobserved) effect in all four studies (*circles*) is the same and equal to a common effect *θ*. The observed effect in each study (*diamond*) deviates from the common effect *θ* because of chance.

Assume that $\theta_1 = \theta_2 = \cdots = \theta_m = \theta$

This means all studies are estimating the same effect size, though estimates will vary.

Random effects

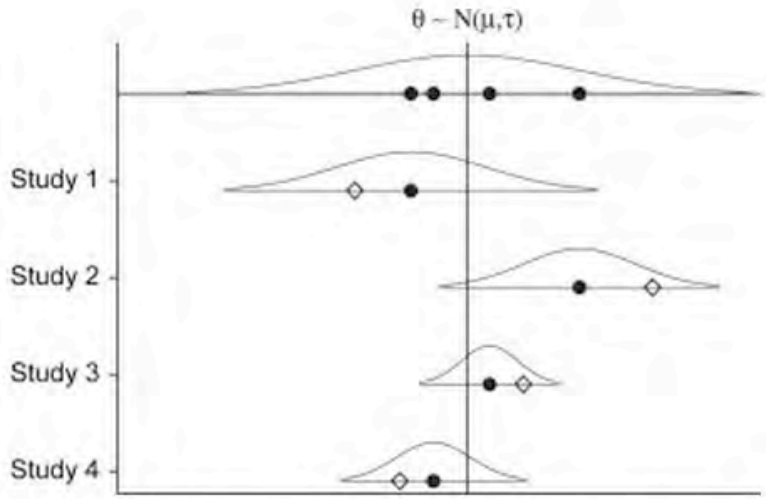


Fig. 2. Schematic representation of random-effects meta-analysis. The figure shows a hypothetical meta-analysis of four studies, each with its own normal distribution. The random-effects model assumes that the study-specific true effects (*circles*) can differ somewhat among the studies, probably reflecting some differences in the populations, the interventions, or some other factor, but that these true effects are distributed around a common grand mean μ with variance τ^2. Note that the observed effect in each study (*diamond*) deviates from its own underlying true effect (*circle*) and from the grand mean μ, as shown in the figure.

Assume instead that $\theta_i \sim N(\mu, \tau^2)$.

This means that instead of all studies estimating the *same* true effect size, there is a *distribution* of true effects.

Random effects is the default

The fixed effects assumption is strong. In general, the random effects model is more appropriate.

In this model, we have:

- The **average effect** is μ
- The **standard deviation** of the effect size distribution is τ

Thus 95% of the true study-specific effect sizes are within the interval:

$$(\mu - 1.96\tau, \mu + 1.96\tau)$$

This is called a **95% Prediction Interval (95PI).**

Aside: Prediction vs Confidence Intervals

People often gets these confused.

A 95% Confidence Interval for μ :

- is an interval estimate that contains the true μ in 95% of random samples.
- is only about inferences regarding the **average effect size**.

A 95% Prediction Interval for θ_i:

- is an interval estimate that contains 95% of the values of θ_i in the population.
- is about making inferences regarding the **distribution of effect sizes across studies.**

Estimation of the mean

We can model the data as:

$$T_i = \theta_i + \varepsilon_i = \mu + \eta_i + \varepsilon_i$$

Study Specific
True Effect

Study Specific
Sampling Error

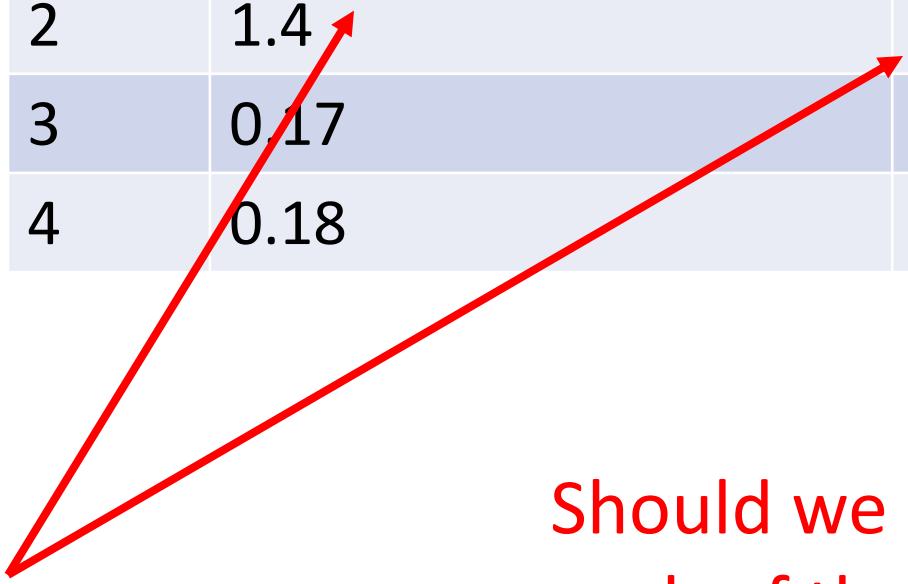
And we want to estimate μ.

We could just estimate this with the **sample mean**.

Data example

Study	Effect Size Estimate (d)	SE(d) Estimate
1	0.20	0.04
2	1.4	0.98
3	0.17	0.10
4	0.18	0.07

The sample mean is:
(0.20+1.4+0.17+0.18)/4
= 0.48



But Study 2 is different.

Its estimate is large.

More importantly, **it isn't very precise**!

Should we really give each of these estimates equal weight?

Precision weighting (aka Inverse Variance Weighting)

The sample mean is not a very precise estimate of the population average.

A more precise estimate gives:

- Greater weight to precise estimates (small standard errors)
- Smaller weight to imprecise estimates (large standard errors)

Formally, this is a **weighted mean.**

Estimate of the population mean

We can use the random effects estimator:

$$T = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}$$

Where

$$w_i = 1/(v_i + \tau^2)$$

We observe this in our data.

We estimate this.

And we can show that:

- $E(T) = \mu$
- $SE(T) = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}$

This is the MVUE **if the weights are exactly inverse-variance.**

Estimation of τ^2

In the RE model, we assume that there is a distribution of ESs, with true variance τ^2. We need to estimate τ^2.

There are many estimators available. These include:

- Method of Moments (MoM)
- Maximum Likelihood (ML)
- Restricted Maximum Likelihood (REML)

We suggest using REML.

Hypothesis Testing: Average effect size

Question: Is the average treatment effect non-zero in the population?

$$H_0: \mu = 0 \; vs \; Ha: \mu \neq 0$$

Hedges (1982): $t = \frac{T}{SE(T)}$

Reject H_0 when $|t| > 1.96$

Default in most software, but not good

Knapp & Hartung (2003): $t = \frac{T}{a * SE(T)}$

where $a = \sqrt{\frac{\sum_{i=1}^{k} w_i (T_i - T)^2}{k-1}}$

Reject H_0 when $|t| > t_{.025,k-1}$

Available. Make it your default.

Hypothesis testing: Heterogeneity

Question: Do true effect sizes vary across studies?

$$H_0: \ \tau^2 = 0 \ vs \ Ha: \ \tau^2 \neq 0$$

Q-test: $\mathrm{Q} = \sum_{i=1}^{k} w_i (T_i - T)^2$

Reject H_0 if $Q > \chi^2_{.05,k-1}$

Notes:

- This test does not have much power in small samples.
- **Do not** use this test to determine if you should use a FE or RE model.

Prediction Interval & Heterogeneity

In addition to the estimate of the average and standard deviation, you should report an estimate of the 95% PI and other heterogeneity statistics.

95% PI Estimate =

$$(T - 1.96\hat{\tau}, T + 1.96\hat{\tau})$$

$$I^2 = \frac{Q-(k-1)}{Q}$$

I^2 is interpreted as the proportion of observed variation that is “true” variation.

Multilevel Model

Combining effect sizes

Hierarchical data

Now we have studies $i = 1, \ldots, k$. Each study has $j = 1, \ldots, k_j$ effect size estimates.

θ_{ij} = parameter for ES j in study i (e.g. δ, ρ, ω)

T_{ij} = estimate of ES j in study i (e.g. d, g, r, o)

v_{ij} = variances of ES j in study i

$$v_{ij} = var(T_{ij}|\theta_{ij}) = SE(T_{ij}|\theta_{ij})^2$$

As before, we assume $T_{ij} \mid \theta_{ij} \sim N(\theta_{ij}, v_{ij})$

and we typically assume that the variances v_{ij} are known.

Hierarchical data

Now we have studies $i = 1, \ldots, k$. Each study has $j = 1, \ldots, k_j$ effect size estimates.

θ_{ij} = parameter for ES j in study i (e.g. δ, ρ, ω)

T_{ij} = estimate of ES j in study i (e.g. d, g, r, o)

v_{ij} = variances of ES j in study i

$$v_{ij} = var(T_{ij}|\theta_{ij}) = SE(T_{ij}|\theta_{ij})^2$$

As before, we assume $T_{ij} \mid \theta_{ij} \sim N(\theta_{ij}, v_{ij})$

and we typically assume that the variances v_{ij} are known.

This is the jth effect size reported in study i

This is the standard error reported for the jth effect size in study i

Assume:
-Unbiased
-Variance known
-Estimate is normally distributed

Hierarchical Model

Can be written:

$$T_{ij} = \theta_{ij} + \varepsilon_{ij} = \mu + \eta_i + \phi_{ij} + \varepsilon_{ij}$$

ES Specific True Effect

ES Specific Sampling Error

where now we assume that:

$$\theta_{ij} \sim N(\mu, \tau^2 + \omega^2)$$

Where

- μ is the **average effect size**
- τ is the standard deviation of the distribution of **true study-average** ES
- ω is the standard deviation of the distribution of **true ES within studies**

Within study independence

Assume that the residuals ε_{ij} are **not** correlated.

This means for two ESs in the same study we have:

ES j: $T_{ij} = \mu + \eta_i + \Phi_{ij} + \varepsilon_{ij}$

ES k: $T_{ik} = \mu + \eta_i + \Phi_{ik} + \varepsilon_{ik}$

Have covariance

$$Cov(T_{ij}, T_{ik}) = \tau^2 + \omega^2$$

Estimation of the average ES μ

The rest follows in a similar form.

- Inverse variance weights are most precise
- We observe the v_{ij} in our data and treat it as known
- We estimate both τ^2 and ω^2

We use the same estimator, but with weights:

$$w_{ij} = 1/(\tau^2 + \omega^2 + v_{ij})$$

Interpretation in this model

T is the estimate of the **average effect size** in the population.

The degree of **heterogeneity** can be summarized using τ^2 and ω^2.

The **total variation** in true effect sizes is now $\tau^2 + \omega^2$.

95% Prediction Intervals

A **95%PI for effect sizes** is:

$$(\mu - 1.96\sqrt{\tau^2 + \omega^2}, \mu + 1.96\sqrt{\tau^2 + \omega^2})$$

A **95%PI for study-average** effect sizes is:

$$(\mu - 1.96\tau, \mu + 1.96\tau)$$

This gives you a sense of the range of true effect sizes (or study average ES) found in your population (a measure of heterogeneity).

Multivariate Model

Can be written:

$$T_{ij} = \theta_{ij} + \varepsilon_{ij} = \mu + \eta_i + \phi_{ij} + \varepsilon_{ij}$$

ES Specific True Effect

ES Specific Sampling Error

where now we assume that:

$$\theta_{ij} \sim N(\mu, \tau^2 + \omega^2)$$

Where

- μ is the **average effect size**
- τ is the standard deviation of the distribution of **true study-average** ES
- ω is the standard deviation of the distribution of **true ES within studies**

Multivariate Model

This looks the same as the Multilevel Model!

Can be written:

$$T_{ij} = \theta_{ij} + \varepsilon_{ij} = \mu + \eta_i + \phi_{ij} + \varepsilon_{ij}$$

ES Specific True Effect

ES Specific Sampling Error

where now we assume that:

$$\theta_{ij} \sim N(\mu, \tau^2 + \omega^2)$$

Where

- μ is the **average effect size**
- τ is the standard deviation of the distribution of **true study-average** ES
- ω is the standard deviation of the distribution of **true ES within studies**

But now there is correlation

Now assume that the residuals ε_{ij} **are** correlated:

$$\text{ES } j: \quad T_{ij} = \mu + \eta_i + \Phi_{ij} + \varepsilon_{ij}$$

$$\text{ES } k: \quad T_{ik} = \mu + \eta_i + \Phi_{ik} + \varepsilon_{ik}$$

Then the covariance is:

$$Cov(T_{ij}, T_{ik}) = \tau^2 + \omega^2 + \rho\sqrt{v_{ij}v_{ik}}$$

Where ρ is the correlation between measurements j & k on the same person within study i.

Like v_{ij} and v_{ik}, ρ *must be provided* from the original study.

Nevermind then…

This Multivariate Model has been around for a long time (it was in the original Hedges & Olkin!).

But studies don't often report these correlations, which has led people to avoid using this model.

But we don' t have to give up! Using a sensitivity + robustness approach is feasible:

- "Guess" at the correlation (e.g., .80) when unreported.
- And use RVE (next up)…

Multivariate is just an extension to Multilevel

So the interpretation is the same:

T is the estimate of the **average effect size** in the population.

The degree of **heterogeneity** can be summarized using τ^2 and ω^2.

The **total variation** in true effect sizes is now $\tau^2 + \omega^2$.

RVE: A safety net

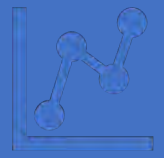
Robust standard errors

What if your model isn't quite right?

What if we have multiple effect sizes per study (so independent effect size model fails)?

But also, some of the effect sizes within studies might be correlated with each other (so multilevel model fails)?

But you also don't have any reported info on the correlation structure (so multivariate model seems impossible)?



Or variance problems

You're having trouble extracting variances from some studies – e.g., correlations not reported between pre-post, or only sample size reported.

You can make assumptions to extract these – but you could be wrong.

And if you're wrong, your weights may not be exactly inverse variance.

Your standard errors could be wrong.

What's wrong with this?

Our estimate T of the population mean μ is ok.

But our **standard error** estimator ***has problems***.

Recall, we used the estimator:

$$SE(T) = \sqrt{\frac{1}{\sum_{i=1}^{k} \sum_{j=1}^{kj} w_{ij}}}$$

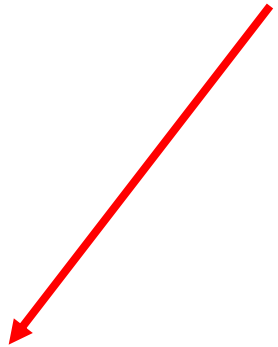
Which was based on the assumptions that the:

- Weights w_{ij} are *exactly* inverse variance.
- *Variances are correctly specified.*
- Effect sizes *within* studies are *independent*.

Misspecification

->

Hypothesis tests might be wrong



Robust variance estimation (RVE)

Up until now, we have been using **Model Standard Errors**.

Instead of using a model to estimate the variance, RVE uses the observed variation in effect sizes to estimate the standard error.

Estimators of this type are also known as:

- Huber-White standard errors
- Cluster Robust Variance Estimation (CRVE)
- Empirical standard errors

Residuals as estimators

We define the observed residuals as: $e_{ij} = T_{ij} - T$

Instead of assuming we know that $V(T_{ij}) = \tau^2 + \omega^2 + v_{ij}$

We now estimate

$$v(T_{ij}) = eij^2$$
$$cov(T_{ij}, T_{ik}) = e_{ij}e_{ik}$$
$$cov(T_{ij}, T_{i'k}) = 0$$

Seems strange right? That's because for each individual study this is a *terrible estimate* of each variance and covariance…

RVE estimator

…but we don't care about how these individual estimates perform. We *only care about them when averaged* (sort of) across studies.

Hedges, Tipton, and Johnson (2010) showed that when estimating the average effect size, if the within-study weights are equal, the RVE estimator can be written:

$$v^R(T) = \frac{\sum_{i=1}^{k} w_i^2 \, \bar{e}_i^{\,2}}{\left(\sum_{i=1}^{k} w_i\right)^2}$$

Hypothesis testing using RVE

We can use RVE to test:

$$H_0\text{: } \mu = 0 \; vs \; Ha\text{: } \mu \neq 0$$

Tipton (2015) shows that the appropriate small-sample test is:

$$t = \frac{T}{SE^R\,(T)}$$

And we can reject H_0 if: $|t| > t_{.025,df}$

where typically $df \cong k - 1$.

What about heterogeneity?

Using RVE means that your hypotheses tests about the **average effect size** hold even if:

- The effect sizes are **not** normally distributed
- The variances are incorrect
- The weights are not inverse variance

But, RVE can't help with heterogeneity parameters. In order to estimate τ^2 and ω^2, 95% Prediction Intervals, I^2, and test hypotheses about these, **the above assumptions are required.**

This means you should think of these heterogeneity statistics as good estimates "if the assumptions hold" and otherwise as approximations.

Final note

If you've heard about RVE before, you may be accustomed to thinking of this as:

RVE vs Multivariate

But what I'm saying here is:

It's not either-or – it can be *both.*

Using the model provides you some structure and thus heterogeneity estimates. Adding RVE makes you robust to misspecification. Everyone wins!

Lab

metafor + clubSandwich

Activity

We will use the R package ‘metafor’ for both the independent effect sizes model and the hierarchical model.

We will use the R package ‘clubSandwich’ to implement RVE.

For both, we will use the TDV_KW.csv data.

Go to RStudio now and open MALab1.R

Wrap up: Three steps

1. **Approximate your “random effects” (covariance) structure.**

2. **Model your “fixed effects” (here just a mean).**

3. **Guard against misspecification**. Use Robust Variance Estimation to estimate our standard errors and hypothesis tests.

Citations (for your proposal)

Why prediction intervals should be reported:

- Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research synthesis methods*, *8*(1), 5-18.

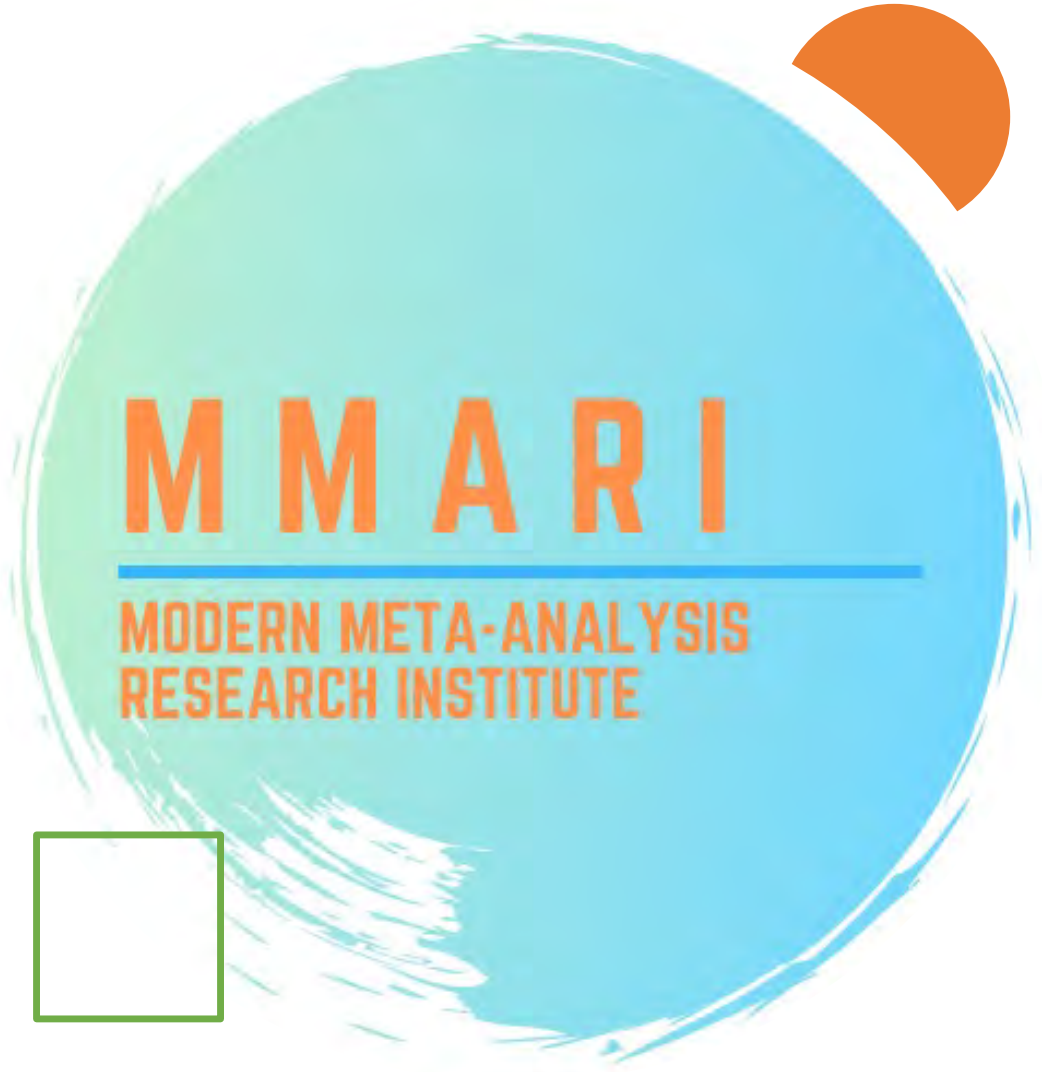
RVE & t-tests:

- Hedges, L.V., Tipton, E., and Johnson, M.C. (2010) Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*. 1(1): 39-65. Erratum in 1(2): 164-165.
- Tipton, E. (2015) Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3): 375 – 393.
- Tanner-Smith, E., Tipton, E., & Polanin, J. (2016) Handling Complex Meta-Analytic Data Structures using Robust Variance Estimates: A Tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2(1): 85-112.
- Pustejovsky, J. & Tipton, E. (2021) Meta-Analysis with Robust Variance Estimation: Expanding the Range of Working Models. Forthcoming in *Prevention Science*.

Meta-Regression: Please use more than one variable

Elizabeth Tipton

Northwestern University



Overview

Moderators and analysis plans

Meta-regression

Lab: Estimating meta-regression models

Moderator analyses

An overview

Heterogeneous effect sizes

In the social sciences and education, effect sizes are most often heterogeneous.

This is indicated by:

- A large degree of variation (τ^2 and ω^2)
- A wide 95%PI
- A large I^2 value

When heterogeneity is detected, it is important to be **cautious in the interpretation of the average effect size.**

Careful interpretation

Whenever there is heterogeneity, the results of a meta-analysis must be carefully interpreted, particularly in a policy context.

- If all effects are the same sign, the question is really about the *magnitude* of the effect.
- If some effects are positive and some are negative, the question is really about *for whom, what, when, where, etc the intervention works*.

Ecological fallacy

The ecological fallacy: the relationship between the aggregate X and aggregate Y may differ from the relationship between each X and Y.

For example, if you find that the effect size of an intervention is larger for studies using older children than for studies using younger children, it could either be that

- in fact the treatment works better for older children than younger children, or
- the treatments used for older children were better than those for younger children, or
- the research groups doing studies on older children had better quality than those using younger children.

Confounding and controls

Recall that meta-analysis is an **observational study**. Even if all of the studies themselves are experiments, the study conditions or populations were not randomly assigned to the experiments.

For this reason, it is important to separate out “focal” variables from “controls”.

Durlak and Lipsey (1991)

	Overall Mean ES	Research Design	
		Comparison group	Pre-post
Primary	.51 (k = 260)	.43 (k = 200)	.76 (k = 60)
Secondary	.58 (k = 240)	.36 (k = 100)	.74 (k = 140)

Durlak and Lipsey (1991)

	Overall Mean ES	Research Design	
		Comparison group	Pre-post
Primary	.51 (k = 260)	.43 (k = 200)	.76 (k = 60)
Secondary	.58 (k = 240)	.36 (k = 100)	.74 (k = 140)

Durlak and Lipsey (1991)

This is methodological confounding!!

	Overall Mean ES	Research Design	
		Comparison group	Pre-post
Primary	.51 (k = 260)	.43 (k = 200)	.76 (k = 60)
Secondary	.58 (k = 240)	.36 (k = 100)	.74 (k = 140)

Hypothesis testing

Whenever possible, control for confounding variables. **There is no reason to conduct analyses 1-variable at a time.**

But, when there are multiple variables, there are two concerns that arise with testing:

- Inflated Type I error (which can arise from “fishing” or “p-hacking”)
- Not having adequate power (inflated Type II error)

We need to be careful.

Strategies to contain error

1. **Pre-specify covariates** of interest and treat these analyses as “confirmatory”.
2. **Include confounders** in your model, but don’t report their p-values or hypothesis tests (that’s not their role).
3. **Use multiple comparisons** corrections for your analyses (i.e., smaller p-value threshold *a la* Bonferoni)
4. Make clear when **analyses are exploratory** and emphasize the need for future studies to test these hypotheses.

Meta-regression

Models and methods

Mixed models

Moving from meta-analysis to meta-regression isn't that difficult.

It just involves adding covariates to the models we've already described.

We will call all of these covariates $\boldsymbol{X}$= $(X_1, X_2, \ldots, X_p)$, regardless of if they are continuous or categorical, focal or controls.

General MR model

Recall that in general we have:

T = fixed + residuals

Where “fixed” is the mean (μ)

and

Residuals =

- $\eta_j + \epsilon_{ij}$ in the Univariate Model
- $\eta_j + \Phi_{ij} + \epsilon_{ij}$ in the Multilevel and Multivariate Models

In MR, we model this **fixed part** using regression.

Three steps

1. Just as in MA, we select our “random effects” structure:

- Univariate model
- Multilevel model
- Multivariate model

2. **Then we model the “fixed” part using covariates.**

3. And use Robust Variance Estimation to estimate our standard errors and hypothesis tests.

Example of a model

$$\theta_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \eta_j + \Phi_{ij}$$

For example:

- X_1 = average age of patients
- X_2 = dummy variable indicating if study is an RCT
- X_3 = length of time patients were in treatment
- X_4 = dummy variable indicating if outcome is physical measure (vs cognitive)

Example of a model

Control/ confounder variable

$$\theta_{ij} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3ij} + \beta_4 X_{4ij} + \eta_j + \Phi_{ij}$$

For example:

- X_1 = average age of patients
- X_2 = dummy variable indicating if study is an RCT
- X_3 = length of time patients were in treatment
- X_4 = dummy variable indicating if outcome is physical measure (vs cognitive)

Estimation of the coefficients

In general, we can use **Weighted Least Squares (WLS)** to estimate the regression coefficients.

These are estimated using: $\boldsymbol{b} = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WT}$

Where we can specify the weights ($\boldsymbol{W}$) based upon the model for the random effects that we use (so as to be approximately inverse variance).

RVE for estimation of the standard errors

And again, we use RVE to estimate the standard errors.

Tipton (2014) shows that we write this estimator more generally as:

$$v^R(\boldsymbol{b}) = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WAee'AWX}\,(\boldsymbol{X'WX})^{-1}$$

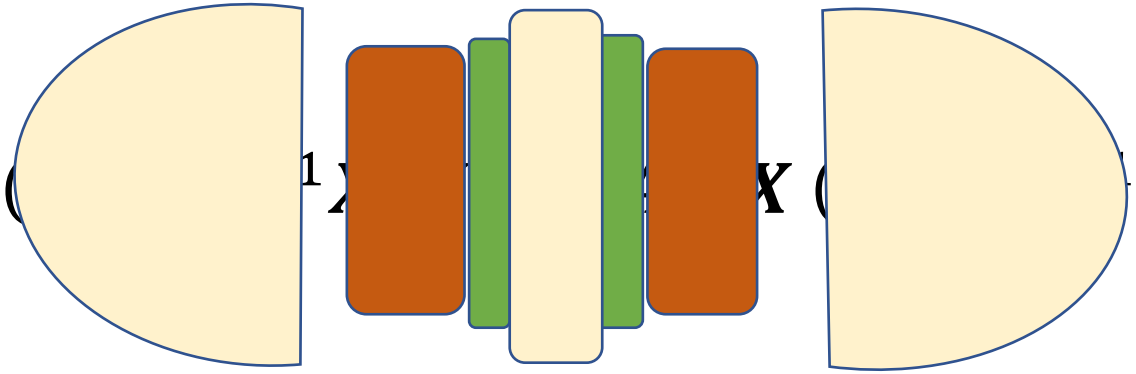
where the $\boldsymbol{A}$ matrices provide adjustments so that the estimator works well in small and moderate samples.

RVE for estimation of the standard errors

And again, we use RVE to estimate the standard errors.

Tipton (2014) shows that we write this estimator more generally as:

$$v^R(\boldsymbol{b}) = ($$



$$1 \boldsymbol{X} ($$

Club Sandwich

where the **A** matrices provide adjustments so that the estimator works well in small and moderate samples.

RVE for estimation of the standard errors

And again, we use RVE to estimate the standard errors.

Tipton (2014) shows that we write this estimator more generally as:

$$v^R(\boldsymbol{b}) = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WAee'AWX}\,(\boldsymbol{X'WX})^{-1}$$

Club Sandwich

where the **A** matrices provide adjustments so that the estimator works well in small and moderate samples.

Individual covariates hypotheses

For each of the focal variables, you could test if there is a relationship or not in the population, i.e.,

$$H_0\text{: } \beta_m = 0 \quad vs \quad Ha\text{: } \beta_m \neq 0$$

Tipton (2015) test: $t_m = \frac{b_m}{SE^R(b_m)}$

And we reject H_0 when $|t_m| > t_{.025,df_m}$

Note: the degrees of freedom (df_m) depend on the covariate.

Joint hypothesis tests

Example: Suppose there are 3 age groups studied: **children, teenagers, and adults**, and that these are included in the model as **2 dummy variables**.

$$H_0\colon \beta_1 = \beta_2 = 0 \text{ vs } \quad H_0\colon \beta_1 \neq 0 \;\&/or\; \beta_2 \neq 0$$

Tipton & Pustejovsky (2015) AHT F-test (here $q = 2$):

$$F_{12} = \frac{d-q+1}{dq}(\boldsymbol{b}_{12} - \mathbf{0})'\left(\boldsymbol{v}^R(\boldsymbol{b}_{12})\right)^{1/2}(\boldsymbol{b}_{12} - \mathbf{0})$$

Reject H_0 if $F_{12} > F_{.025,q,df_m=d-q+1}$ where again, $df_m = d - q + 1$ are estimated.

Degrees of freedom

In both the **t-test** and **F-tests** there are degrees of freedom that are estimated.

These degrees of freedom depend on:

- The number of *independent studies*
- The *number of covariates* tested
- The **type of covariates** and **features of the covariate**, including skewness and balance

If $df_m > 4$ you can trust the p-values that are given. **If $\boldsymbol{df_m < 4}$**, the approximation isn't as good and you should use a higher standard of evidence (e.g., $p < .01$).

Descriptive comparisons

How much of the ES variability do the moderator variables explain?

To answer this, we can calculate an R^2 value, like in regression:

1. Run the model without any variables and get τ^2
2. Run the model with variables and get ${\tau_c}^2$
3. Then the variables explain 100*R^2 of the variation in ESs, where

$$R^2 = 1 - {\tau_c}^2/\tau^2$$

Caveat: This doesn't always turn out great – for various reasons, you can end up with $R^2 < 0$.

Lab

metafor + clubSandwich

Activity

MR is straightforward – now we build off of what you already know how to do in **metafor** and **clubSandwich**.

We will use the same data. But now we will add moderators and F-tests.

Go to RStudio now and open MALab2.R.

Wrap up: Four steps

1. **Approximate your “random effects” covariance structure**

2. **Model your “fixed effects”**

3. **Identify your focal variables and your confounding variables.**

3. **Guard against misspecification**. Use Robust Variance Estimation to estimate our standard errors and hypothesis tests.

Remember these strategies to contain error

1. **Pre-specify covariates** of interest and treat these analyses as “confirmatory”.
2. **Include confounders** in your model, but don’t report their p-values or hypothesis tests (that’s not their role).
3. **Use multiple comparisons** corrections for your confirmatory analyses (i.e., smaller p-value threshold *a la* Bonferoni)
4. Make clear that **other analyses are exploratory** and emphasize the need for future studies to test these hypotheses.