UiO : CEMO – Centre for Educational Measurement
University of Oslo

Meta-Analyses in Educational Research

A Hands-On Workshop

Diego G. Campos & Ronny Scherer

QRM Conference 2024
Gothenburg, 11 June 2024


BEGIN.

Workshop Content

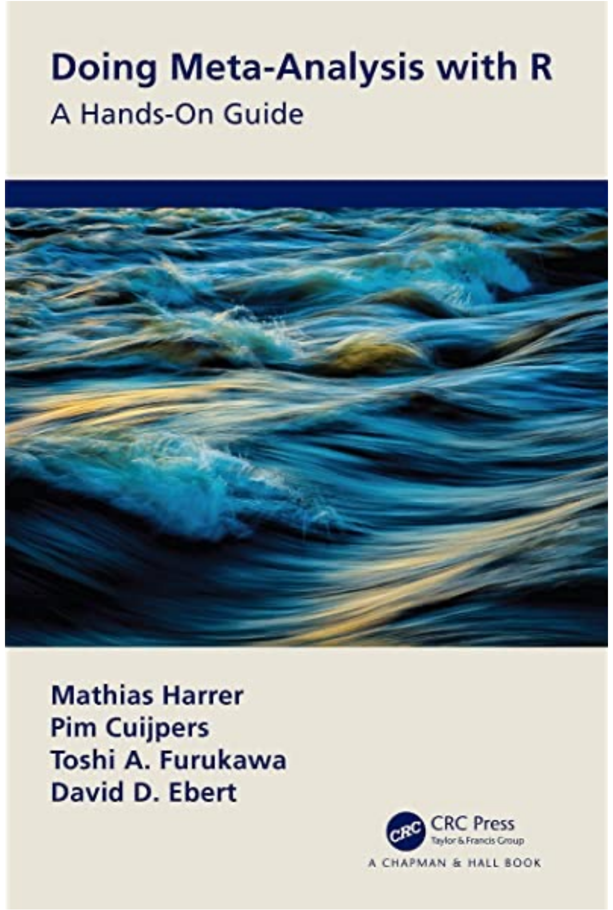
Basic concepts of synthesizing effect sizes

- Standard random-effects models
- Meta-analytic data structure
- Multilevel meta-analysis
- Moderator analyses
- Publication bias and influential effect sizes

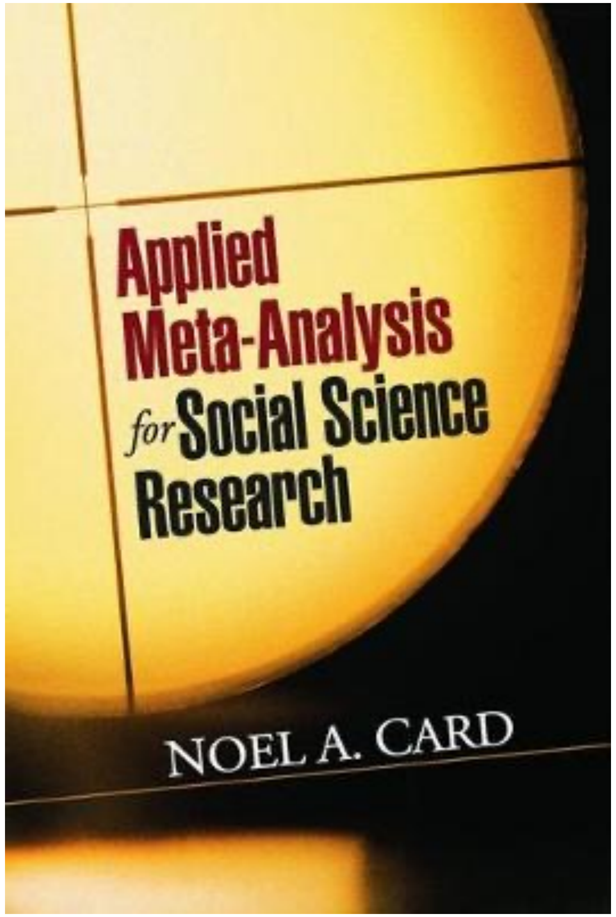
Data analyses and illustrative example using the R packages `metafor` and some supplementary packages

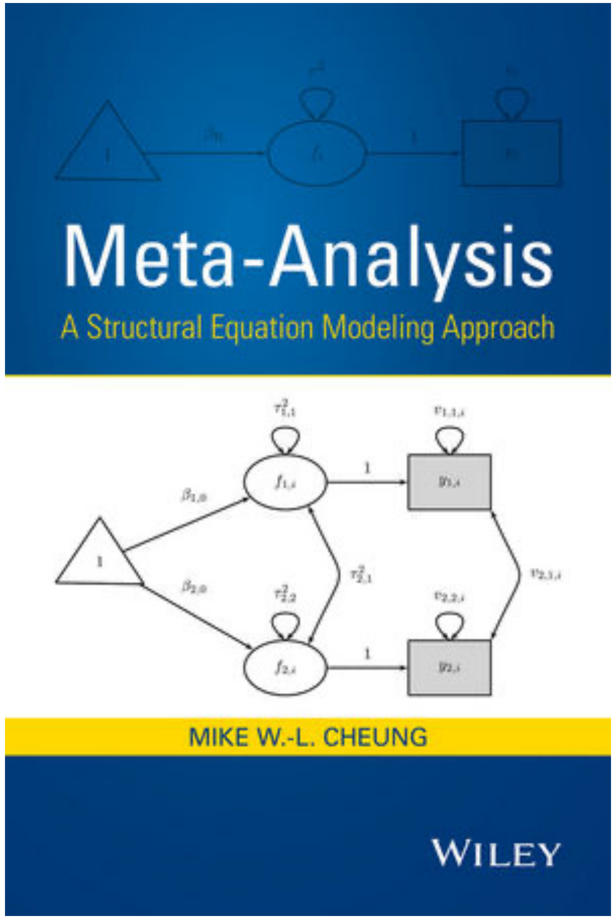
Complexities of including data from international large-scale assessments (ILSAs)

Recommended Literature



https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

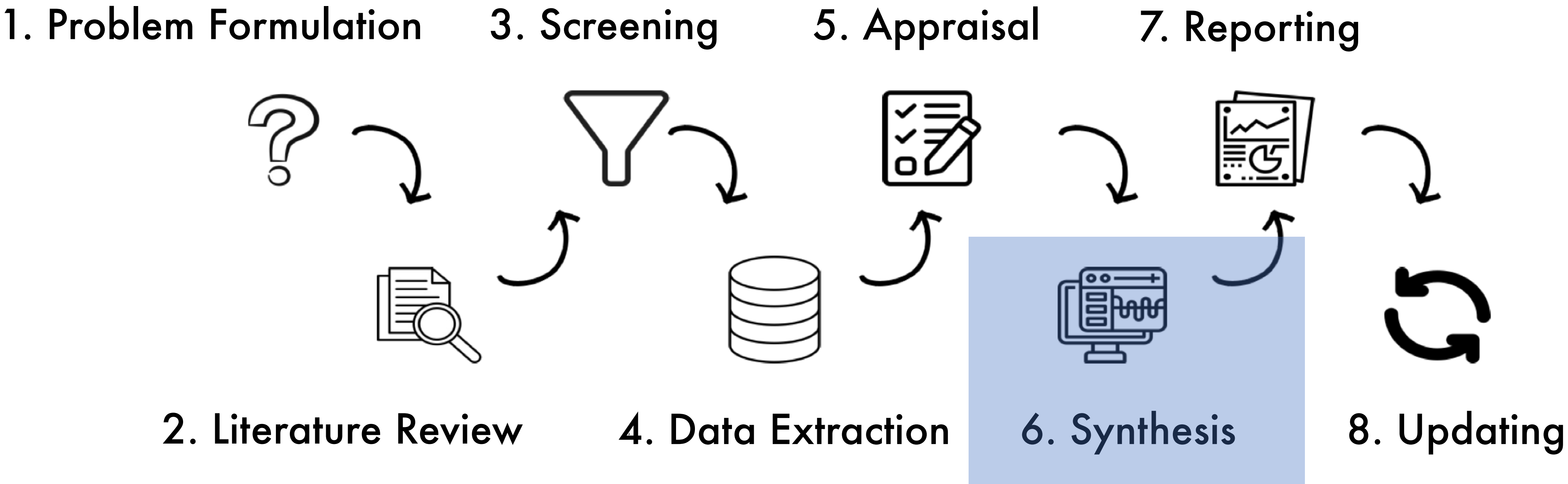




https://onlinelibrary.wiley.com/doi/book/10.1002/9781118957813

Steps in a Meta-Analysis

(Borenstein et al., 2009; Card, 2012)



Focus of this workshop:
Quantitative synthesis of effect sizes

Main Purposes of Meta-Analyses

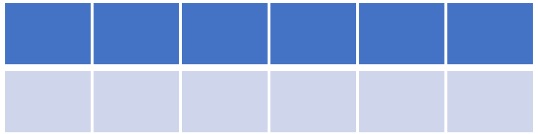
Three key outcomes of a meta-analysis—pooled effect size, heterogeneity, and moderator effects

Pooled Effect Size and Heterogeneity

(Borenstein et al., 2009)

Typical univariate meta-analysis

m effect sizes
m studies



Study 1 θ_1 v_1
Study 2 θ_2 v_2
… … …
Study m θ_m v_m

Sampling variances v_i

Effect sizes θ_i
(e.g., correlations, standardized mean differences, regression coefficients)

Pooling & Weighting →

$\bar{\theta}$

Weighted average effect size

τ^2

Between-study heterogeneity

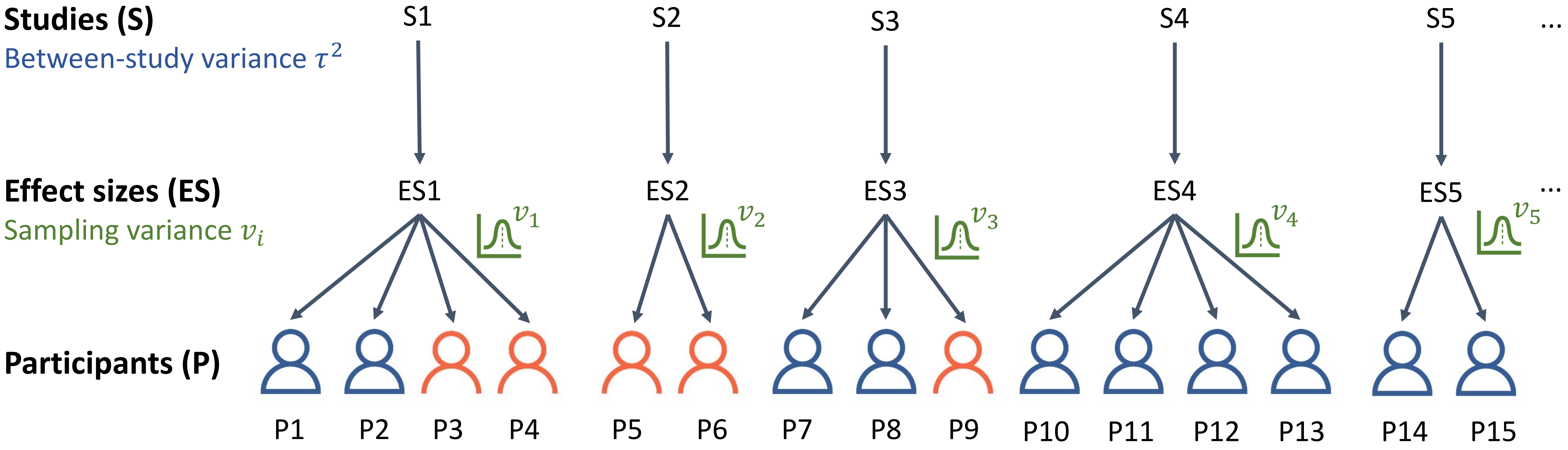
x_i →

Moderators
(e.g., study, sample, measurement characteristics)

Variance Components in a Meta-Analysis

(CMM, 2019)

Two-level hierarchical structure



Data Example: Gender Differences in Digital Literacy

(Campos et al., 2023; Scherer et al., 2024)

Meta-analytic data set of the gender differences in students' digital literacy

- Performance-based measures of digital literacy in K-12 student samples
- 59 effect sizes from 24 studies and 31 countries (both ILSA and non-ILSA)
- Standardized mean differences (female-male) Hedges' g (`g`) and sampling variance (`Var.g`)
- Identifiers of effect sizes (`ESID`), primary studies (`IDSTUDY`), and countries (`IDCOUNTRY`)
- Sample sizes (`N`, `nF`, `nM`)
- Contextual variables: Publication year (`PubYear`), availability of individual participant data (`IPD`), countries' power distance index (`PDI`), and GDP (`cGCP`)

Research questions:

1. To what extent do boys and girls differ in their digital literacy performance?
2. To what extent do the gender effects vary across studies and countries?
3. Which contextual variables explain the possible heterogeneity in the effects?

Pooled effect size

Heterogeneity

Moderators

Standard Random-Effects Model

Meta-analytic baseline model assuming heterogeneity between effect sizes

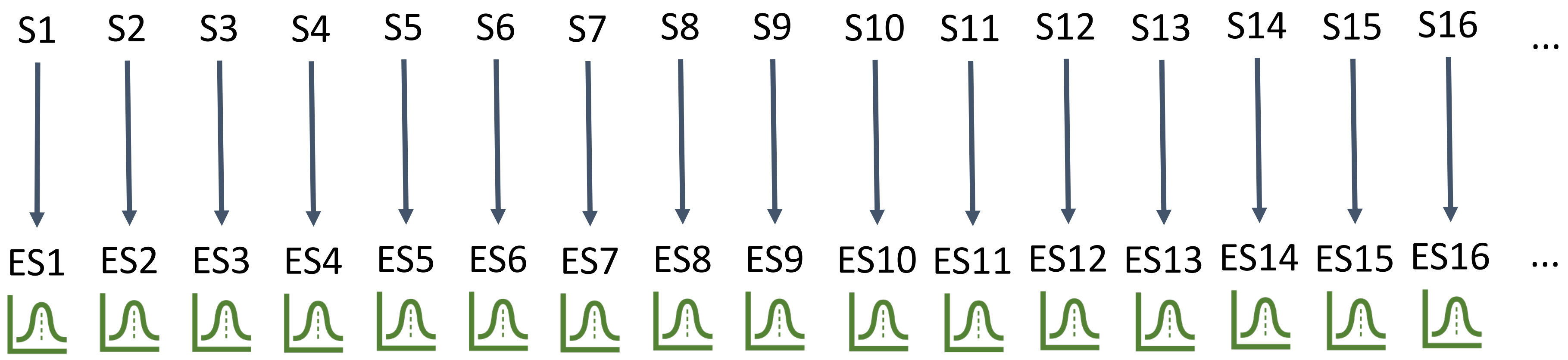
Standard Meta-Analytic Data Structure

(CMM, 2019)

Two-level hierarchical structure: the «ideal» scenario

Studies (S)
Between-study variance τ^2

Effect sizes (ES)
Sampling variance v_i



Level 1 is the level containing the sampling variances for each study (ES).
Level 2 is the level containing the heterogeneity between studies (S).

Standard Random-Effects Model

(Card, 2012; Cheung, 2015; Pastor & Lazowski, 2018)

Univariate random-effects model (REM)

for each study $i = 1, \dots, m$

Level 1:

$$\theta_i = f_i + e_i$$
$$e_i \sim N(0, v_i)$$

Level 2:

$$f_i = \beta_R + u_i$$
$$u_i \sim N(0, \tau^2)$$

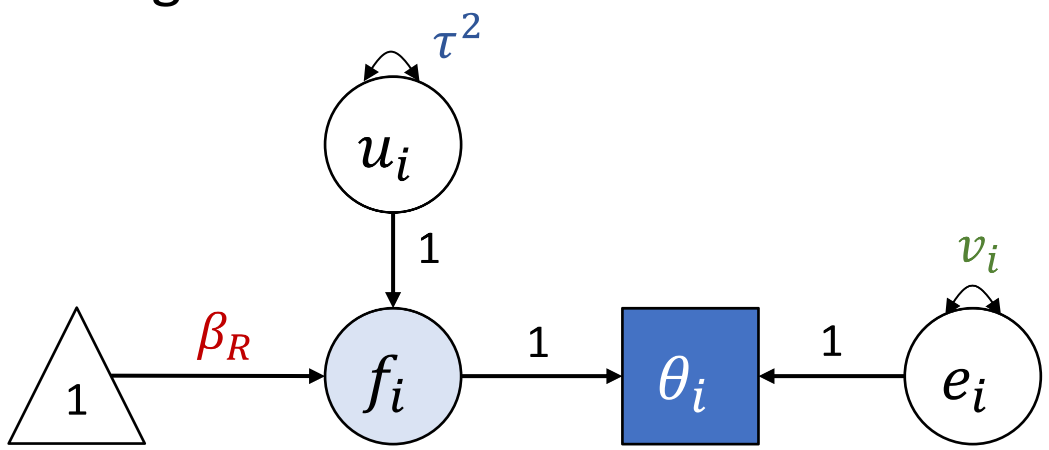
Total:

$$\theta_i = \beta_R + u_i + e_i$$

Weights:

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

Path diagram:



β_R: Weighted average effect size under the REM

Standard Random-Effects Model

Univariate REM

Model estimation in `metafor`

```
## Model estimation
REM <- rma.mv(g,
              Var.g,
              random = list(~ 1 | ESID),
              method = "REML",
              data = dat)

## Model summary
summary(REM, digits = 4)
```

Alternative specification:

```
REM2 <- rma(g, Var.g, data = dat, method = "REML")
summary(REM2)
```

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
##   26.7148   -53.4296   -49.4296   -45.3088   -49.2115
##
## Variance Components:
##
##              estim    sqrt  nlvls  fixed  factor
## sigma^2     0.0156  0.1247     59     no    ESID
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.1456  0.0178  8.1906  <.0001  0.1107  0.1804  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\tau^2 = 0.0156$ heterogeneity

$\bar{\theta} = 0.1456$ pooled effect size

Testing for Heterogeneity

(Higgins & Thomson, 2002; Sánchez-Meca & Marín-Martínez, 2008)

I^2 statistic

- The proportion of observed variation reflecting true variation between effect sizes.
- It only indicates what proportion of the variation between effect sizes is true variation.
- 25% small, 50% medium, 75% large heterogeneity

```
## Number of effect sizes
k <- REM$k
## Weights from the model
REM.wi <- 1/REM$vi
REM.vt <- (k-1) * sum(REM.wi) / (sum(REM.wi)^2 - sum(REM.wi^2))
100 * REM$sigma2 / (REM$sigma2 + REM.vt)
```

```
## [1] 91.72972
```

$I^2 = 91.7\%$ large amount of heterogeneity

Confidence interval CI of the variance indicating the heterogeneity between effect sizes τ^2

95% CI of τ^2

```
## Variances: 95 % confidence intervals
confint(REM, digits = 4)
```

```
##            estimate  ci.lb  ci.ub
## sigma^2      0.0156 0.0099 0.0253
## sigma        0.1247 0.0993 0.1591
```

Zero is not included.

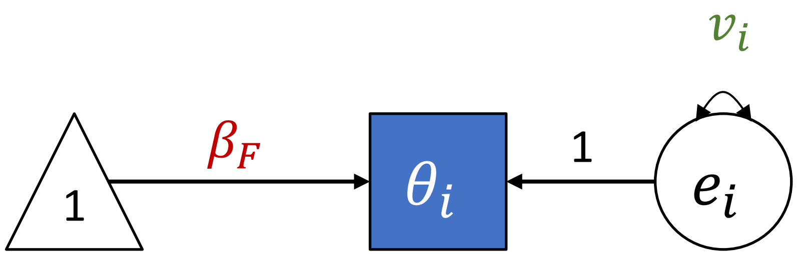
Testing for Heterogeneity

(Cheung, 2015; Hedges & Vevea 1998)

Model comparison

Fixed-effects model ($\tau^2 = 0$) vs. random-effects model (τ^2 freely estimated)

FEM



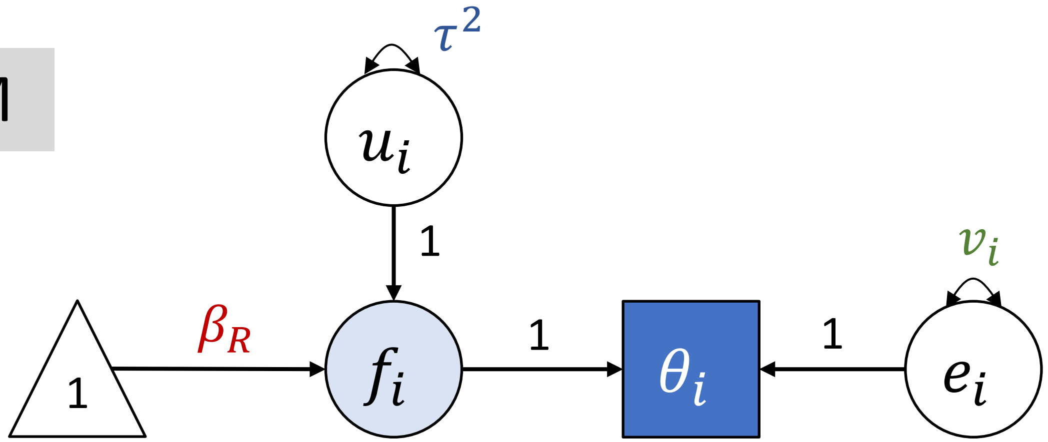
$$\theta_i = \beta_F + e_i$$

$$e_i \sim N(0, v_i)$$

Weights:

$$w_i^* = 1/v_i$$

REM



$$\theta_i = \beta_R + u_i + e_i$$

$$e_i \sim N(0, v_i), u_i \sim N(0, \tau^2)$$

Weights:

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

Testing for Heterogeneity

Model comparison

Random-effects model (τ^2 freely estimated) vs. fixed-effects model ($\tau^2 = 0$)

Estimate the FEM (in addition to the REM)

```
FEM <- rma.mv(g,
              Var.g,
              random = list(~ 1 | ESID),
              sigma2 = 0,  # No heterogeneity
              method = "REML",
              data = dat)
```

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -170.9519   341.9037   343.9037   345.9642   343.9752
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed  factor
## sigma^2    0.0000  0.0000     59    yes    ESID
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Model Results:
##
## estimate      se     zval    pval   ci.lb   ci.ub
##   0.1535  0.0049  31.6424  <.0001  0.1440  0.1630  ***
```

$\tau^2 = 0$ no heterogeneity

Testing for Heterogeneity

Model comparison

Random-effects model (τ^2 freely estimated) vs. fixed-effects model ($\tau^2 = 0$)

Perform a likelihood-ratio test and compare the information criteria

```
## Model comparison: REM vs. FEM
anova(REM, FEM)
```

```
##            df      AIC      BIC     AICc    logLik      LRT   pval       QE
## Full        2 -49.4296 -45.3088 -49.2115   26.7148                 592.4533
## Reduced     1 343.9037 345.9642 343.9752 -170.9519 395.3334 <.0001 592.4533
```

```
metafor::fitstats.rma(REM, FEM)
```

```
##                   REM       FEM
## logLik:      26.71482 -170.9519
## deviance:   -53.42964  341.9037
## AIC:        -49.42964  343.9037
## BIC:        -45.30875  345.9642
## AICc:       -49.21146  343.9752
```

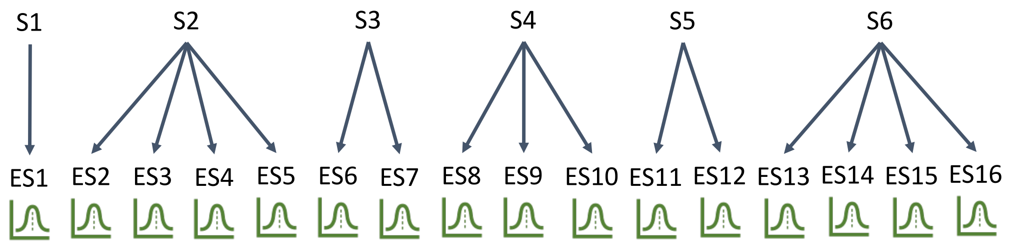
Result: The REM is preferred over the FEM. This is evidence for heterogeneity between effect sizes.

Effect Size Multiplicity

Multiple effects sizes available per study

Effect Size Multiplicity—Why?

Sources of multiple effect sizes per study ("Effect size multiplicity")



- Multiple populations or sub-populations (e.g., different samples or age groups)
- Multiple treatment or control groups (e.g., multiple intervention arms, active/passive controls)
- Multiple outcome variables (e.g., multiple constructs, different ways of measuring the same construct)
- Multiple time points (e.g., multiple post-tests)
- …

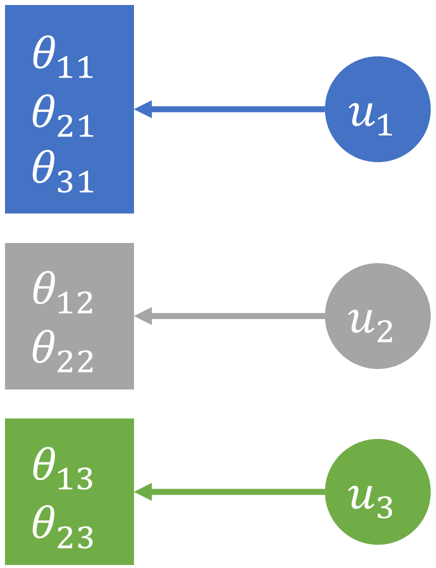
…or combinations ☺

(López-López et al., 2018, https://doi.org/10.1002/jrsm.1310)

Types of Effect Size Multiplicity

(Pustejovsky & Tipton, 2021, https://doi.org/10.1007/s11121-021-01246-3;
Illustration inspired by J. E. Pustejovsky, UseR! Oslo Talk, 02.09.2021)

Correlated Effects.

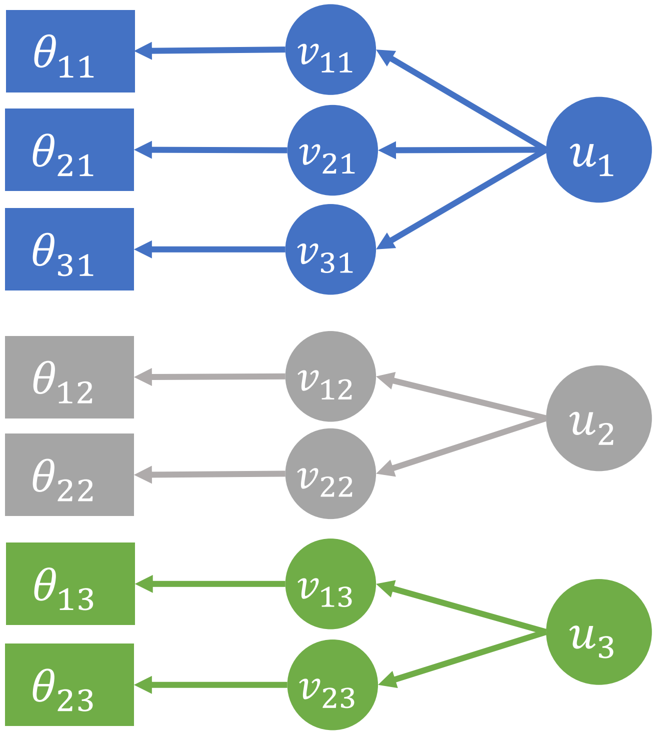


Correlated effect sizes
Between-study heterogeneity
No within-study heterogeneity

Examples: Multiple measures of the same construct, multiple measurement occasions of the same sample

Strategy
Specify the correlation between effect sizes

Hierarchical Effects.



Effect sizes not correlated
Between-study heterogeneity
Within-study heterogeneity

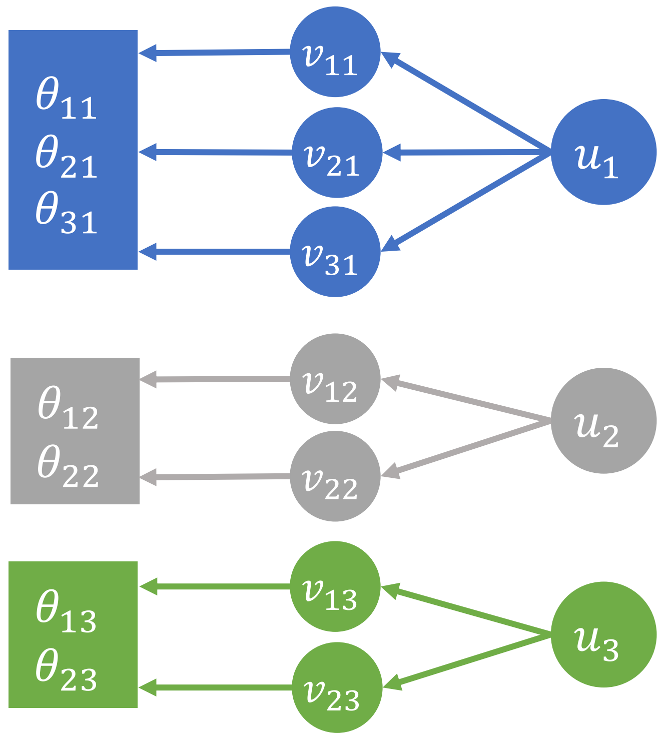
Examples: Multiple samples included in a study, multi-lab replications

Strategy
Estimate variances at different levels

Types of Effect Size Multiplicity

(Pustejovsky & Tipton, 2021, https://doi.org/10.1007/s11121-021-01246-3;
Illustration inspired by J. E. Pustejovsky, UseR! Oslo Talk, 02.09.2021)

Correlated *and* Hierarchical Effects.



Correlated effect sizes
Between-study heterogeneity
Within-study heterogeneity

Example: Multiple measures available from multiple samples within a study

Strategy
Estimate variances at different levels &
Specify the correlation between effect sizes

Multilevel Random-Effects Models

Model heterogeneity at different levels in the presence of hierarchical effect size multiplicity

Three-Level Hierarchical Structure

(CMM, 2019)

Effect sizes nested in primary studies

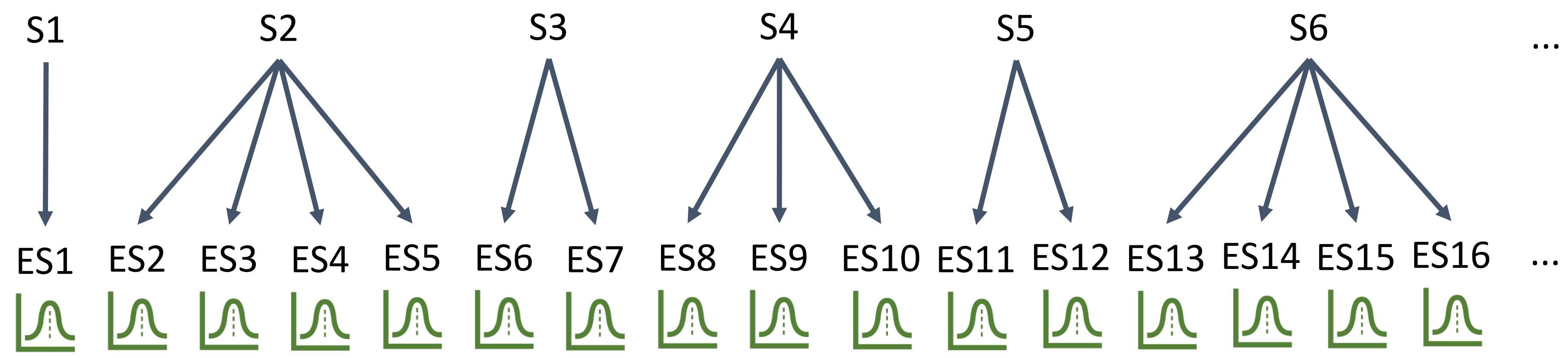
Studies (S)

Between-study variance $\tau^2_{(3)}$

Within-study variance $\tau^2_{(2)}$

Effect sizes (ES)

Sampling variance v_i



Typical meta-analytic structure if multiple independent samples are included in primary studies.

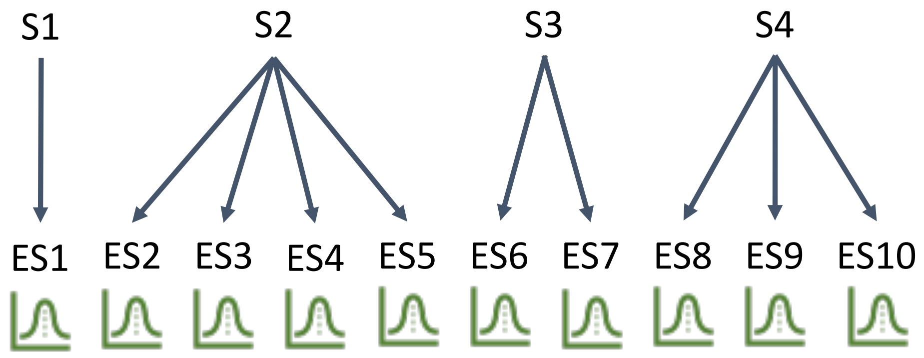
Multilevel Meta-Analysis

(Van den Noortgate et al., 2015)

Model the nested data structure

- Effect sizes $\theta_{ij}, i = 1, \ldots, k; j = 1, \ldots, m$ studies
- Sampling variances v_{ij}
- Multiple independent effect sizes nested in studies (e.g., multiple samples)

Main idea
Model the different levels of analysis and the respective variance components



Multilevel Meta-Analysis

(Cheung, 2015)

Three-level random-effects model (3LREM)

i: Effect sizes, j: Studies

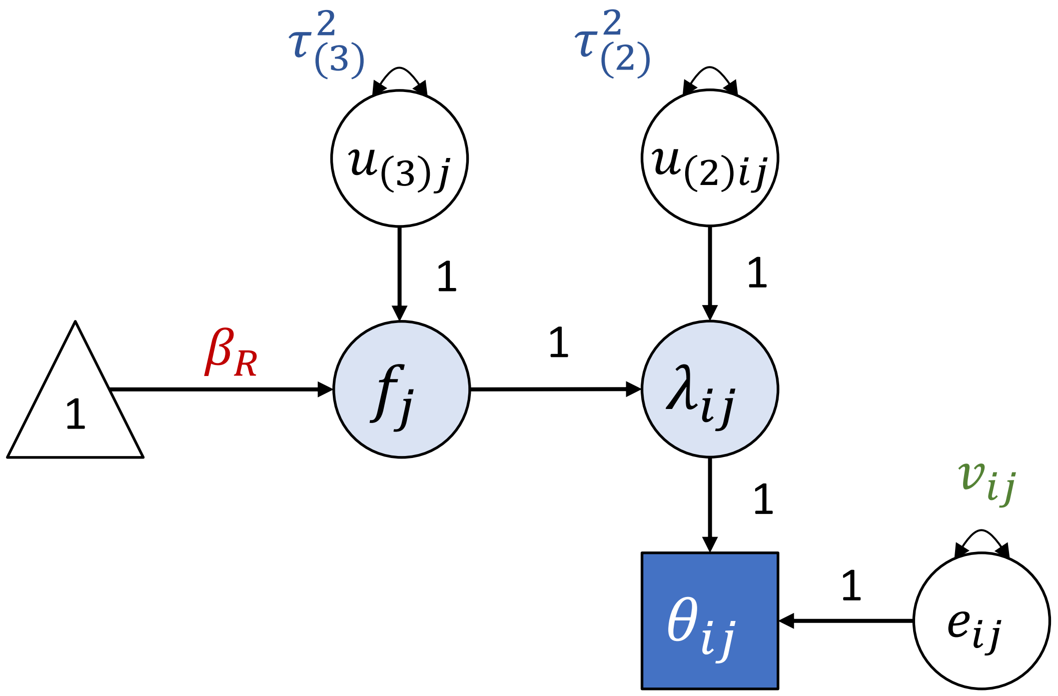
Level 1: $$\theta_{ij} = \lambda_{ij} + e_{ij}$$
$$e_{ij} \sim N(0, v_{ij})$$

Level 2: $$\lambda_{ij} = f_j + u_{(2)ij}$$
$$u_{(2)ij} \sim N(0, \tau^2_{(2)})$$

Level 3: $$f_j = \beta_R + u_{(3)j}$$
$$u_{(3)j} \sim N(0, \tau^2_{(3)})$$

Total: $$\theta_{ij} = \beta_R + u_{(2)ij} + u_{(3)j} + e_{ij}$$

Path diagram:



β_R: Weighted average population effect size under the 3LREM

Multilevel Meta-Analysis

Three-level random-effects model

Effect sizes nested in studies

```
REM3a <- rma.mv(g,
                Var.g,
                random = list(~ 1 | IDSTUDY/ESID),
                method = "REML",
                data = dat,
                tdist = TRUE,
                test = "t")

## Model summary
summary(REM3a, digits = 4)
```

Level 1: Sampling variation

Level 2: Within-study heterogeneity variance

Level 3: Between-study heterogeneity variance

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
##   31.6963   -63.3927   -57.3927   -51.2113   -56.9482
##
## Variance Components:
##
##                estim    sqrt  nlvls  fixed        factor
## sigma^2.1     0.0268  0.1637     24     no       IDSTUDY
## sigma^2.2     0.0073  0.0852     59     no  IDSTUDY/ESID
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Model Results:
##
## estimate      se    tval  df    pval   ci.lb   ci.ub
##   0.1291  0.0408  3.1643  58  0.0025  0.0474  0.2108  **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\tau^2_{(3)}$ (sigma^2.1)

$\tau^2_{(2)}$ (sigma^2.2)

β_R (estimate)

Multilevel Meta-Analysis

Three-level random-effects model

Effect sizes nested in studies

```
## Confidence intervals of the variance component(s)
confint.rma.mv(REM3a)
```

```
##             estimate  ci.lb  ci.ub
## sigma^2.1     0.0268 0.0072 0.0707
## sigma.1       0.1637 0.0851 0.2659
##
##             estimate  ci.lb  ci.ub
## sigma^2.2     0.0073 0.0041 0.0134
## sigma.2       0.0852 0.0641 0.1157
```

```
## Visualization of the variance components
REM3a.var <- dmetar::var.comp(REM3a)
plot(REM3a.var)
```

Multilevel Meta-Analysis

Three-level random-effects model

Effect sizes nested in studies

Obtain cluster-robust standard errors to avoid misspecification in the error structure

```
## Cluster-robust standard errors
robust(REM3a, cluster = IDSTUDY, clubSandwich = TRUE)
```

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
## Variance Components:
##
##                estim    sqrt  nlvls  fixed        factor
## sigma^2.1   0.0268  0.1637     24     no       IDSTUDY
## sigma^2.2   0.0073  0.0852     59     no  IDSTUDY/ESID
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Number of estimates:   59
## Number of clusters:    24
## Estimates per cluster: 1-21 (mean: 2.46, median: 1)
##
## Model Results:
##
## estimate      se¹     tval¹     df¹     pval¹    ci.lb¹    ci.ub¹
##   0.1291  0.0407   3.1762   21.22   0.0045   0.0446   0.2136   **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## 1) results based on cluster-robust inference (var-cov estimator: CR2,
##    approx t-test and confidence interval, df: Satterthwaite approx)
```

β_R

Multilevel Meta-Analysis

Three-level random-effects model

Effect sizes nested in countries

```r
## Model estimation
REM3b <- rma.mv(g,
                Var.g,
                random = list(~ 1 | IDCOUNTRY/ESID),
                method = "REML",
                data = dat,
                tdist = TRUE,
                test = "t")
```

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
##   26.7718   -53.5436   -47.5436   -41.3623   -47.0992
##
## Variance Components:
##
##              estim    sqrt  nlvls  fixed           factor
## sigma^2.1   0.0009  0.0297     31     no        IDCOUNTRY
## sigma^2.2   0.0144  0.1199     59     no   IDCOUNTRY/ESID
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Model Results:
##
## estimate      se    tval  df    pval   ci.lb   ci.ub
##   0.1457  0.0184  7.9075  58  <.0001  0.1088  0.1826  ***
```

$\tau^2_{(3)}$ (sigma^2.1), $\tau^2_{(2)}$ (sigma^2.2), β_R (estimate)

Level 1: Sampling variation

Level 2: Within-country heterogeneity variance

Level 3: Between-country heterogeneity variance

Multilevel Meta-Analysis

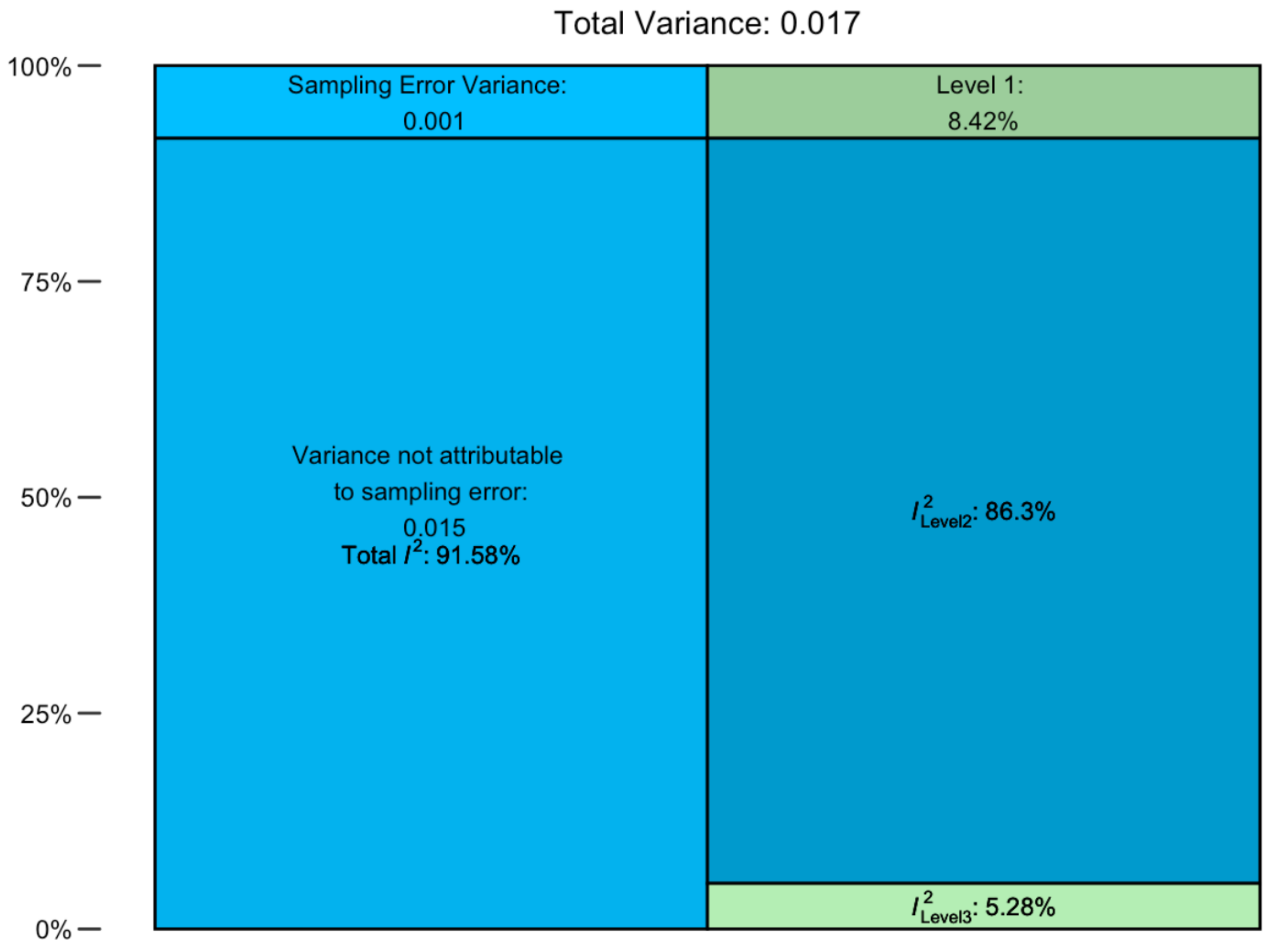
Three-level random-effects model

Effect sizes nested in countries

```
## Confidence intervals of the variance component(s)
confint.rma.mv(REM3b)
```

```
##              estimate  ci.lb  ci.ub
## sigma^2.1    0.0009 0.0000 0.0088
## sigma.1      0.0297 0.0000 0.0940
##
##              estimate  ci.lb  ci.ub
## sigma^2.2    0.0144 0.0068 0.0251
## sigma.2      0.1199 0.0824 0.1586
```

```
## Visualization of the variance components
REM3b.var <- dmetar::var.comp(REM3b)
plot(REM3b.var)
```



Cross-Classified Data Structure

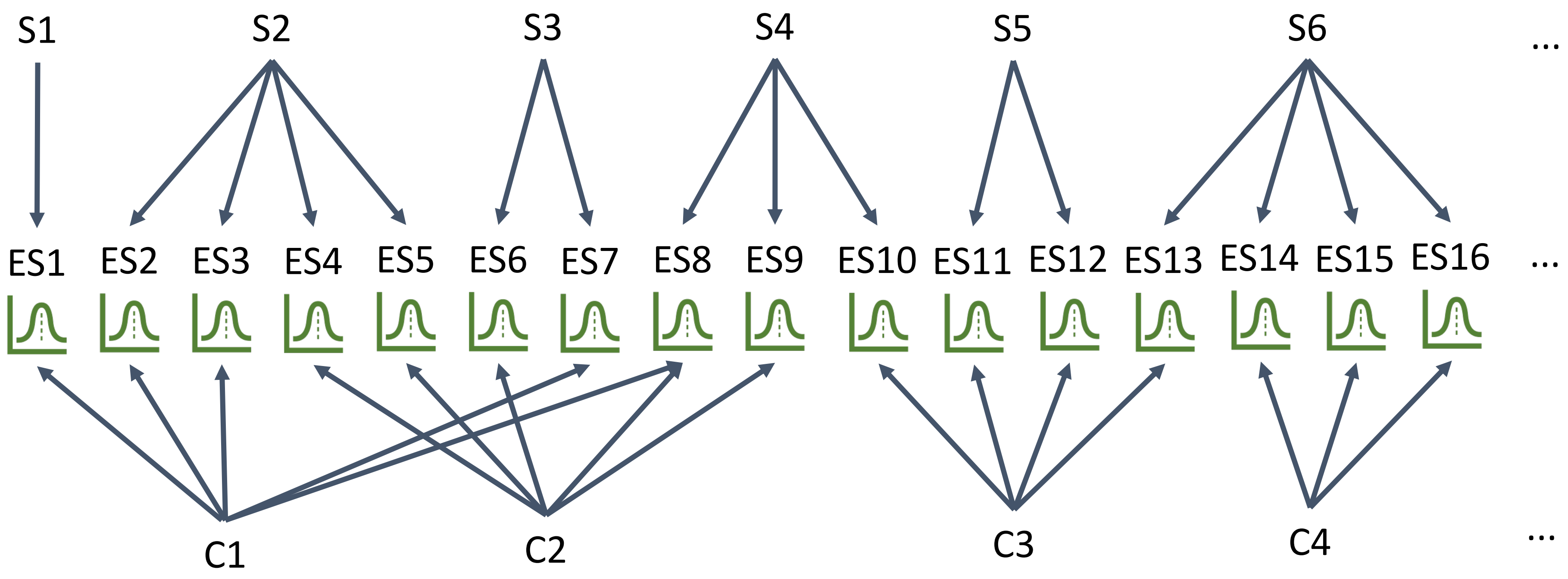
(CMM, 2019)

Four-level non-hierarchical structure with cross-classification

Studies (S)
Between-study variance $\tau^2_{(3)}$
Within-study variance $\tau^2_{(2)}$

Effect sizes (ES)
Sampling variance v_i

Countries (C)
Between-country variance $\tau^2_{(4)}$



Typical meta-analytic structure if ILSA data are included.

Cross-Classified Random-Effects Model

Four-level random-effects model

Effect sizes nested in studies and countries

```r
## Model estimation
CCREM <- rma.mv(g,
                Var.g,
                random = list(~ 1 | IDSTUDY/ESID,
                              ~ 1 | IDCOUNTRY),
                method = "REML",
                data = dat,
                tdist = TRUE,
                test = "t",
                time = TRUE)
```

Independent nesting of effect sizes in studies and countries

(More details about CCREMs: Fernández-Castilla et al., 2019)

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
##   33.8044   -67.6088   -59.6088   -51.3670   -58.8541
##
## Variance Components:
##
##               estim    sqrt  nlvls  fixed        factor
## sigma^2.1    0.0308  0.1754     24     no       IDSTUDY
## sigma^2.2    0.0016  0.0397     59     no  IDSTUDY/ESID
## sigma^2.3    0.0055  0.0743     31     no     IDCOUNTRY
##
## Test for Heterogeneity:
## Q(df = 58) = 592.4533, p-val < .0001
##
## Model Results:
##
## estimate      se    tval  df    pval   ci.lb   ci.ub
##   0.0951  0.0444  2.1424  58  0.0364  0.0062  0.1839  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\tau^2_{(3)}$ (sigma^2.1), $\tau^2_{(2)}$ (sigma^2.2), $\tau^2_{(4)}$ (sigma^2.3)

β_R (estimate)

Cross-Classified Random-Effects Model

Four-level random-effects model

Effect sizes nested in studies and countries

```
## Confidence intervals of the variance component(s)
confint.rma.mv(CCREM)
```

```
##              estimate  ci.lb  ci.ub
## sigma^2.1    0.0308 0.0112 0.0770
## sigma.1      0.1754 0.1057 0.2775
##
##              estimate  ci.lb  ci.ub
## sigma^2.2    0.0016 0.0000 0.0079
## sigma.2      0.0397 0.0000 0.0886
##
##              estimate  ci.lb  ci.ub
## sigma^2.3    0.0055 0.0003 0.0122
## sigma.3      0.0743 0.0175 0.1105
```

Level-specific I^2

```
## [1] 78.371363  4.008578 14.047175
```

$I^2_{(3)}$	$I^2_{(2)}$	$I^2_{(4)}$
Between studies	Within studies	Between countries

Model Selection

Decide on a meta-analytic baseline model

Possible Criteria for Model Selection

Proportion of Variance Components

A well-fitting model should have a reasonable distribution of variance across the levels. Small proportions of variance components suggest that this level may not be needed.

Information Criteria

Lower values of the AIC or BIC suggest a better-fitting model.

Likelihood-Ratio Test (LRT)

Direct comparison of nested models. An insignificant LRT suggests that the simpler model (with fewer levels) might be sufficient.

Conceptual Considerations

Nature of effect size multiplicity, research questions and goals.

Proportion of Variance Components

Weighted average effect sizes and variance estimates

	Standard REM	3LREM/Studies	3LREM/Countries	CCREM
$\bar{g}$	0.146	0.129	0.146	0.095
95 % CI	[0.110, 0.181]	[0.045, 0.214]	[0.108, 0.183]	[0.006, 0.184]
Heterogeneity variances	τ^2=0.016 (between effect sizes)	$\tau^2_{(3)}$=0.027 (between studies) $\tau^2_{(2)}$=0.007 (within studies)	$\tau^2_{(3)}$=0.001 (between countries) $\tau^2_{(2)}$=0.014 (within countries)	$\tau^2_{(3)}$=0.031 (between studies) $\tau^2_{(2)}$=0.002 (within studies) $\tau^2_{(4)}$=0.006 (between countries)

Information Criteria

AIC and BIC for the random-effects models

	Standard REM	3LREM/Studies	3LREM/Countries	CCREM
logLik	26.71482	31.69633	26.77180	33.80439
AIC	-49.42964	-57.39265	-47.54360	-59.60879
BIC	-45.30875	-51.21133	-41.36227	-51.36702

Result: Model `CCREM` shows the lowest AIC and BIC.

Likelihood-Ratio Tests of Nested Models

LRTs between the random-effects models

(a) REM3a preferred over REM

(b) REM preferred over REM3b

(c) CCREM preferred over REM3a

```
## Likelihood-ratio tests
## Standard REM vs. three-level REMs
metafor::anova.rma(REM, REM3a)
```

```
##
##         df      AIC      BIC     AICc  logLik    LRT   pval       QE
## Full     3 -57.3927 -51.2113 -56.9482 31.6963                592.4533
## Reduced  2 -49.4296 -45.3088 -49.2115 26.7148 9.9630 0.0016 592.4533
```

```
metafor::anova.rma(REM, REM3b)
```

```
##
##         df      AIC      BIC     AICc  logLik    LRT   pval       QE
## Full     3 -47.5436 -41.3623 -47.0992 26.7718                592.4533
## Reduced  2 -49.4296 -45.3088 -49.2115 26.7148 0.1140 0.7357 592.4533
```

```
## Three-level REM vs. CCREM
metafor::anova.rma(REM3a, CCREM)
```

Result: Model CCREM may serve as a baseline model.

```
##
##         df      AIC      BIC     AICc  logLik    LRT   pval       QE
## Full     4 -59.6088 -51.3670 -58.8541 33.8044                592.4533
## Reduced  3 -57.3927 -51.2113 -56.9482 31.6963 4.2161 0.0400 592.4533
```

Moderator Analyses

Mixed-effects meta-regression models to explain heterogeneity

Moderator Analysis

(Cheung, 2015)

Three-level mixed-effects meta-regression

i: Effect sizes, j: Studies

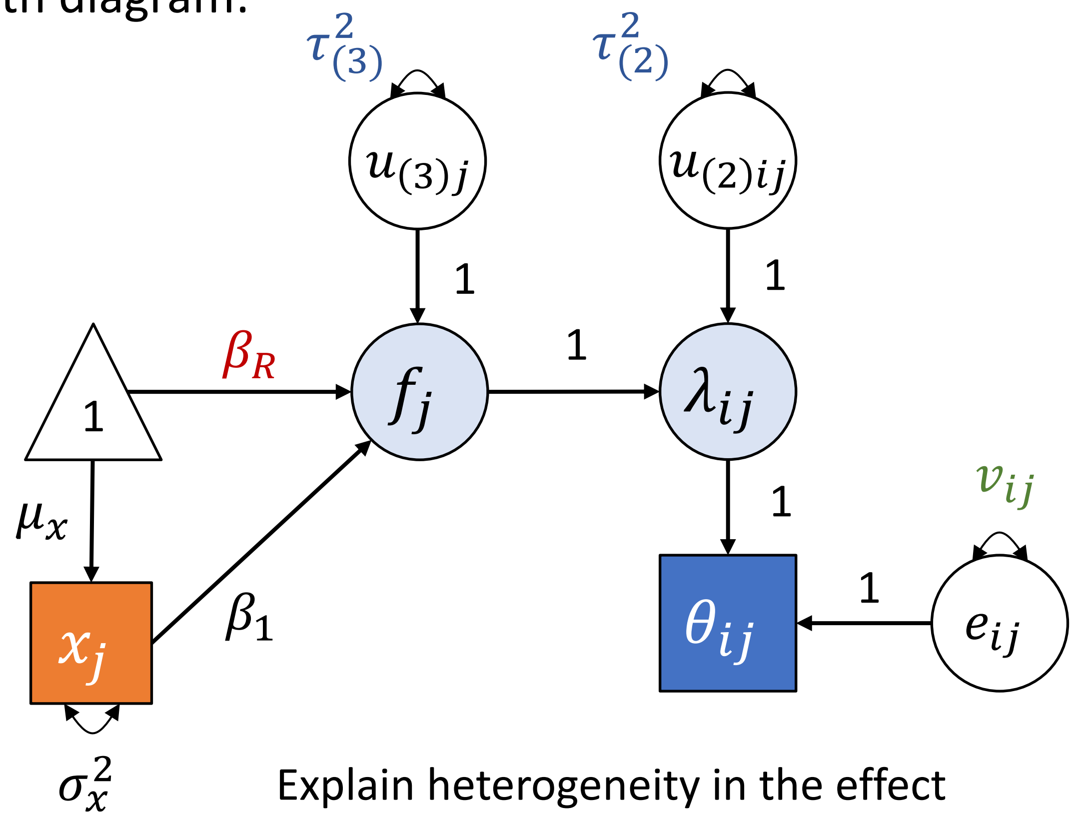
Level 1: $$\theta_{ij} = \lambda_{ij} + e_{ij}$$
$$e_{ij} \sim N(0, v_{ij})$$

Level 2: $$\lambda_{ij} = f_j + u_{(2)ij}$$
$$u_{(2)ij} \sim N(0, \tau^2_{(2)})$$

Level 3: $$f_j = \beta_R + \beta_1 x_j + u_{(3)j}$$
$$u_{(3)j} \sim N(0, \tau^2_{(3)})$$

Total: $$\theta_{ij} = \beta_R + \beta_1 x_j + u_{(2)ij} + u_{(3)j} + e_{ij}$$

Path diagram:



Explain heterogeneity in the effect sizes by a moderating variable x_j

Moderator Analysis

Mixed-effects meta-regression

Moderation by the availability of IPD (binary variable)

```
## Mixed-effects meta-regression model
CCREM.ipd <- rma.mv(g,
                    Var.g,
                    random = list(~ 1 | IDSTUDY/ESID,
                                  ~ 1 | IDCOUNTRY),
                    method = "REML",
                    data = dat,
                    tdist = TRUE,
                    test = "t",
                    mods =~ IPD)
```

Result: No evidence on the differences in effect sizes between studies with and without IPD.

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##   logLik  Deviance       AIC       BIC      AICc
##  34.2303  -68.4606  -58.4606  -48.2454  -57.2842
##
## Variance Components:
##
##              estim    sqrt  nlvls  fixed        factor
## sigma^2.1   0.0330  0.1818     24     no       IDSTUDY
## sigma^2.2   0.0015  0.0389     59     no  IDSTUDY/ESID
## sigma^2.3   0.0056  0.0748     31     no     IDCOUNTRY
##
## Test for Residual Heterogeneity:
## QE(df = 57) = 578.8574, p-val < .0001
##
## Test of Moderators (coefficient 2):
## F(df1 = 1, df2 = 57) = 0.3452, p-val = 0.5592
##
## Model Results:
##
##          estimate      se    tval  df    pval    ci.lb   ci.ub
## intrcpt    0.0854  0.0481  1.7753  57  0.0812  -0.0109  0.1816  .
## IPD        0.0805  0.1370  0.5875  57  0.5592  -0.1939  0.3549
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

β_1

Moderator Analysis

Mixed-effects meta-regression

Moderation by the availability of IPD (binary variable)

Model estimation without the intercept

```
CCREM.ipd2 <- rma.mv(g,
                    Var.g,
                    random = list(~ 1 | IDSTUDY/ESID,
                                  ~ 1 | IDCOUNTRY),
                    method = "REML",
                    data = dat,
                    tdist = TRUE,
                    test = "t",
                    mods =~ factor(IPD) - 1 # remove the intercept
                    )
```

This way, the weighted average effect sizes for each group are estimated directly (assuming the same amount of heterogeneity for each group).

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik  Deviance       AIC       BIC      AICc
##  34.2303  -68.4606  -58.4606  -48.2454  -57.2842
##
## Variance Components:
##
##              estim    sqrt  nlvls  fixed        factor
## sigma^2.1  0.0330  0.1818     24     no       IDSTUDY
## sigma^2.2  0.0015  0.0389     59     no  IDSTUDY/ESID
## sigma^2.3  0.0056  0.0748     31     no     IDCOUNTRY
##
## Test for Residual Heterogeneity:
## QE(df = 57) = 578.8574, p-val < .0001
##
## Test of Moderators (coefficients 1:2):
## F(df1 = 2, df2 = 57) = 2.3311, p-val = 0.1064
##
## Model Results:
##
##               estimate      se    tval  df    pval    ci.lb   ci.ub
## factor(IPD)0    0.0854  0.0481  1.7753  57  0.0812  -0.0109  0.1816  .
## factor(IPD)1    0.1659  0.1297  1.2789  57  0.2061  -0.0938  0.4256
```

Moderator Analysis

Mixed-effects meta-regression

Moderation by the PDI (continuous variable)

```r
## Mixed-effects meta-regression model
CCREM.pdi <- rma.mv(g,
                    Var.g,
                    random = list(~ 1 | IDSTUDY/ESID,
                                  ~ 1 | IDCOUNTRY),
                    method = "REML",
                    data = dat,
                    tdist = TRUE,
                    test = "t",
                    mods =~ scale(PDI, ## standardized
                                  center = TRUE,
                                  scale = TRUE))
```

Result: Some evidence on the negative relation between the effects and countries' `PDI`.

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik  Deviance        AIC        BIC       AICc
##  34.3599  -68.7198   -58.7198   -48.5046   -57.5434
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed        factor
## sigma^2.1  0.0321  0.1792     24     no       IDSTUDY
## sigma^2.2  0.0015  0.0385     59     no  IDSTUDY/ESID
## sigma^2.3  0.0048  0.0694     31     no     IDCOUNTRY
##
## Test for Residual Heterogeneity:
## QE(df = 57) = 592.4506, p-val < .0001
##
## Test of Moderators (coefficient 2):
## F(df1 = 1, df2 = 57) = 3.7615, p-val = 0.0574
##
## Model Results:
##
##                                    estimate      se     tval  df    pval
## intrcpt                              0.1001  0.0447   2.2403  57  0.0290
## scale(PDI, center = TRUE, scale = TRUE)  -0.0292  0.0150  -1.9395  57  0.0574
##                                       ci.lb   ci.ub
## intrcpt                              0.0106  0.1896  *
## scale(PDI, center = TRUE, scale = TRUE)  -0.0593  0.0009  .
```

β_1

Moderator Analysis

Mixed-effects meta-regression

Moderation by the PDI (continuous variable)

Level-specific variance explanation as the proportional reduction of variance (R^2)

```
## Level: Effect sizes
max(1-CCREM.pdi$sigma2[2]/CCREM$sigma2[2] , 0)
```

```
## [1] 0.05710977
```

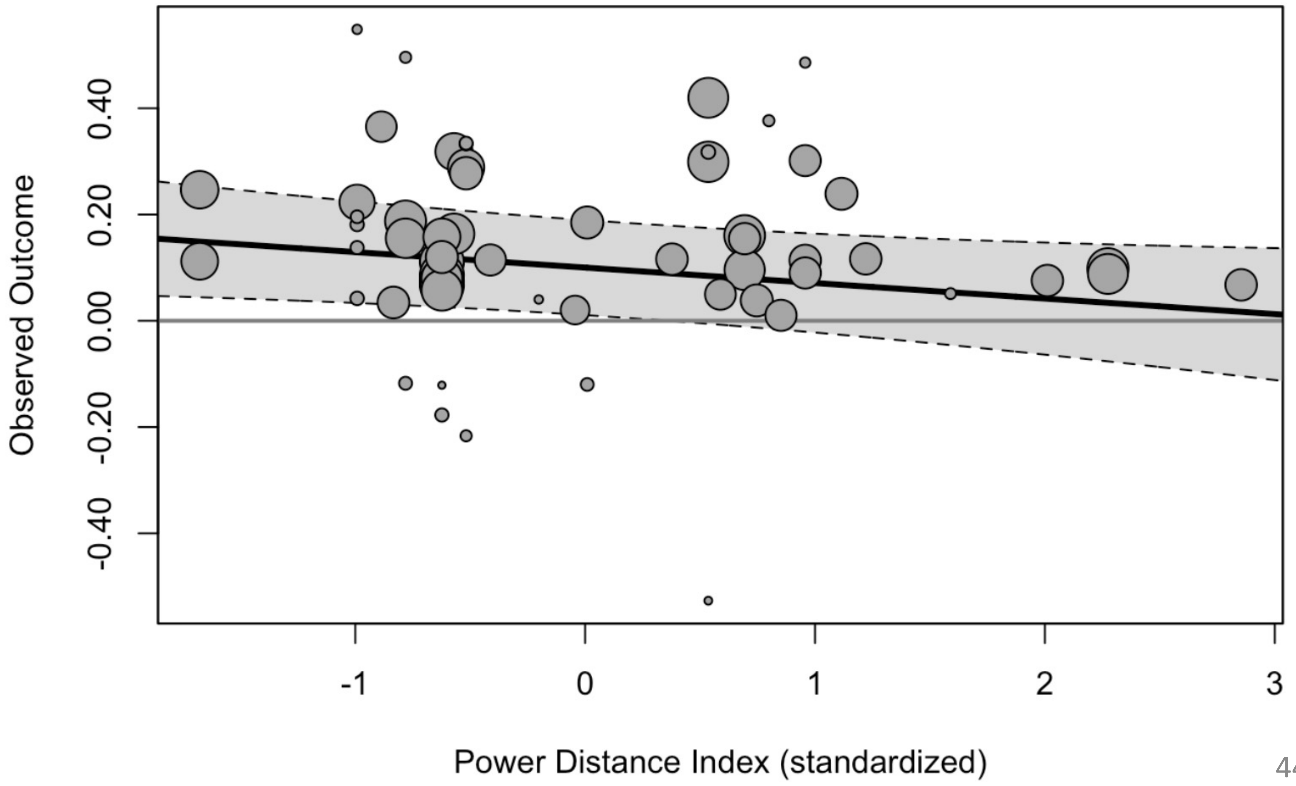
```
## Level: Primary studies
max(1-CCREM.pdi$sigma2[1]/CCREM$sigma2[1] , 0)
```

```
## [1] 0
```

```
## Level: Countries
max(1-CCREM.pdi$sigma2[3]/CCREM$sigma2[3] , 0)
```

```
## [1] 0.1275372
```

```
## Regression plot
regplot(CCREM.pdi,
        xlab = "Power Distance Index (standardized)",
        refline = 0,
        digits = 2)
```



Publication Bias and Influential Effect Sizes

Ways of quality assurance and testing the robustness/sensitivity of findings

Publication Bias

Publication Bias is a Form of Non-Reporting Bias.

The probability of publishing a primary study is affected by its results: Statistically significant or hypothesis-confirming results are more likely to be published (see Harrer et al., 2021, chap. 9).

Detecting Publication Bias

- Funnel plot asymmetry and Egger's regression test
- Precision-effect test (PET) and precision-effect estimate with standard errors (PEESE)
- Funnel plot test
- Begg's correlation test
- Trim-and-fill analyses with the estimators L_0^+ and R_0^+
- Moderation by publication year and/or publication status (e.g., grey vs. published)
- Worst-case sensitivity analyses
- …

(e.g., Fernández-Castilla et al., 2021, https://doi.org/10.1080/00220973.2019.1582470;
Harrer et al., 2021, https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/pub-bias.html)

Detecting Publication Bias

Funnel Plot Symmetry

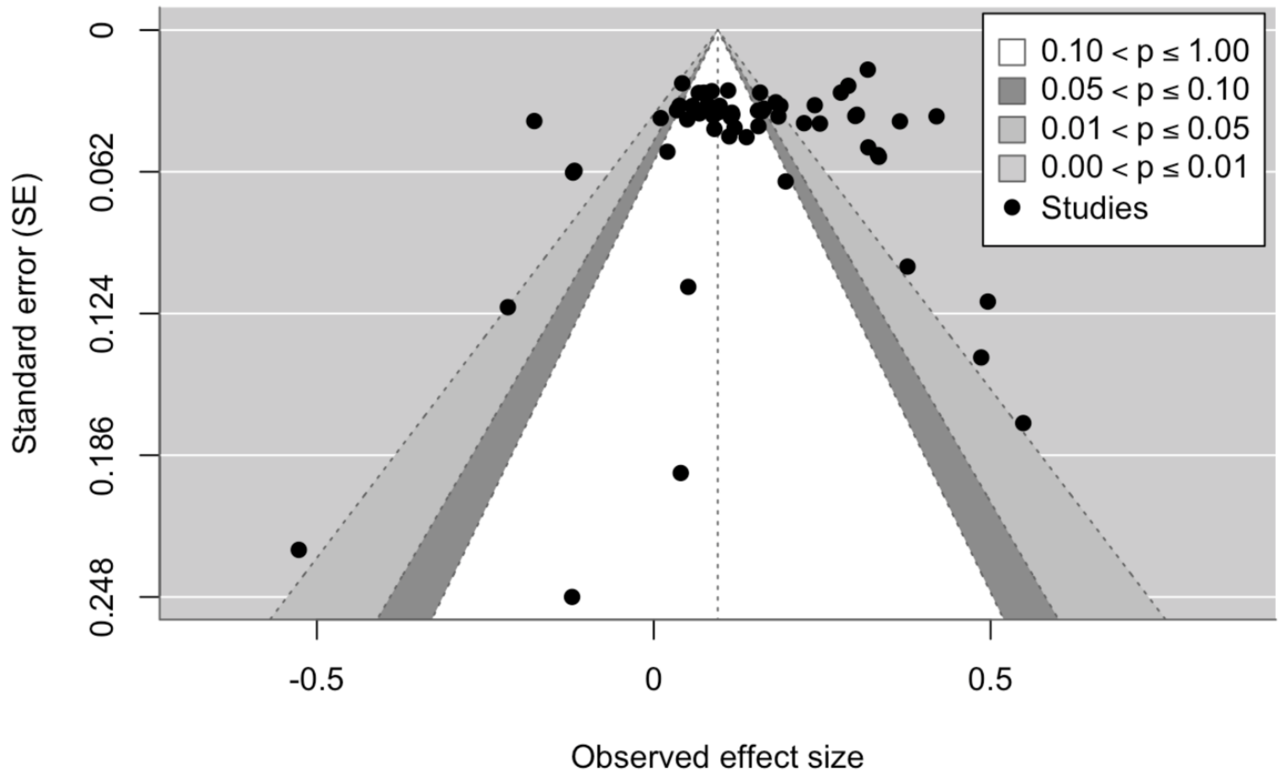
Plot the observed effect sizes against their standard errors to examine small-study effects

Graphical inspection of the plot asymmetry

Publication bias might be indicated by an asymmetric funnel plot.

```
metafor::funnel(CCREM,
                yaxis = "sei",
                level = c(90, 95, 99),
                shade = c("white", "gray55", "gray75"),
                legend = TRUE,
                xlab = "Observed effect size",
                ylab = "Standard error (SE)")
```

Contour-enhanced funnel plot



Detecting Publication Bias

Egger's Regression Test

Test the asymmetry of the funnel plot via a linear regression of the scaled effect sizes (z-score) on the precision ($1/SE$):

$$\frac{\hat{\theta}_i}{SE_{\hat{\theta}_i}} = \beta_0 + \beta_1 \frac{1}{SE_{\hat{\theta}_i}} + r_i$$

$$r_i \sim N(0, \sigma_r^2)$$

Test $\hat{\beta}_0$ against zero.

Significant $\hat{\beta}_0$ indicates funnel plot asymmetry.

Original Egger's test (Egger et al., 1997)

```
dat %>%
  mutate(y = g/sqrt(Var.g), x = 1/sqrt(Var.g)) %>%
  lm(y ~ x, data = .) %>%
  summary()
```

Modified Egger's test (Pustejovsky & Rodgers, 2019)

```
dat$SEg_modified <- sqrt((dat$nF + dat$nM)/(dat$nF*dat$nM))

dat %>%
  mutate(y = g/SEg_modified, x = 1/SEg_modified) %>%
  lm(y ~ x, data = .) %>%
  summary()
```

Detecting Publication Bias

Egger's Regression Test

Test the asymmetry of the funnel plot via a linear regression of the scaled effect sizes (z-score) on the precision ($1/SE$):

$$\frac{\hat{\theta}_i}{SE_{\hat{\theta}_i}} = \beta_0 + \beta_1 \frac{1}{SE_{\hat{\theta}_i}} + r_i$$

Test $\hat{\beta}_0$ against zero.

Significant $\hat{\beta}_0$ indicates funnel plot asymmetry.

Original Egger's test (Egger et al., 1997)

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1390     1.0976  -1.038    0.304
## x             0.1945     0.0414   4.698  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.149 on 57 degrees of freedom
## Multiple R-squared:  0.2791, Adjusted R-squared:  0.2665
## F-statistic: 22.07 on 1 and 57 DF,  p-value: 1.701e-05
```

Result: No evidence of publication bias.

Detecting Publication Bias

Precision-effect test (PET) and Precision-effect estimate with standard error (PEESE)

Control for the sampling error (PET) or variance (PEESE) in the weighted average effect size and extract a limit effect (i.e., the effect with $SE = 0$).

Extract the intercepts from the two regressions:
PET estimate: $\hat{\theta}_{\text{PET}} = \hat{\beta}_{0_{\text{PET}}}$
PEESE estimate: $\hat{\theta}_{\text{PEESE}} = \hat{\beta}_{0_{\text{PEESE}}}$

Decision for an overall (controlled) estimate:

$$\hat{\theta}_{\text{PEESE}} = \begin{cases} P(\hat{\beta}_{0_{\text{PET}}} = 0) < 0.1 \text{ and } \hat{\beta}_{0_{\text{PET}}} > 0\text{:} \ \hat{\beta}_{0_{\text{PEESE}}} \\ \text{else: } \hat{\beta}_{0_{\text{PET}}} \end{cases}$$

(Harrer et al., 2021, chap. 9; Stanley et al., 2014)

```
## Multilevel MEM
CCREM.pet <- rma.mv(g,
                    Var.g,
                    random = list(~ 1 | IDSTUDY/ESID,
                                  ~ 1 | IDCOUNTRY),
                    method = "REML",
                    data = dat,
                    tdist = TRUE,
                    test = "t",
                    mods =~ sqrt(Var.g))

# Summarize the results
summary(CCREM.pet, digits=4)
```

For PEESE: `mods =~ Var.g`

Detecting Publication Bias

Precision-effect test (PET) and Precision-effect estimate with standard error (PEESE)

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik  Deviance       AIC       BIC      AICc
##  33.4927  -66.9854  -56.9854  -46.7702  -55.8090
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed        factor
## sigma^2.1  0.0330  0.1816     24     no       IDSTUDY
## sigma^2.2  0.0016  0.0401     59     no  IDSTUDY/ESID
## sigma^2.3  0.0055  0.0740     31     no     IDCOUNTRY
##
## Test for Residual Heterogeneity:
## QE(df = 57) = 583.2291, p-val < .0001
##
## Test of Moderators (coefficient 2):
## F(df1 = 1, df2 = 57) = 0.0872, p-val = 0.7688
##
## Model Results:
##
##              estimate      se     tval  df    pval    ci.lb   ci.ub
## intrcpt        0.1111  0.0706   1.5732  57  0.1212  -0.0303  0.2525
## sqrt(Var.g)   -0.2225  0.7534  -0.2953  57  0.7688  -1.7312  1.2862
```

$$\hat{\beta}_{0_{\mathrm{PET}}} = 0.11, p = .12$$

Result: Decide for the PET estimate.

```
## Multivariate Meta-Analysis Model (k = 59; method: REML)
##
##    logLik  Deviance       AIC       BIC      AICc
##  33.8299  -67.6599  -57.6599  -47.4446  -56.4834
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed        factor
## sigma^2.1  0.0329  0.1814     24     no       IDSTUDY
## sigma^2.2  0.0016  0.0401     59     no  IDSTUDY/ESID
## sigma^2.3  0.0055  0.0739     31     no     IDCOUNTRY
##
## Test for Residual Heterogeneity:
## QE(df = 57) = 589.7801, p-val < .0001
##
## Test of Moderators (coefficient 2):
## F(df1 = 1, df2 = 57) = 0.6443, p-val = 0.4255
##
## Model Results:
##
##          estimate      se     tval  df    pval    ci.lb   ci.ub
## intrcpt    0.1167  0.0528   2.2119  57  0.0310   0.0110  0.2223  *
## Var.g     -2.6613  3.3155  -0.8027  57  0.4255  -9.3006  3.9779
```

$$\hat{\beta}_{0_{\mathrm{PEESE}}} = 0.12, p = .03$$

Detecting Publication Bias

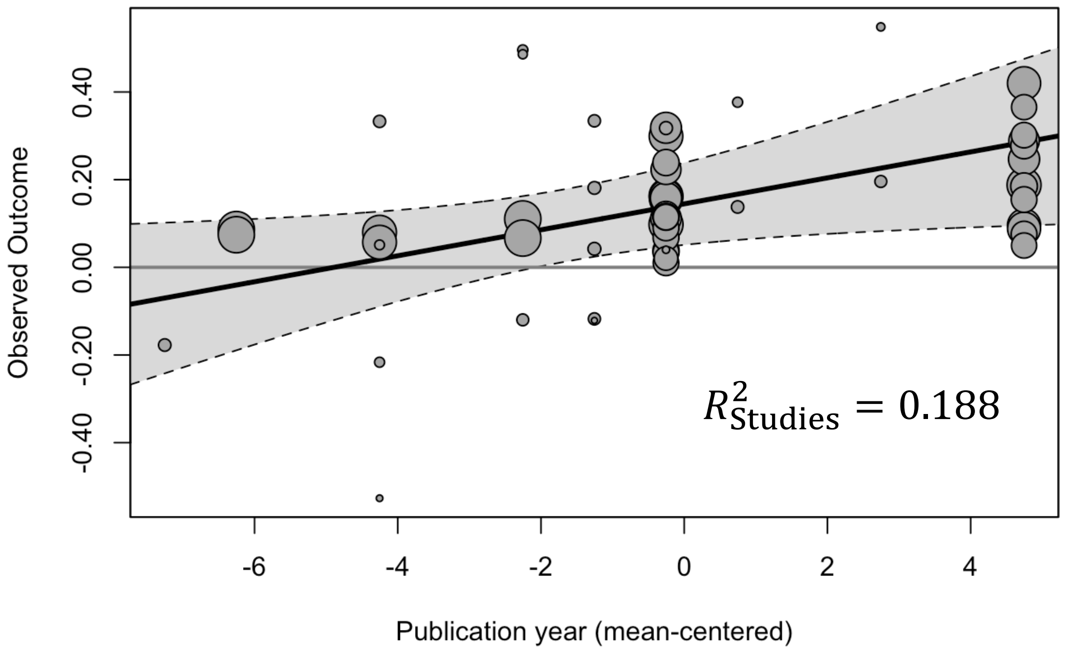
Moderation by publication characteristics

Examine the relation between publication characteristics (e.g., type of publication, peer-review status, publication year) and the effects.

Significant relations may indicate publication bias.

```r
## Mixed-effects meta-regression model
CCREM.pub <- rma.mv(g,
                    Var.g,
                    random = list(~ 1 | IDSTUDY/ESID,
                                  ~ 1 | IDCOUNTRY),
                    method = "REML",
                    data = dat,
                    tdist = TRUE,
                    test = "t",
                    mods =~ scale(PubYear, ## mean-centered
                                  center = TRUE,
                                  scale = FALSE))
```

$\hat{\beta}_1 = 0.03, SE = 0.01, 95\,\%CI\ [0.01, 0.06]$



Result: Significant relation between publication year and the effects. More recent studies exhibit larger effect sizes.

Detecting Influential Effect Sizes

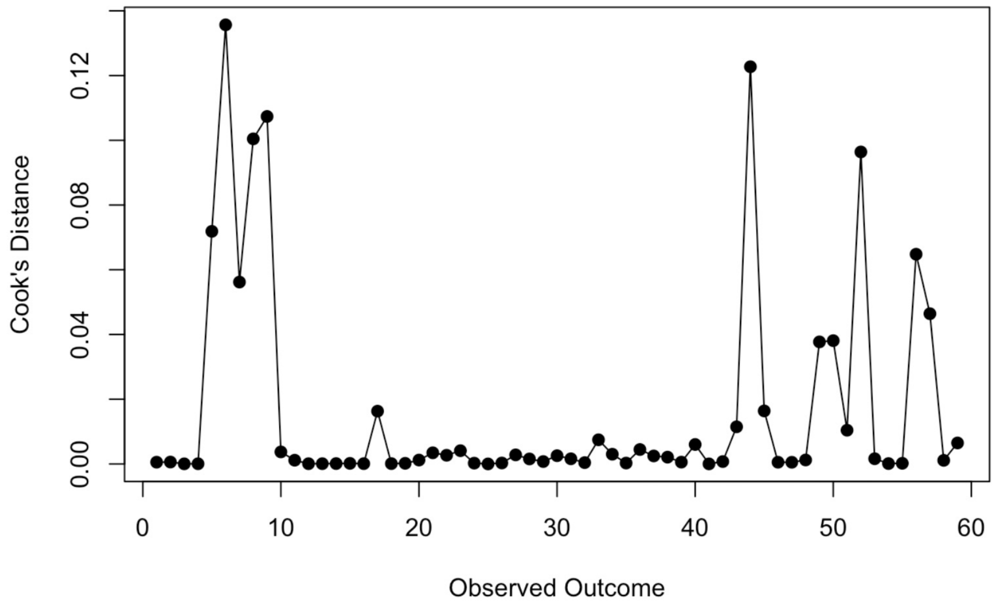
Standard methods to detect «outliers» at multiple levels

Great overview: Cheung & Viechtbauer (2010), https://doi.org/10.1002/jrsm.11

Level: Effect sizes

```r
## Source: https://wviechtb.github.io/metafor/reference/influence.rma.mv.html
## Compute Cook's distances
CCREM.cook1 <- metafor::cooks.distance.rma.mv(CCREM,
                                              reestimate = FALSE,
                                              parallel = "multicore",
                                              ncpus = ncores - 1)


## Plot the D values
plot(CCREM.cook1,
     type="o",
     pch=19,
     xlab="Observed Outcome",
     ylab="Cook's Distance")
```



Need for sensitivity analyses (effects and heterogeneity with vs. without the influential effect sizes)

Detecting Influential Effect Sizes

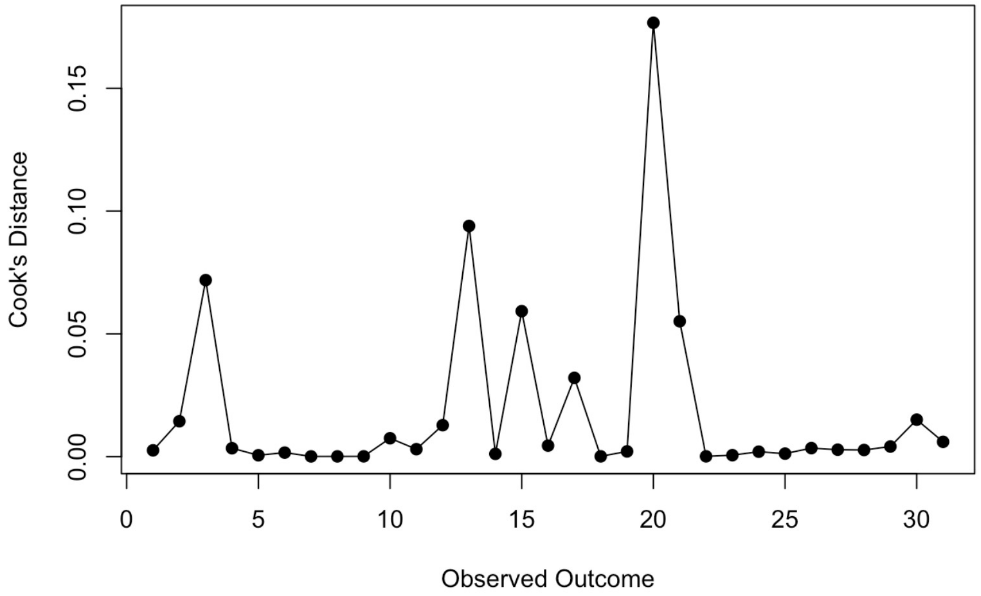
Standard methods to detect «outliers» at multiple levels

Great overview: Cheung & Viechtbauer (2010), https://doi.org/10.1002/jrsm.11

Level: Countries

```
## Source: https://wviechtb.github.io/metafor/reference/influence.rma.mv.html
## Compute Cook's distances
CCREM.cook3 <- metafor::cooks.distance.rma.mv(CCREM,
                                              reestimate = FALSE,
                                              parallel = "multicore",
                                              ncpus = ncores - 1,
                                              cluster = IDCOUNTRY)


## Plot the D values
plot(CCREM.cook3,
     type="o",
     pch=19,
     xlab="Observed Outcome",
     ylab="Cook's Distance")
```



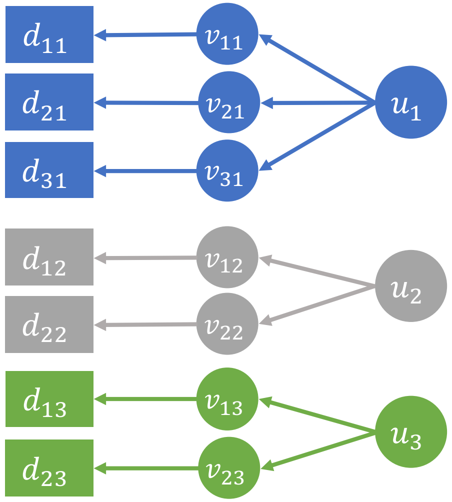
Need for sensitivity analyses (effects and heterogeneity with vs. without the influential effects from some countries)

Thank you.

Appendix A. Robust Variance Estimation (RVE)

Correct standard errors and test statistics for effect size multiplicity

Meta-Analysis with Robust Variance Estimation



Weights for each effect size *i* in study *j*

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

Model 1: Hierarchical Effects (HE)

The first of the original working models is the hierarchical effects (HE) model, which has the form

$$T_{ij} = \mathbf{x}_{ij}\beta + u_j + v_{ij} + e_{ij}, \tag{2}$$

where $\mathrm{Var}\left(u_j\right) = \tau^2$, $\mathrm{Var}\left(v_{ij}\right) = \omega^2$, $\mathrm{Var}\left(e_{ij}\right) = s_{ij}^2$, and $\mathrm{Cov}\left(e_{hj}, e_{ij}\right) = 0$. Here, τ^2 is the between-study variation in study-average true effect sizes, ω^2 is the within-study variation in true effect sizes, and s_{ij} is the known standard error from estimation.

(Pustejovsky et al., 2021, https://doi.org/10.1007/s11121-021-01246-3)

This is a multilevel meta-analytic model, but between- and within-study heterogeneity are only incidental.

Meta-Analysis with Robust Variance Estimation

Random-effects model with RVE

Effect sizes nested in countries

```
## Clustering: Countries
RVE.count <- robu(g ~ 1,
                  data = dat,
                  studynum = IDCOUNTRY,
                  var.eff.size = Var.g,
                  small = FALSE,
                  modelweights = "HIER")

## Model summary
print(RVE.count)
```

```
## RVE: Hierarchical Effects Model
##
## Model: g ~ 1
##
## Number of clusters = 31
## Number of outcomes = 59 (min = 1 , mean = 1.9 , median = 1 , max = 7 )
## Omega.sq = 0.01131715
## Tau.sq = 0.001700087
##
##                Estimate StdErr t-value dfs       P(|t|>) 95% CI.L 95% CI.U Sig
## 1 X.Intercept.    0.146 0.0186    7.85  30 0.00000000936    0.108    0.184 ***
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
```

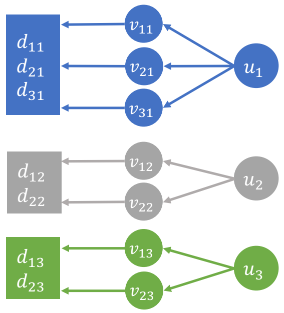
β_R

Appendix B. Multilevel Meta-Analysis with Correlated Effects

Account for possible correlational *and* hierarchical effect size multiplicity

(Pustejovsky & Tipton, 2021, https://doi.org/10.1007/s11121-021-01246-3)

Multilevel Meta-Analysis with CE (MLMA-CE)



Model correlated and hierarchical effects

- Effect sizes y_{ij}, $i = 1, \ldots, k; j = 1, \ldots, m$
- Sampling variances v_{ij} with average $v_{.j}$
- Constant sampling correlation ρ
- Within-study variation ω^2
- Between-study variation τ^2

$$y_{ij} = \beta_R + u_{(2)ij} + u_{(3)j} + e_{ij}$$

Main idea
Incorporate a constant sampling correlation in the model and estimate variances at multiple levels

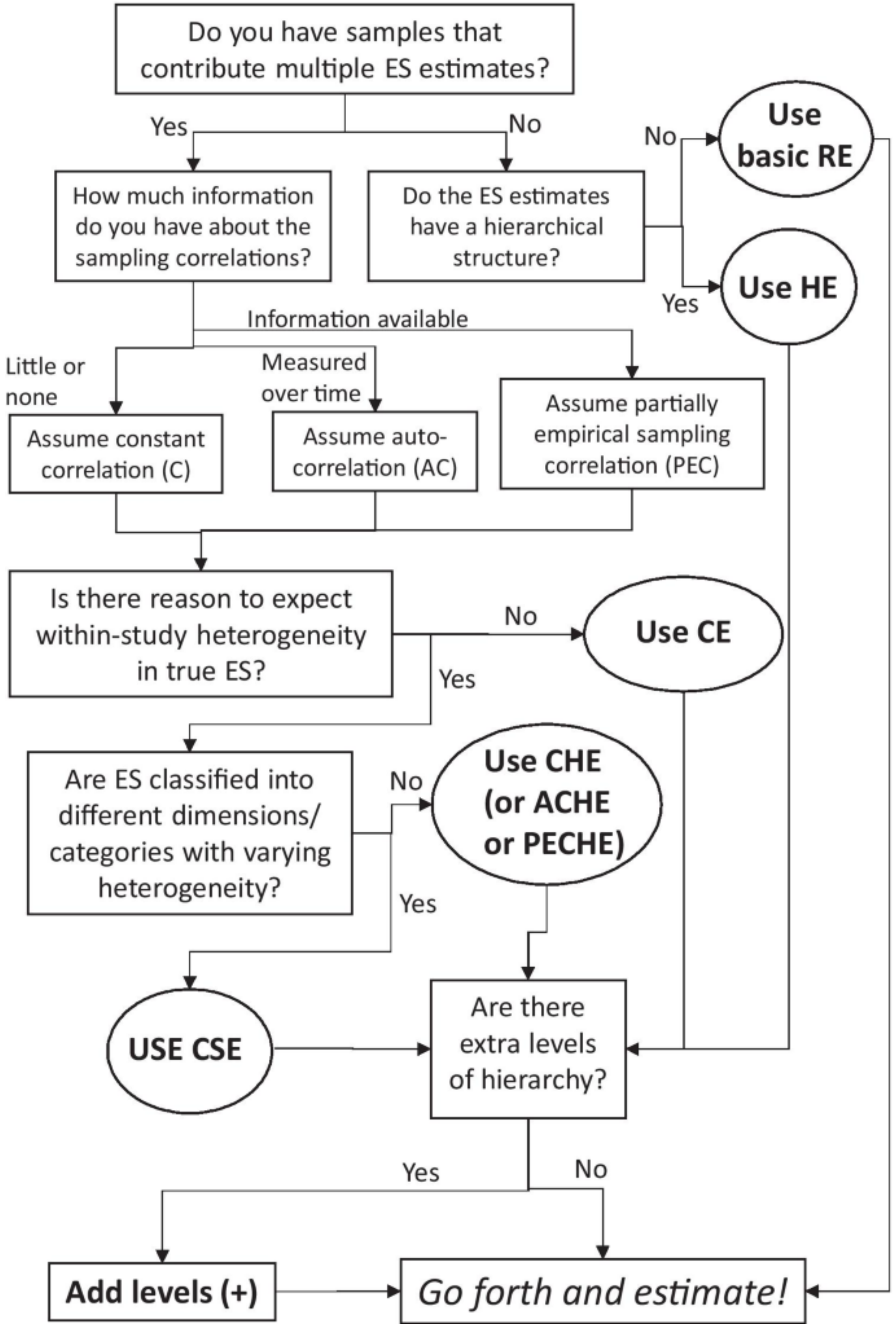
$$Var(u_{(3)j}) = \tau^2$$
$$Var(u_{(2)ij}) = \omega^2$$
$$Var(e_{ij}) = v_{.j}$$
$$Cov(e_{hj}, e_{ij}) = \rho v_{.j}$$

Note: This model can be extended to a model in which the study-specific sampling covariance matrices are incorporated.

Decision Scheme

What is the nature of the multiple effects?

- Hierarchical effects (HE)
- Correlated effects (CE)
- Correlated and hierarchical effects (CHE)



(Pustejovsky et al., 2021, https://doi.org/10.1007/s11121-021-01246-3)

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to Meta-Analysis. John Wiley & Sons, Ltd.

Campos, D. G., Cheung, M. W. -L., & Scherer, R. (2023). A primer on synthesizing individual participant data obtained from complex sampling surveys: A two-stage IPD meta-analysis approach. Psychological Methods. https://doi.org/10.1037/met0000539

Card, N. A. (2012). Applied meta-analysis for social science research. The Guilford Press.

Cheung, M. W.-L. (2015). Meta-Analysis: A Structural Equation Modeling Approach. John Wiley & Sons Ltd.

Cheung, M. W.-L. (2019). A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes. Neuropsychology Review, 29(4), 387-396. https://doi.org/10.1007/s11065-019-09415-6

CMM. (2019). Multilevel structures and classifications. Retrieved 21 August 2019, from http://www.bristol.ac.uk/cmm/learning/multilevel-models/data-structures.html

Egger, M., Smith, G. D., Schneider, M. & Minder, C. (1997). Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ,* 315 (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, 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(5), 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(3), 1286-1304. https://doi.org/10.3758/s13428-018-1063-2

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press. ISBN 978-0-367-61007-4. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/#citing-this-guide

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39-65. https://doi.org/10.1002/jrsm.5

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

Higgins, J.P.T. and Thompson, S.G. (2002), Quantifying heterogeneity in a meta-analysis. Statist. Med., 21: 1539-1558. https://doi.org/10.1002/sim.1186

Pictures by unsplash.com

References

Moeyaert, M., Ugille, M., Natasha Beretvas, S., 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

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(1), 152-171. https://doi.org/10.3758/s13428-018-1156-y

Pastor, D. A., & Lazowski, R. A. (2018). On the Multilevel Nature of Meta-Analysis: A Tutorial, Comparison of Software Programs, and Discussion of Analytic Choices. Multivariate Behavioral Research, 53(1), 74-89. https://doi.org/10.1080/00273171.2017.1365684

Pustejovsky JE, & Rodgers MA. Testing for funnel plot asymmetry of standardized mean differences. *Res Syn Meth*. 2019; 10: 57–71. https://doi.org/10.1002/jrsm.1332

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods, 13*(1), 31–48. https://doi.org/10.1037/1082-989X.13.1.31

Scherer, R., Siddiq, F. & Nilsen, T. (2024). The potential of international large-scale assessments for meta-analyses in education. *Large-scale Assess Educ* **12**, 4. https://doi.org/10.1186/s40536-024-00191-1

Stanley, T.D., & Doucouliagos, H. (2014), Meta-regression approximations to reduce publication selection bias. Res. Syn. Meth., 5: 60-78. https://doi.org/10.1002/jrsm.1095

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. Journal of Developmental and Life-Course Criminology, 2(1), 85-112. https://doi.org/10.1007/s40865-016-0026-5

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. (2015). Meta-analysis of multiple outcomes: a multilevel approach. Behavior Research Methods, 47(4), 1274-1294. https://doi.org/10.3758/s13428-014-0527-2

Viechtbauer, W., & Cheung, M.W.-.-L. (2010), Outlier and influence diagnostics for meta-analysis. Res. Synth. Method, 1: 112-125. https://doi.org/10.1002/jrsm.11

Pictures by unsplash.com