Biodiversity knowledge synthesis: an introduction to meta-analyses and systematic reviews

How to calculate and combine effect-sizes?

October 2022

Beillouin Damien
Chercheur CIRAD- Hortsys

What is an effect-size?

A metric quantifying the relationship between two entities:

- captures the **direction** and **magnitude** of this relationship
- directly extracted from the publication, or calculated

? Difference with the p-value?

What is an effect-size?

A metric quantifying the relationship between two entities:

- captures the **direction** and **magnitude** of this relationship
- directly extracted from the publication, or calculated

? Difference with the p-value?

-It is possible to generate very significant p-values for effect sizes with little practical importance and vice versa

-With enough observations, even tiny differences in parameters become statistically significant

True or estimated effect-size?

An effect-size is commonly noted with the greek letter **theta** (θ):

-θ_k the 'true' effect-size of study k

- $\widehat{\theta_k}$ the observed effect-size of study k

The true and estimated effect-size differs because of the sampling error:

-$\theta_k = \widehat{\theta_k} + \varepsilon_k$

Aim of any study: to be as close as possible to the true effect size

True or estimated effect-size?

Unknown $\longrightarrow$ $\theta_k = \widehat{\theta_k} + \varepsilon_k$ $\longleftarrow$ Unknown

$\widehat{\theta_k}$: Estimated through the mean value
(of a sampling distribution)

ε_k : Estimated through the standard error (SE).
(i.e. The standard deviation of the sampling distribution)
standard error of the **mean** : $SE = \frac{s}{\sqrt{n}}$; with n: sample size, s: standard dev.

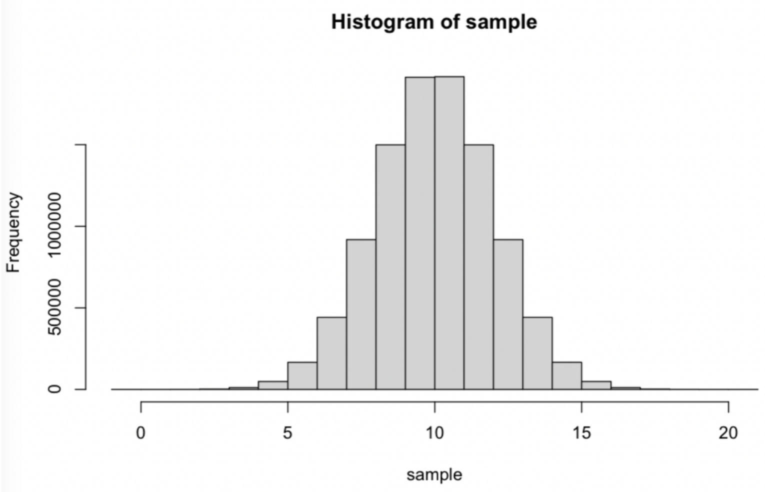
True or estimated effect-size?

$$\theta_k = \widehat{\theta_k} + \varepsilon_k$$

« The perfect world »

A random variable Mean Standard dev.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



$$X \sim \mathcal{N}(10, 2)$$

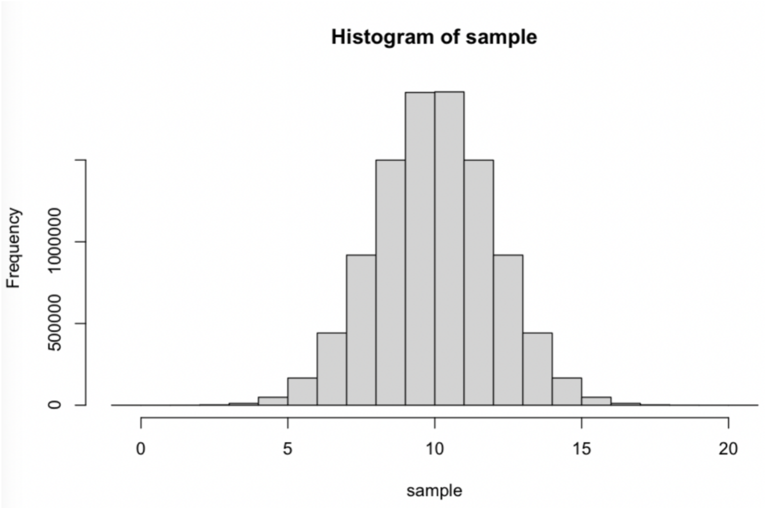
True or estimated effect-size?

$$\theta_k = \widehat{\theta_k} + \varepsilon_k$$

« The perfect world »

A random variable Mean Standard dev.

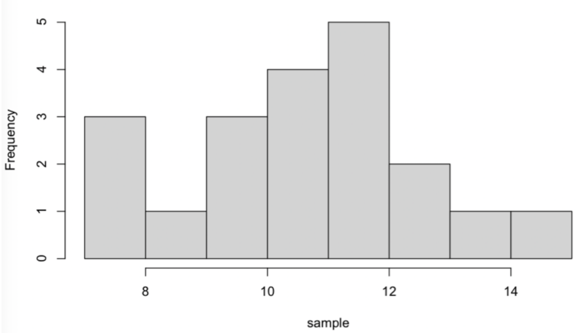
$$X \sim \mathcal{N}(\mu, \sigma^2)$$



$$X \sim \mathcal{N}(10, 2)$$

Experiments

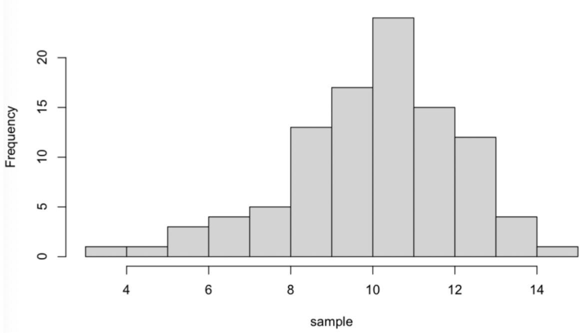
20 exp. units



Estimated Mean : 10.671

SE of the mean : 0.436

100 exp. units



Estimated Mean : 10.063

SE of the mean : 0.199

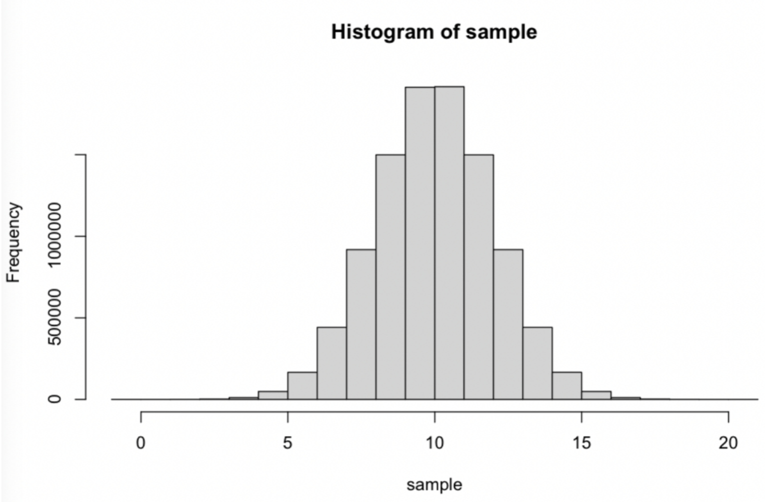
True or estimated effect-size?

$$\theta_k = \widehat{\theta_k} + \varepsilon_k$$

« The perfect world »

A random variable Mean Standard dev.

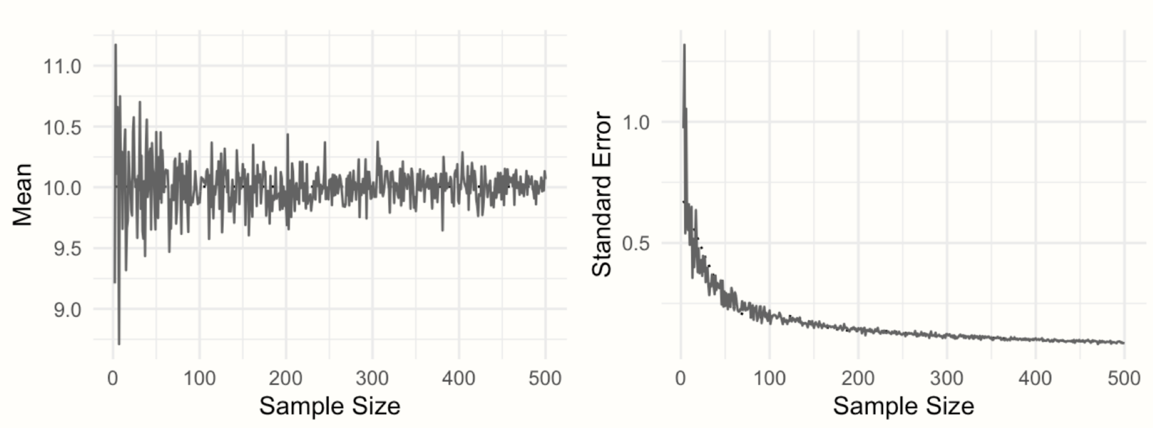
$$X \sim \mathcal{N}(\mu, \sigma^2)$$



$$X \sim \mathcal{N}(10, 2)$$

Experiments

The precision of the estimates increase with sample size



The different type of effect-sizes

Proportions: *(k, n)*

$p = \frac{k}{n}$; with k: number of individuals in a subgroup and n: total sample size

$SE_p = \sqrt{\frac{p(1-p)}{n}}$;

BUT : $p \in [0,1]$ → logit transformation (i.e. log (odds ratio)) : $L \in (-\infty,\infty)$

$p_{logit} = log_e\left(\frac{p}{1-p}\right)$;

$SE_{logit} = \sqrt{\frac{1}{np} + \frac{1}{n(1-p)}}$;

To be retrieved:

(k, n)

Interpretation

Negative logit : *p<0.5*,
positive logits : *p<0.5*

For ex. used in epidemiological studies

The different type of effect-sizes

Correlations:

$$r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} = \frac{COV(x,y)}{\sigma_x \sigma_y}; \text{with x and y two variables}$$

$$SE_{r_{xy}} = \frac{1 - r_{xy}^2}{\sqrt{n-2}};$$

To be retrieved:

(r, n)

BUT : the range of proportions is restricted between 0 and 1 : problematic -> Fisher's z

$$z = 0.5 log_e \left(\frac{1+r}{1-r}\right);$$

$$SE_z = \frac{1}{\sqrt{n-3}};$$

e.g. Correlation of native and exotic species richness (Peng et al., 2019)

The different type of effect-sizes

Mean differences:

$$MD_{between} = \overline{x_1} - \overline{x_2}; \text{with} x_1 \ and \ x_2 \ two \ independant \ groups$$

$$SE_{MD_{between}} = s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}};$$

$$s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$$

To be retrieved:

$(\overline{x_1}, n_1, s_1^2, \overline{x_2}, n_2, s_2^2)$

e.g. Cocoa agroforestry systems versus monocultures (Neither et al., 2020)

The different type of effect-sizes

Mean standardized differences = Cohen's *d*:

$$SMD_{between} = \frac{\overline{x_1} - \overline{x_2}}{s_{pooled}}; \text{with} x_1 \; and \; x_2 \; two \; independant \; groups$$

$$SE_{SMD_{between}} = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{SMD_{between}^2}{2(n_1 + n_2)}};$$

To be retrieved:

$(\overline{x_1}, n_1, s_1^2,$
$\overline{x_2}, n_2, s_2^2)$

BUT bias when the sample size of a study is small, especially when n≤ 20 (L. V. Hedges 1981).

Interpretation

SMD= 2 -> a difference of 2 standard deviations

Hedges'g^* =$SMD_{between}\ X\ (1 - \frac{3}{4n-9})$

The different type of effect-sizes

Ratio:

$$R_{xy} = \frac{\overline{x_1}}{\overline{x_2}};\ \text{with} x_1\ and\ x_2\ two\ independant\ groups$$

$$\log(R) = \log(\overline{x_1})\ - \log(\overline{x_2})$$

$$SE_R = s_{pooled}\sqrt{\frac{1}{n_1(\overline{x_1})^2} + \frac{1}{n_2(\overline{x_2})^2}}$$

To be retrieved:

$(\overline{x_1}, n_1, s_1^2,$
$\overline{x_2}, n_2, s_2^2)$

The different type of effect-sizes

Risk Ratio:

	Event	No Event	
Treatment	a	b	n_{treat}
Control	c	d	n_{control}
	n_E	$n_{\neg E}$	

To be retrieved:

(a, b, c, d)

$$p_{E-treat} = \frac{a}{a+b};\ p_{E-control} = \frac{c}{c+d}; \log(RR) = \log(\frac{p_{E-treat}}{p_{E-control}})$$

$$SE_{\log(RR)} = \sqrt{\frac{1}{a} + \frac{1}{c} - \frac{1}{a+c} - \frac{1}{c+d}}$$

How to pool effect-sizes?

Vote counting :

Should be avoided whenever possible.

-do not account for different weights given to each study

-Do not inform and the magnitude of the effect

How to pool effect-sizes?

Fixed-effect model:

all effect sizes stem from a single population

-> all studies share the **same** true effect size

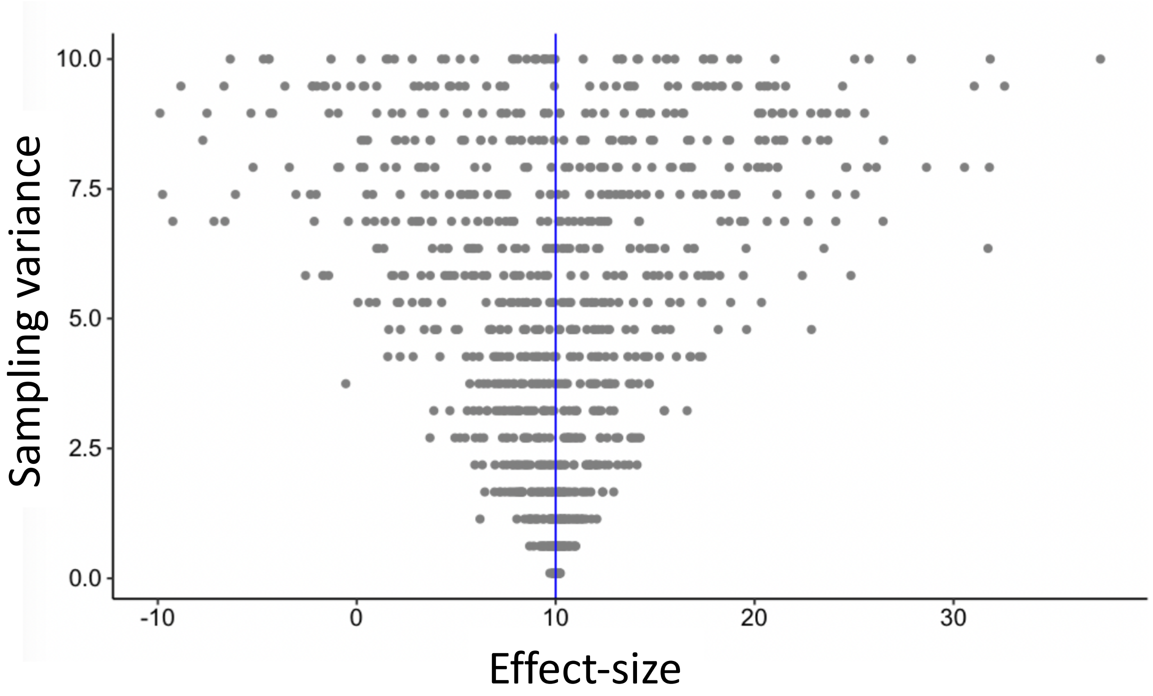
$$\widehat{\theta_k} = \theta + \varepsilon_k$$

The true effect size for study k is not only true for k specifically, but for **all** studies in our meta-analysis

How to pool effect-sizes?

Fixed-effect model:

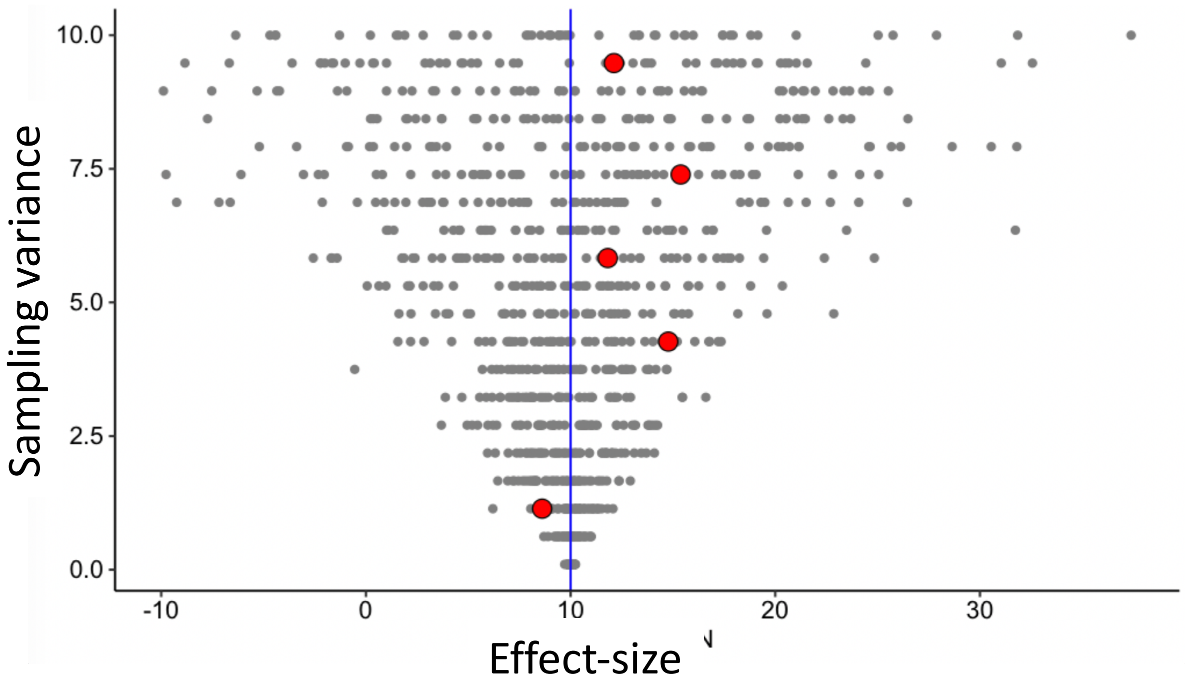
« The perfect world »



How to pool effect-sizes?

Fixed-effect model:

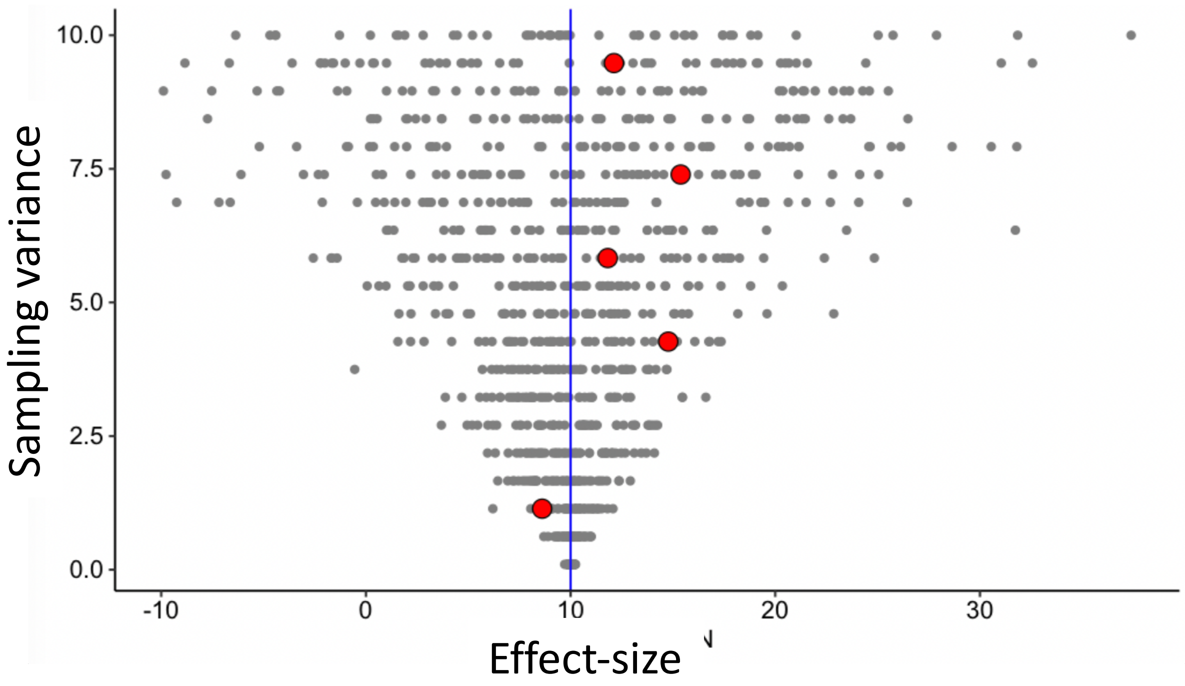
« The perfect world »



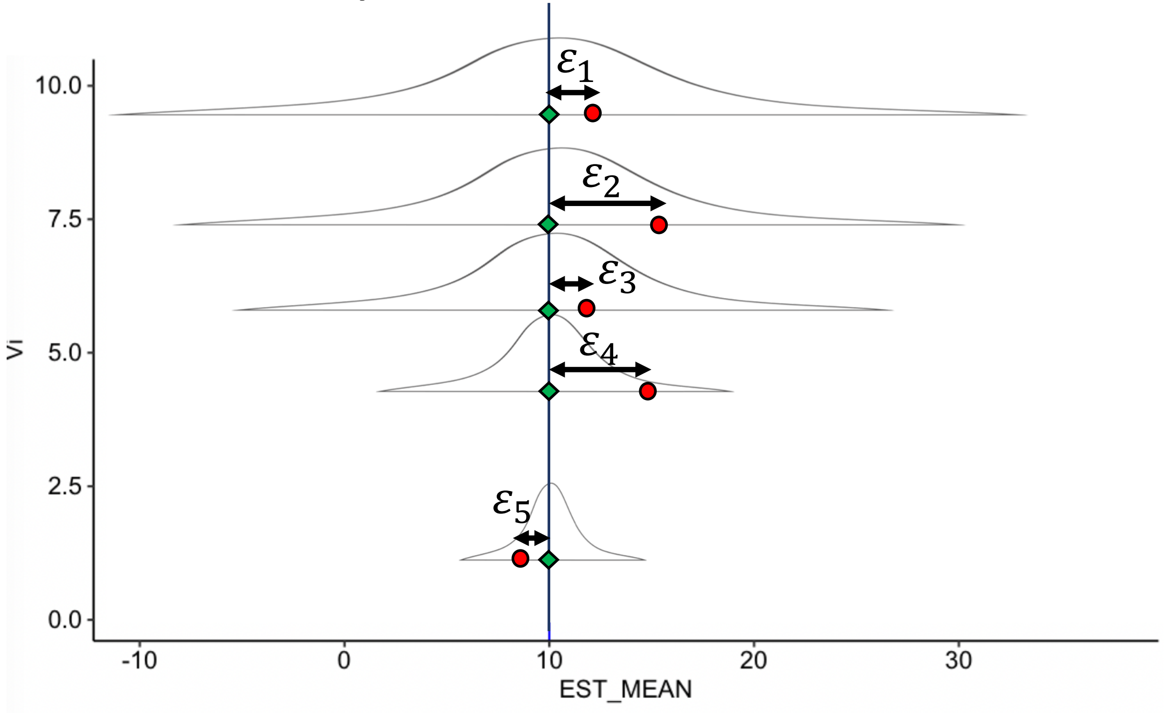
How to pool effect-sizes?

Fixed-effect model:

« The perfect world »



« The experimental world »



How to pool effect-sizes?

Fixed-effect model:

Weights w_k

The usual statistical method for combining results of multiple studies is to weight studies by the amount of information they contribute

$$w_k = \frac{1}{s_k^2}$$

How to pool effect-sizes?

Fixed-effect model:

BUT does not account for:

The outcome of interest could have been measured in many ways.
The type of treatment may not have been exactly the same.
The intensity and duration of treatment could differ.
The target population of the studies may not have been exactly the same for each study.
The control groups used may have been different.

→ between-study **heterogeneity**

How to pool effect-sizes?

Fixed-effect model:

all effect sizes stem from a single population
-> all studies share the **same** true effect size

$$\widehat{\theta_k} = \theta + \varepsilon_k$$

Random-effect model:

There is a distribution of true effect-sizes

$$\widehat{\theta_k} = \theta_k + \varepsilon_k$$

How to pool effect-sizes?

Fixed-effect model:

all effect sizes stem from a single population
-> all studies share the **same** true effect size

$$\widehat{\theta_k} = \theta + \varepsilon_k$$

Random-effect model:

There is a distribution of true effect-sizes

$$\widehat{\theta_k} = \theta_k + \varepsilon_k$$

$$\theta_k = \mu + \zeta_k$$

How to pool effect-sizes?

Fixed-effect model:

all effect sizes stem from a single population
-> all studies share the **same** true effect size

$$\widehat{\theta_k} = \theta + \varepsilon_k$$

Random-effect model:

There is a distribution of true effect-sizes

$$\widehat{\theta_k} = \theta_k + \varepsilon_k$$

$$\theta_k = \mu + \zeta_k$$

$$\widehat{\theta_k} = \mu + \zeta_k + \varepsilon_k$$

How to pool effect-sizes?

Random-effect model:

$$\widehat{\theta_k} = \mu + \zeta_k + \varepsilon_k$$

μ is the overall mean (or meta-analytic mean)

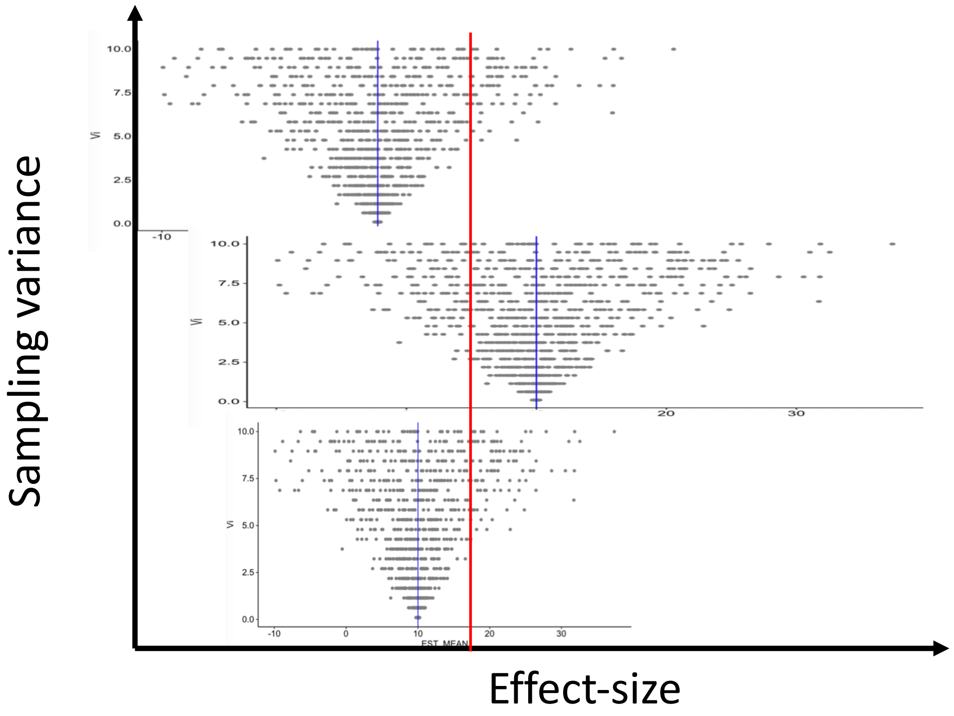
$\zeta_k \sim \mathcal{N}(0, \mathcal{T}^2)$ $\mathcal{T}^2$ is the between study variance

$\varepsilon_k \sim \mathcal{N}(0, v_k)$ v_k is the within study variance

How to pool effect-sizes?

Mixed-effect model:

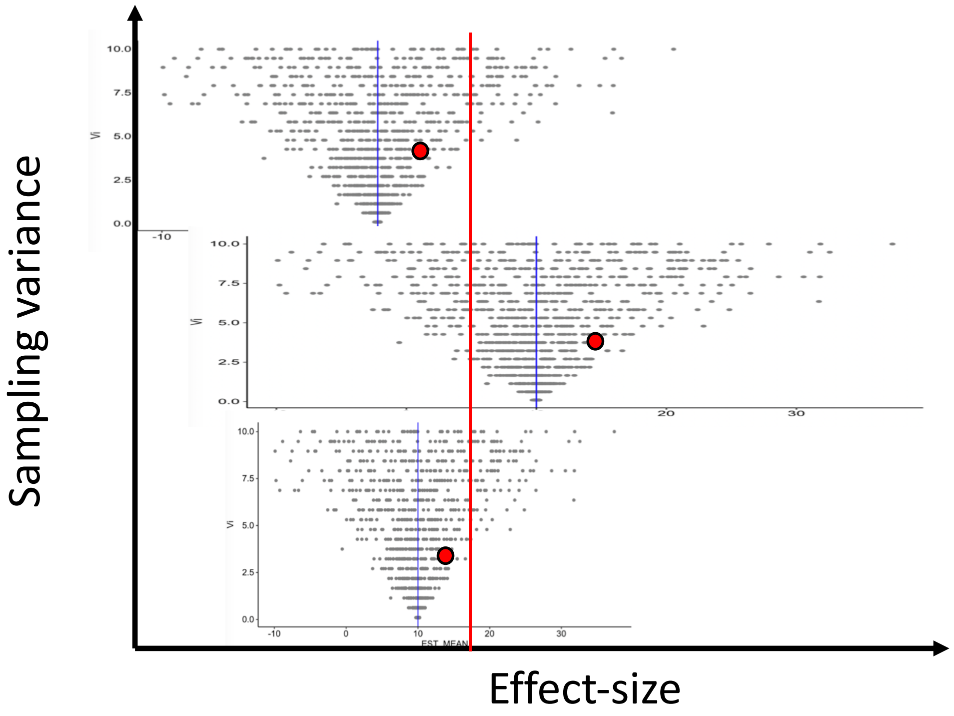
« The perfect world »



How to pool effect-sizes?

Mixed-effect model:

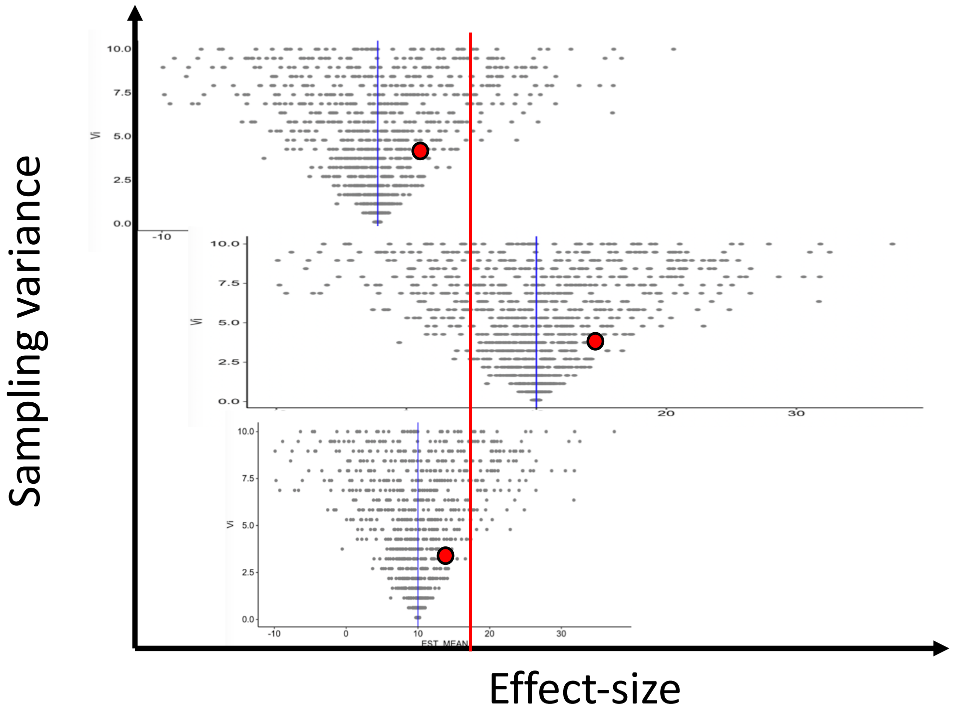
« The perfect world »



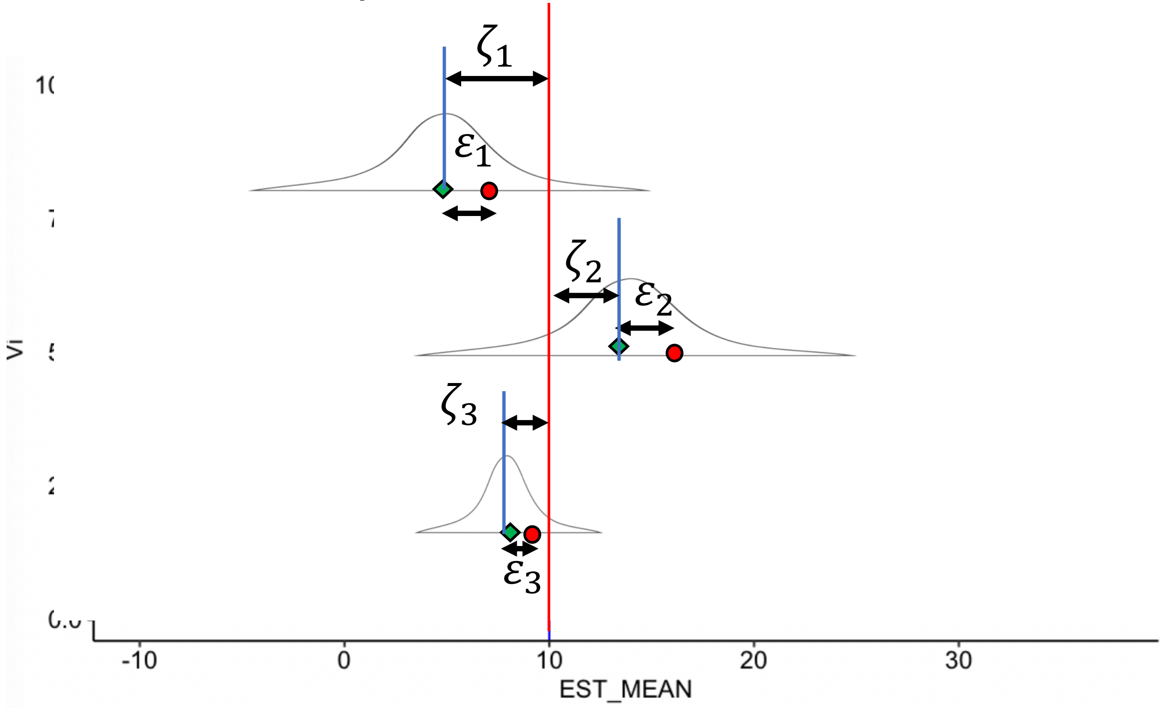
How to pool effect-sizes?

Mixed-effect model:

« The perfect world »



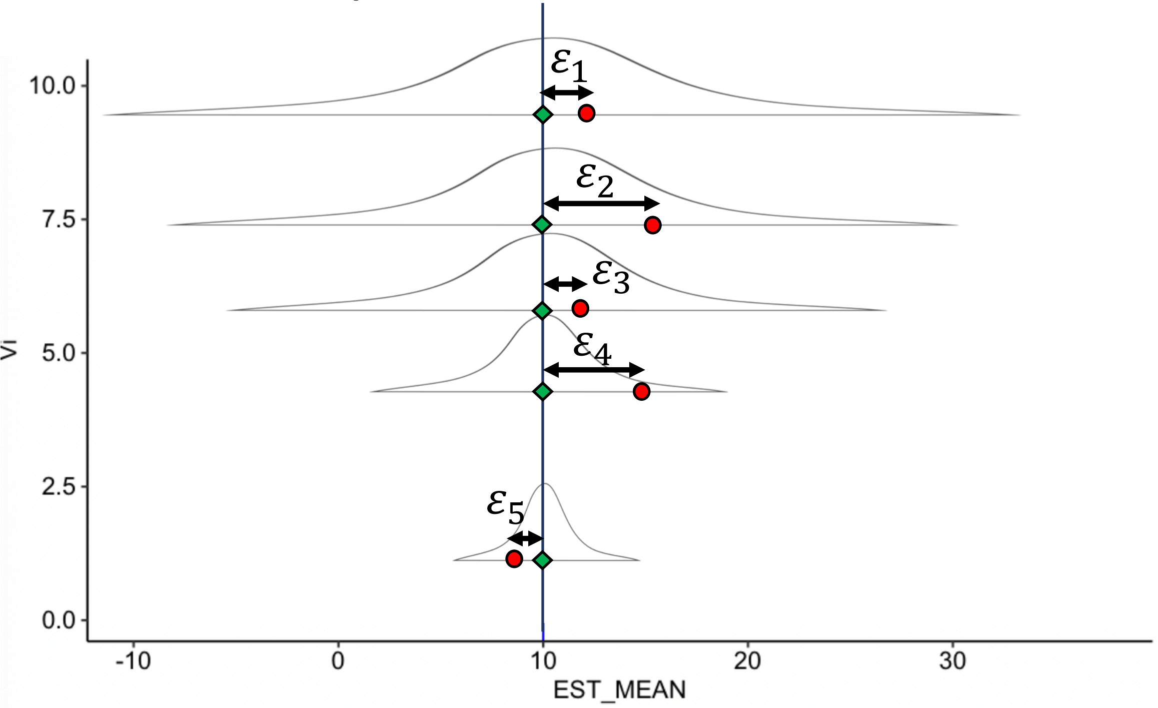
« The experimental world »



How to pool effect-sizes?

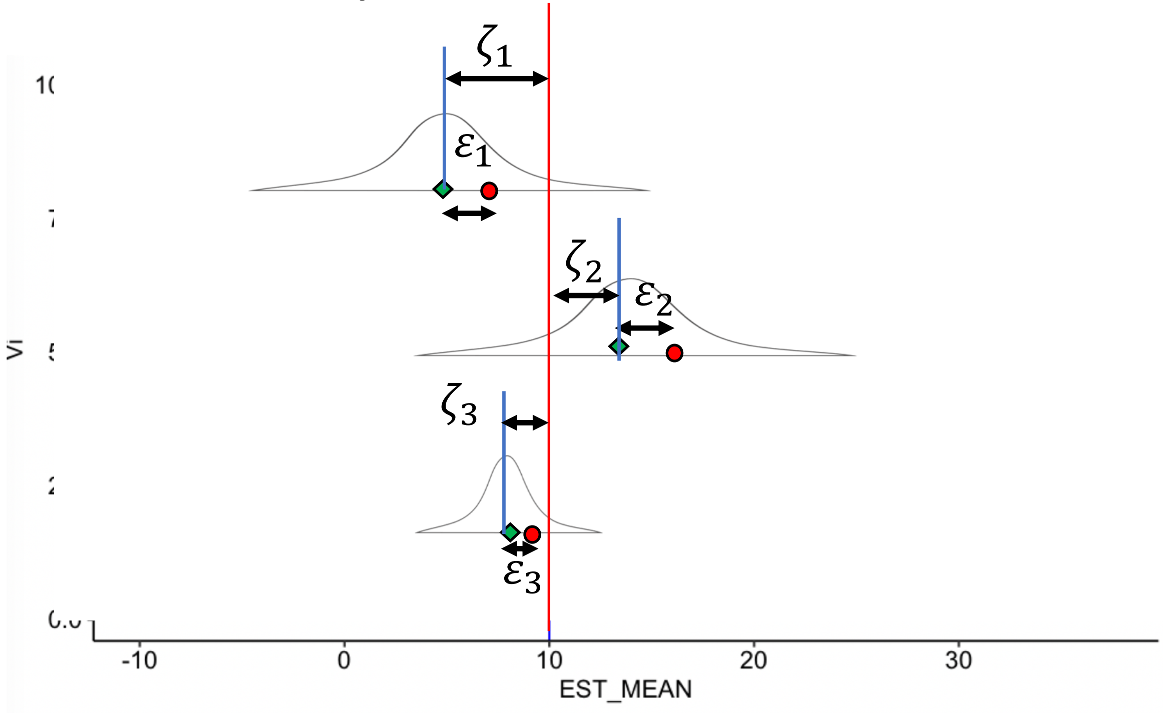
Fixed-effect model:

« The experimental world »



Mixed-effect model:

« The experimental world »



How to pool effect-sizes?

Mixed-effect model:

Weights w_k

The usual statistical method for combining results of multiple studies is to weight studies by the amount of information they contribute

$$w_k = \frac{1}{v_k^2 + \mathcal{T}^2}$$

How to pool effect-sizes?

Mixed-effect model:

Methods to estimates $\mathcal{T}^2$

DerSimonian-Laird (DL) (default estimator in Revman, omprehensive Meta-Analysis, meta package)

Restricted Maximum Likelihood (REML) (default estimator in metafor package)

Paule-Mandel (PM)

…

How to pool effect-sizes?

Mixed-effect model:

Methods to better consider the distribution estimates $\mathcal{T}^2$

(a correction applicable after each of the previous mentionned estimators of $\mathcal{T}^2$*)*

- Knapp-Hartung Adjustments

How to pool effect-sizes?

Three levels meta-analysis :

Statistical independence of the effect sizes is one of the core assumptions when we pool effect sizes in a meta-analysis

BUT:

-One author could report several effect-sizes (multiples experiments, control, …)
-An overall structure of the data (Climate effect, country effect, ….)

→ Nested Three levels met-analyses

How to pool effect-sizes?

Random-effect model (2 levels):

There is a distribution of true effect-sizes

$$\widehat{\theta_k} = \theta_k + \varepsilon_k$$

$$\theta_k = \mu + \zeta_k$$

$$\widehat{\theta_k} = \mu + \zeta_k + \varepsilon_k$$

Random-effect model (3 levels):

There is a distribution of true effect-sizes

$$\widehat{\theta_{jk}} = \theta_{jk} + \varepsilon_{jk}$$

$$\theta_{jk} = \kappa_j + \zeta_{(2)jk}$$

$$\kappa_j = \mu + \zeta_{(3)j}$$

$$\widehat{\theta_k} = \mu + \zeta_{(2)jk} + \zeta_{(3)j} + \varepsilon_{jk}$$

How to pool effect-sizes?

Random-effect model (three levels):

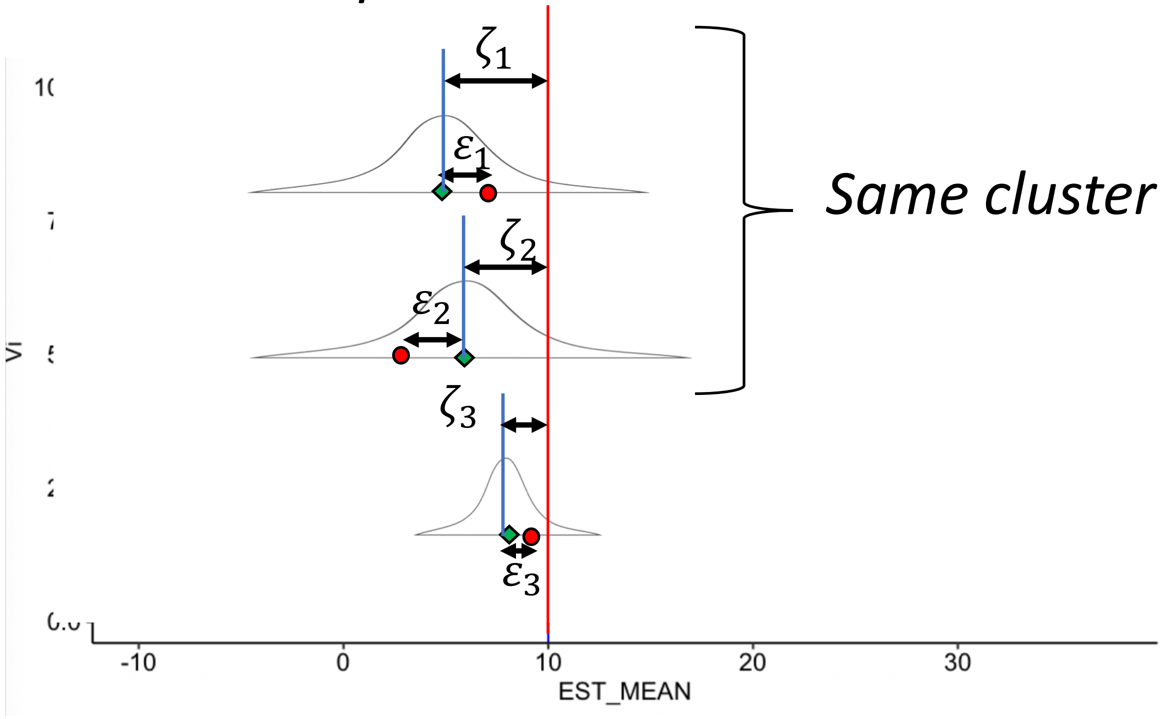
$$\widehat{\theta_k} = \mu + \zeta_{(2)jk} + \zeta_{(3)j} + \varepsilon_{jk}$$

μ	is the overall mean (or meta-analytic mean)
$\zeta_{(2)kj} \sim \mathcal{N}(0, \mathcal{T}^2_{(2)})$	$\mathcal{T}^2_{(2)}$ is the within cluster variability
$\zeta_{(3)j} \sim \mathcal{N}(0, \mathcal{T}^2_{(3)})$	$\mathcal{T}^2_{(3)}$ is the between cluster variability
$\varepsilon_{kj} \sim \mathcal{N}(0, v_{jk})$	v_{jk} is the within study in cluster variance

How to pool effect-sizes?

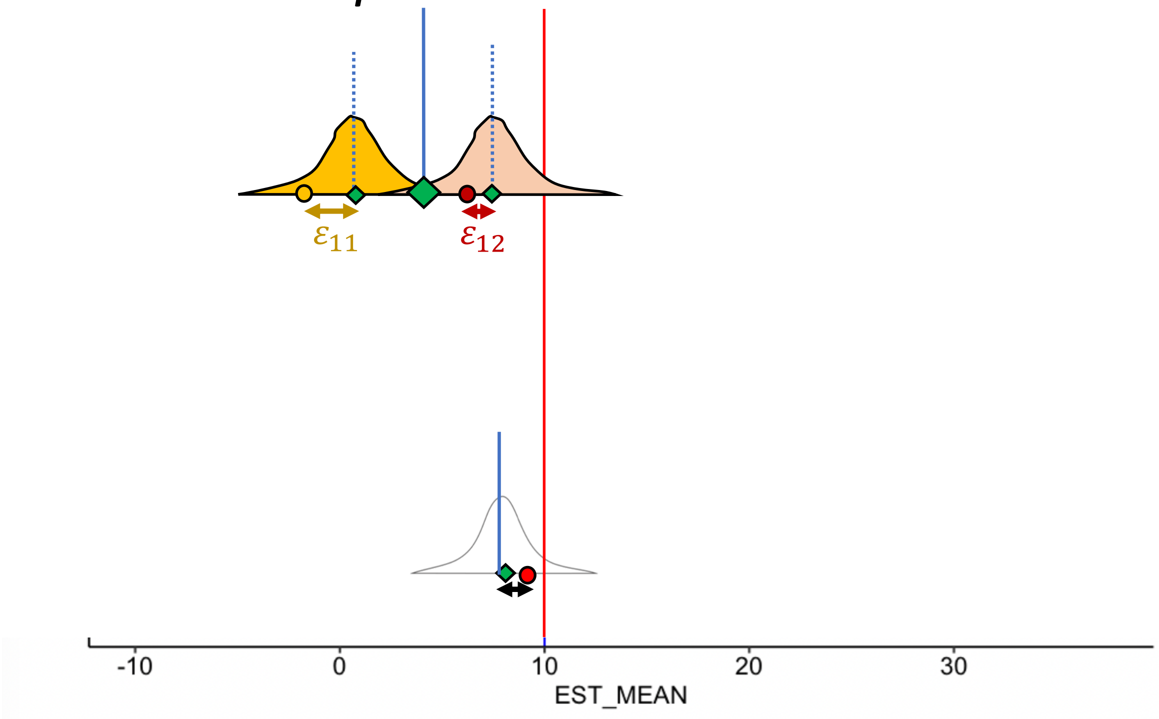
Mixed-effect model (3 levels):

« The experimental world »



Mixed-effect model (3 levels):

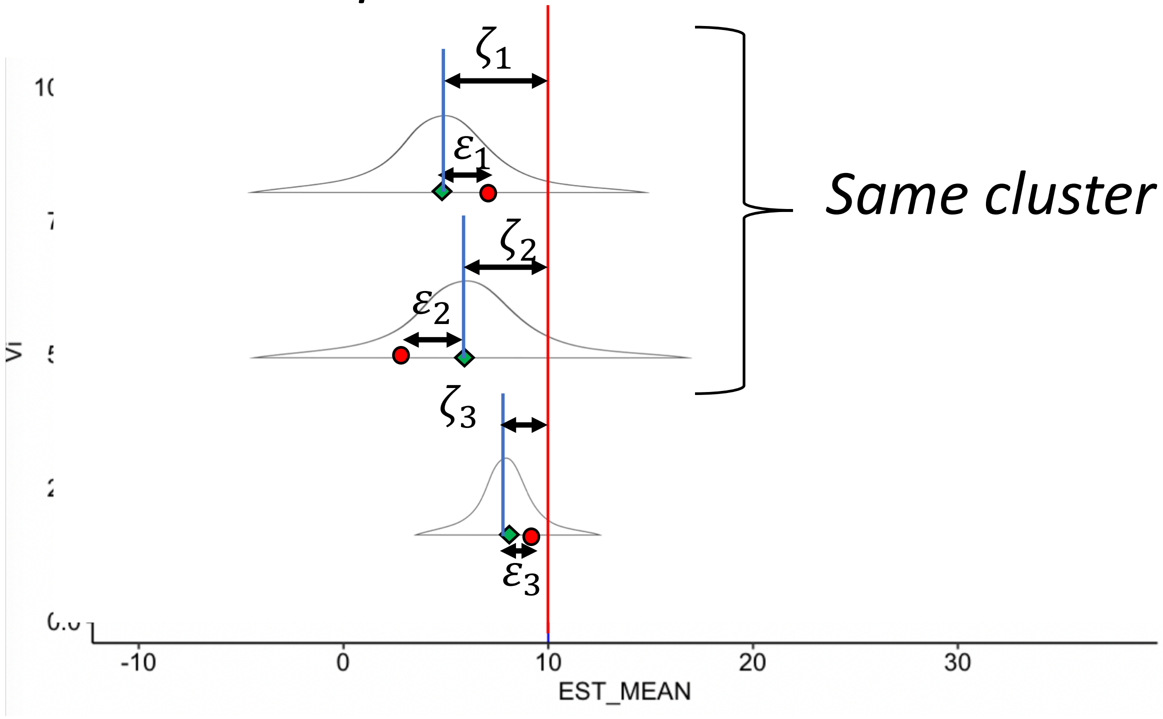
« The experimental world »



How to pool effect-sizes?

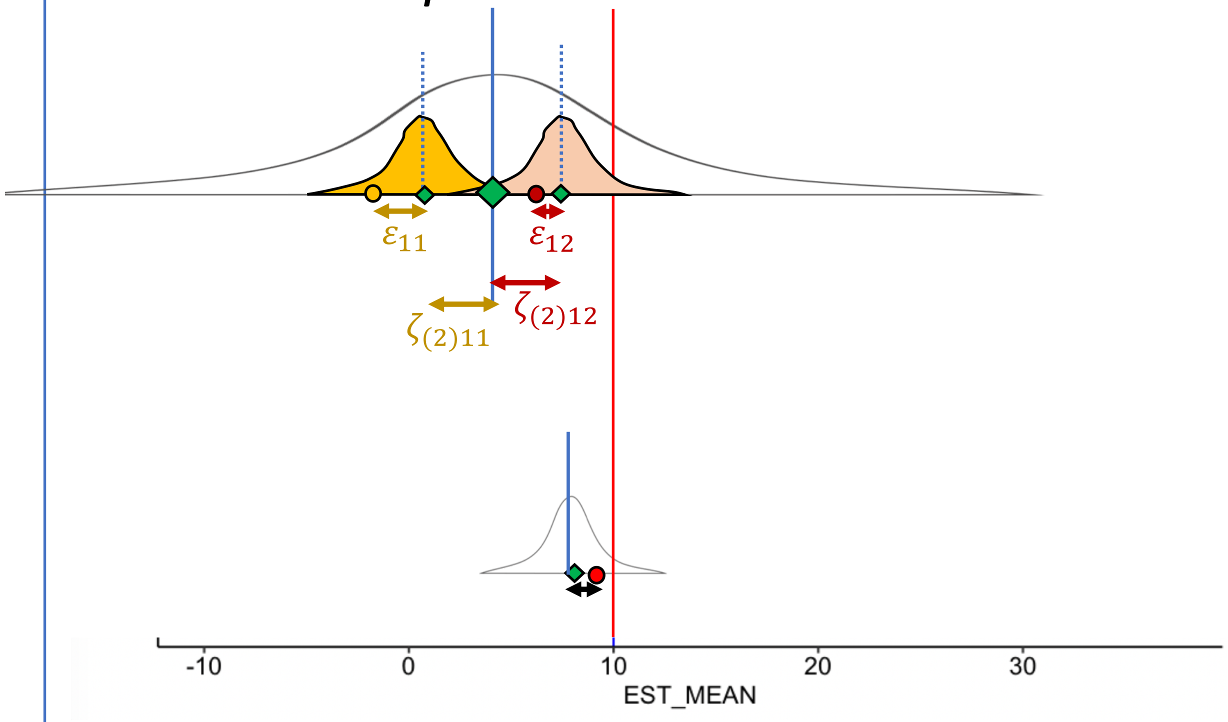
Mixed-effect model (3 levels):

« The experimental world »



Mixed-effect model (3 levels):

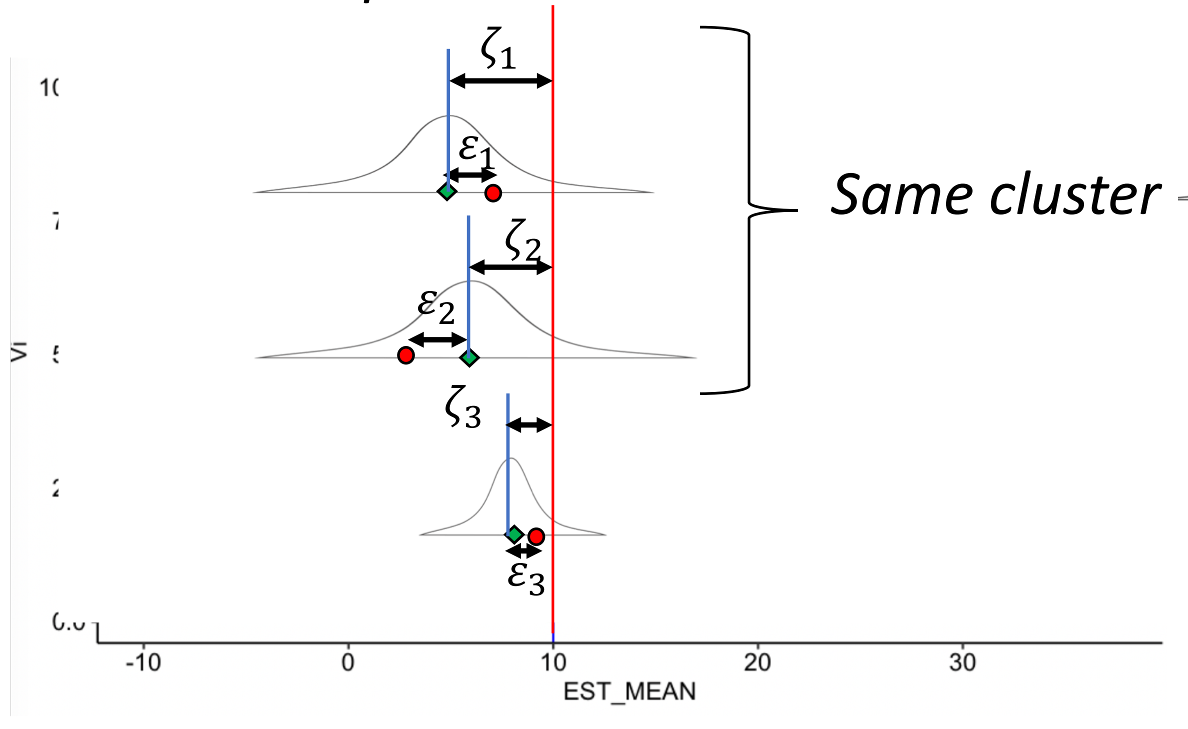
« The experimental world »



How to pool effect-sizes?

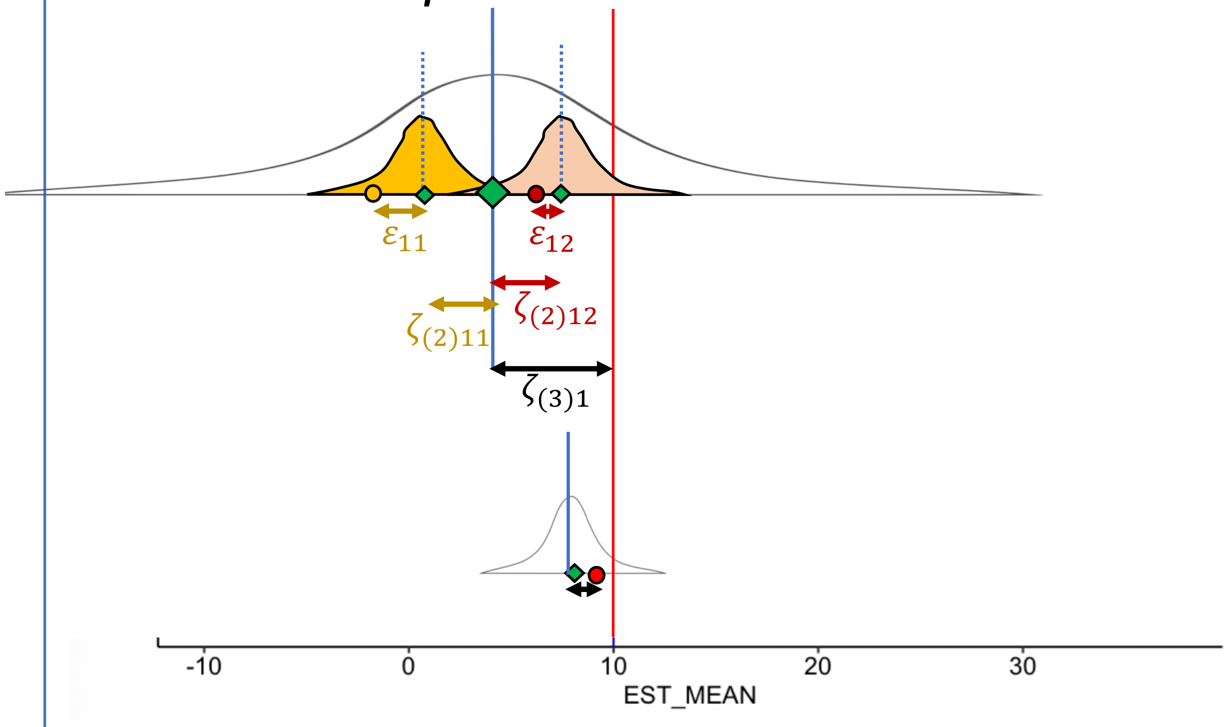
Mixed-effect model (3 levels):

« The experimental world »



Mixed-effect model (3 levels):

« The experimental world »



How to pool effect-sizes?

Sub-group analyses

We assume that the different effect-sizes fall into different **subgroups** and that each subgroup has its own true overall effect

Possible to analyse subgroups in fixed- and random-effect models

How to pool effect-sizes?

Sub-group analyses

Why don’t we run a separate meta-anaysis for each group?

Estimates of the various variances (i.e. $\mathcal{T}^2$, ..) will also differ from subgroup to subgroup, but could be very imprecise (when n is low)

Rather : We consider a **common** estimate of the between-study heterogeneity for each subgroups (better estimated).

How to pool effect-sizes?

Sub-group analyses

Assess if there is a **true** difference between the groups

Perform a statistical hypothesis test (compare the variance between vs. Within the groups)
In meta-analyses-> Cochran's Q test (available by default).

Interpretation : at least one subgroup is part of a different population of studies (or not)

How to pool effect-sizes?

Meta-regressions:

Subgroup analyses are a special form of **meta-regression**

Use the value of some variable x to predict the value of another variable y, and applied to **entire studies**

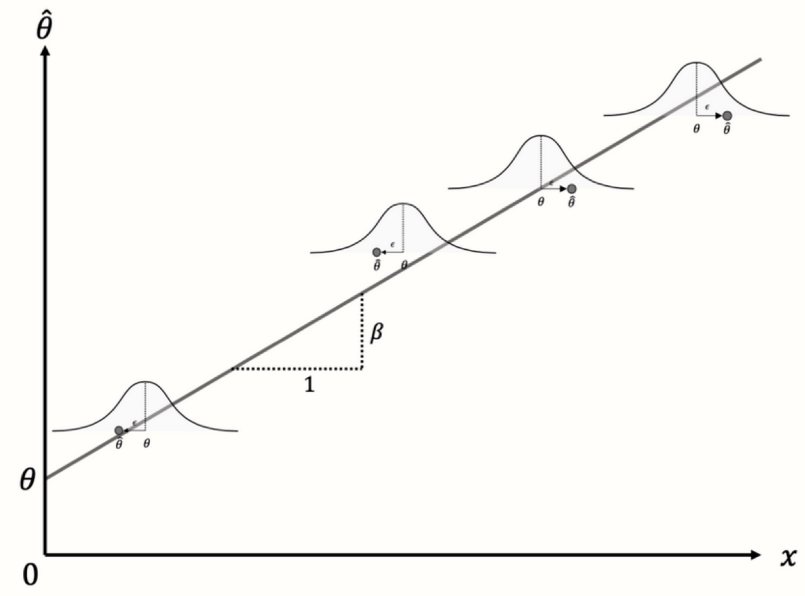
How to pool effect-sizes?

Meta-regressions (mixed effect models):

$$\widehat{\theta_{k}} = \mu + \beta x_k + \zeta_k + \varepsilon_k$$

β the regression coefficient

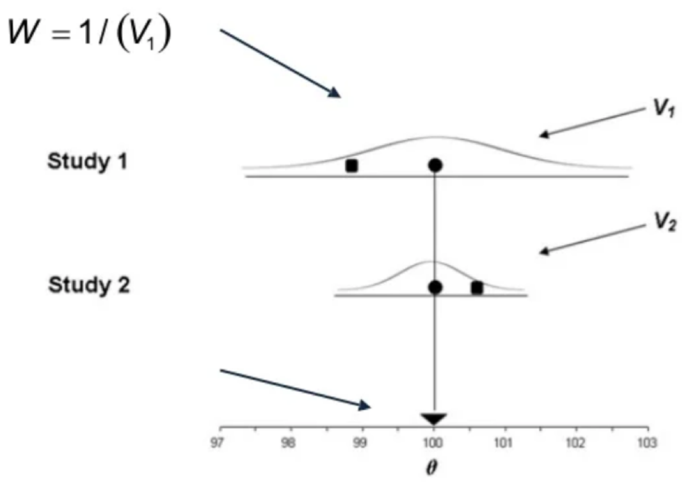
Estimated through weighted least squares (WLS)



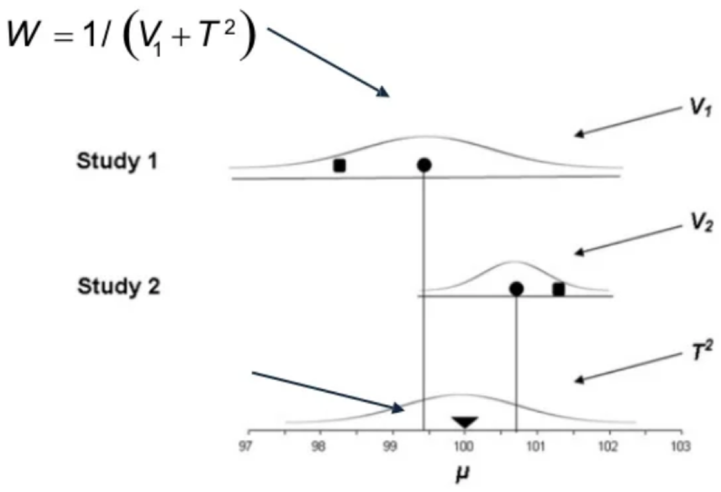
Analysis of heterogeneity

Remember, heterogenity impact weights (of random effect meta-analyses)

Weights when $T^2 = 0$

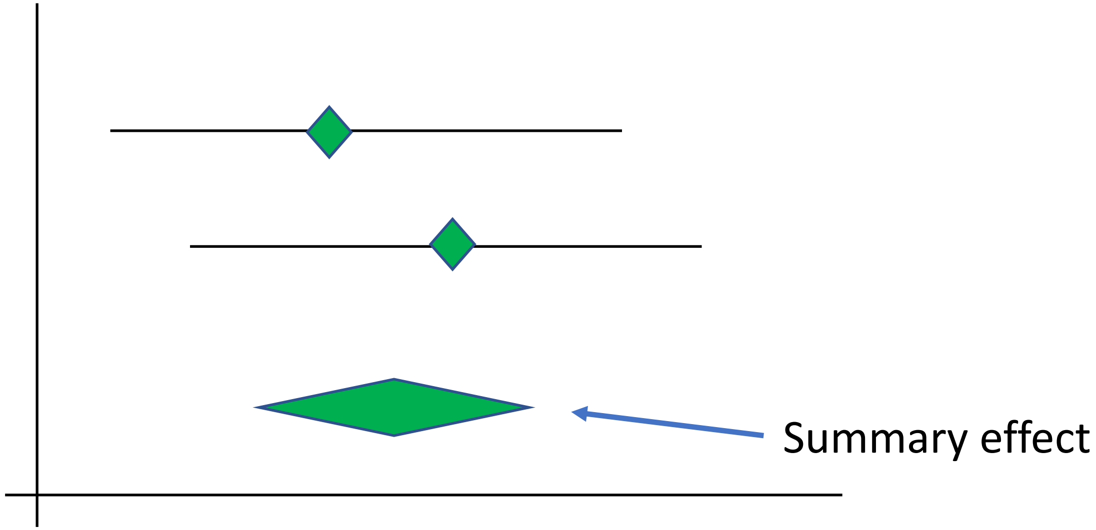


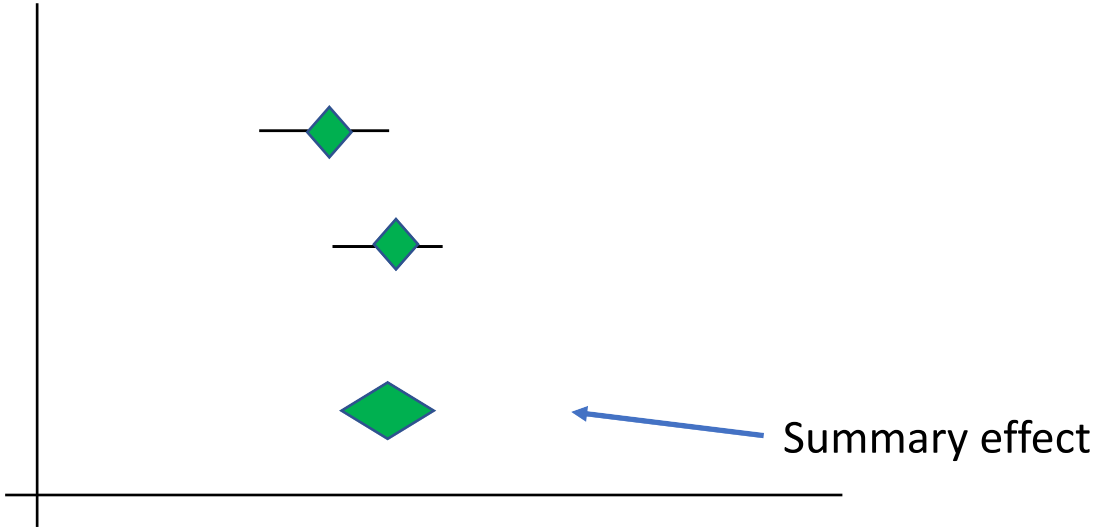
Weights when $T^2 > 0$



Analysis of heterogeneity

Remember, it impacts the estimated variances, thus the confidence intervals, and p-value of the mean effect.





Analysis of heterogeneity

Cochran's Q: which is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies,

The I^2 statistic describes the percentage of variation across studies that is due to heterogeneity rather than chance

$$I^2 = \frac{V_{TRUE}}{V_{OBS}} = \frac{T^2}{V_{OBS}}.$$

$\mathcal{T}^2$ is the between study variance

Analysis of heterogeneity

How to interpret I2?

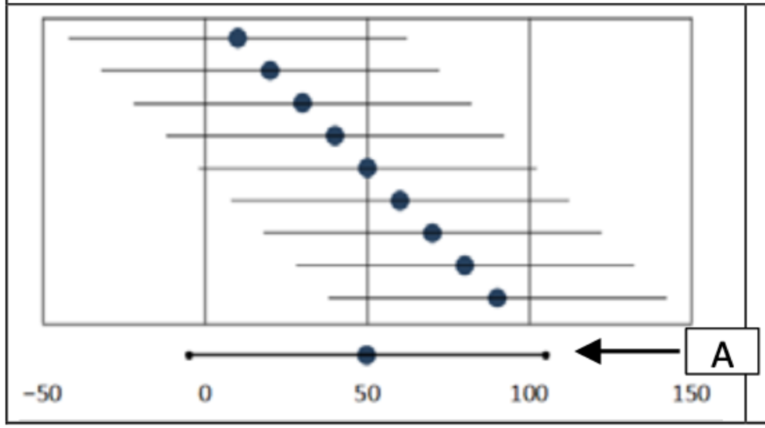
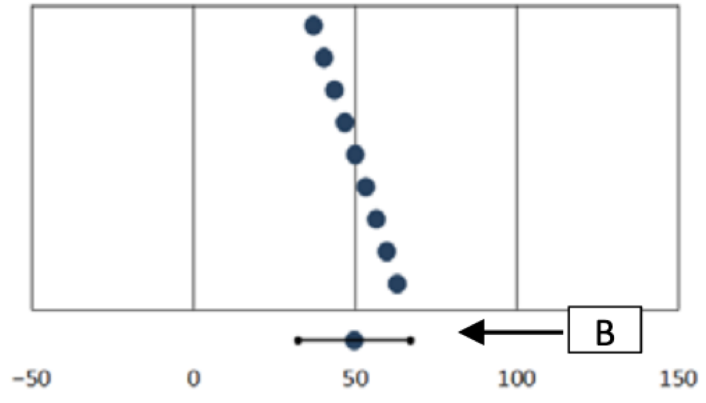
Dispersion of observed effects			Dispersion of true effects	
	V_{OBS}=750 V_{ERR}=675 T^2=75 S_{OBS}=27.4 $M \pm 2S$ −5 | 105	I^2=10% I=32%	T^2=75 T=8.66 $M \pm 2T$ 33 | 67	

Figure 1 | Dispersion of observed effects and dispersion of true effects

Mise en pratique

- library(metafor)

```
> head(dat.bcg)
  trial               author year tpos  tneg cpos  cneg ablat     alloc
1     1              Aronson 1948    4   119   11   128    44    random
2     2     Ferguson & Simes 1949    6   300   29   274    55    random
3     3      Rosenthal et al 1960    3   228   11   209    42    random
4     4    Hart & Sutherland 1977   62 13536  248 12619    52    random
5     5 Frimodt-Moller et al 1973   33  5036   47  5761    13 alternate
6     6      Stein & Aronson 1953  180  1361  372  1079    44 alternate
>
```

Results from 13 clinical trials examining the effectivess of the bacillus Calmette-Guerin (BCG) vaccine for preventing tuberculosis

Mise en pratique

- On calcule l'effect size. Ici le log du relative risque entre vaccins et non vaccins

dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos, di=cneg, data=dat.bcg)

```
n1i <- ai + bi
n2i <- ci + di
ni <- n1i + n2i
p1i.u <- ai.u/n1i.u
p2i.u <- ci.u/n2i.u
p1i <- ai/n1i
p2i <- ci/n2i
if (measure == "RR") {
  if (addyi) {
    yi <- log(p1i) - log(p2i)
  }
```

Mise en pratique

- La nouvelle table de résultats

author	year	tpos	tneg	cpos	cneg	ablat	alloc	yi	vi
Aronson	1948	4	119	11	128	44	random	-0.8893	0.3256
Ferguson & Simes	1949	6	300	29	274	55	random	-1.5854	0.1946
Rosenthal et al	1960	3	228	11	209	42	random	-1.3481	0.4154
Hart & Sutherland	1977	62	13536	248	12619	52	random	-1.4416	0.0200
Frimodt-Moller et al	1973	33	5036	47	5761	13	alternate	-0.2175	0.0512
Stein & Aronson	1953	180	1361	372	1079	44	alternate	-0.7861	0.0069
Vandiviere et al	1973	8	2537	10	619	19	random	-1.6209	0.2230
TPT Madras	1980	505	87886	499	87892	13	random	0.0120	0.0040
Coetzee & Berjak	1968	29	7470	45	7232	27	random	-0.4694	0.0564
Rosenthal et al	1961	17	1699	65	1600	42	systematic	-1.3713	0.0730
Comstock et al	1974	186	50448	141	27197	18	systematic	-0.3394	0.0124
Comstock & Webster	1969	5	2493	3	2338	33	systematic	0.4459	0.5325
Comstock et al	1976	27	16886	29	17825	33	systematic	-0.0173	0.0714

Mise en pratique

- An equal-effects model can be fitted to these data using the rma.uni() function with:

```
rma(yi, vi, method="EE", data=dat)

rma(yi, vi, data=dat)
```

Mise en pratique

- Les résultats

```
Random-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.3132 (SE = 0.1664)
tau (square root of estimated tau^2 value):      0.5597
I^2 (total heterogeneity / total variability):   92.22%
H^2 (total variability / sampling variability):  12.86

Test for Heterogeneity:
Q(df = 12) = 152.2330, p-val < .0001

Model Results:

estimate       se      zval      pval     ci.lb     ci.ub
 -0.7145   0.1798   -3.9744   <.0001   -1.0669   -0.3622      ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Mise en pratique

- Les résultats

```
Random-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 0.3132 (SE = 0.1664)
tau (square root of estimated tau^2 value):      0.5597
I^2 (total heterogeneity / total variability):   92.22%
H^2 (total variability / sampling variability):  12.86

Test for Heterogeneity:
Q(df = 12) = 152.2330, p-val < .0001

Model Results:

estimate       se      zval      pval     ci.lb     ci.ub
 -0.7145   0.1798  -3.9744   <.0001  -1.0669  -0.3622      ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```