Understanding meta-analysis through data simulation with applications to power analysis

Filippo Gambarota [1] & Gianmarco Altoè [1]

[1] Department of Developmental and Social Psychology, University of Padova, Italy

Author Note

The authors made the following contributions. Filippo Gambarota: Conceptualization, Methodology, Formal Analysis, Software, Writing – Original Draft; Gianmarco Altoè: Conceptualization, Methodology, Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Filippo Gambarota, Via Venezia 8, 35131 Padova (PD). E-mail: filippo.gambarota@unipd.it

Abstract

Meta-analysis is a powerful tool to combine evidence from existing literature. Despite several introductory and advanced materials about organizing, conducting, and reporting a meta-analysis, to our knowledge, there are no introductive materials about simulating the most common meta-analysis models. Data simulation is essential for developing and validating new statistical models and procedures. Furthermore, data simulation is a powerful educational tool for understanding a statistical method. In this tutorial, we show how to simulate equal-effects, random-effects, and meta-regression models and illustrate how to estimate statistical power. In the supplementary materials, we extended these simulations also for multilevel and multivariate models. All materials associated with this article can be accessed on Open Science Framework (https://osf.io/54djn/).

Keywords: meta-analysis, monte carlo simulations, power analysis

Word count: 7869

Understanding meta-analysis through data simulation with applications to power analysis

1 Introduction

"If you do not simulate it, you have not understood it."

A meta-analysis is an essential tool for combining knowledge from multiple studies quantitatively. Meta-analysis is commonly used together with a systematic review of the literature. The meta-analysis has several advantages. Firstly, it allows combining evidence from multiple studies assigning more weight to studies with lower estimation variability. Then using meta-regression, it is possible to include variables (i.e., moderators) to explain the observed heterogeneity (Borenstein, Hedges, Higgins, & Rothstein, 2009, pp. 187–203). More recently, location-scale models have been developed to include predictors also on the residual heterogeneity (i.e., Viechtbauer & López-López, 2022). Finally, considering the replication crisis, there are statistical methods to determine the presence and extent of the publication bias. Despite the advantages, meta-analysis implementation is not always straightforward, especially for complex data structures. Additionally, the fact that there are several introductory and advanced resources to understand meta-analysis (Borenstein et al., 2009; Harrer, Cuijpers, Furukawa, & Ebert, 2021; Schmid, Stijnen, & White, 2022), to our knowledge, there are no introductory resources about how to simulate realistic meta-analytic data.

Simulating data has several advantages because it requires understanding the statistical method and the data-generation process. Furthermore, data simulation is the primary tool when it comes to evaluating a new analysis method, estimating the statistical power, or understanding the long-run behavior of our data generation process (Gelman & Hill, 2006, pp. 155–176; Gelman, Hill, & Vehtari, 2020, pp. 69–76; Ingalls, 2011). A recent paper by DeBruine and Barr (2021), which deeply inspired the current work, proposed a stimulating way to understand linear mixed-effects models via data simulation. Simulating

data is also a powerful educational tool within this framework.

For these reasons, this work aims to introduce the basic concepts of meta-analysis and Monte Carlo simulations for equal-effects, random-effects, and meta-regression models with applications also to statistical power calculation. In the first section, we will introduce basic concepts of the meta-analysis that are useful for setting up the simulation. We will evaluate the effect size, variance calculation, and the equal vs. random effects model distinction. Then we will describe how to simulate data for these models and simulate a meta-regression with categorical and numerical predictors. Finally, we will introduce the power analysis extending the previous examples to estimate the statistical power. We used the R statistical programming language (version 4.3.1; R Core Team, 2023).

The aim of the tutorial is not to provide a complete theoretical introduction to meta-analysis but rather to present core topics using a simulation-based approach. Readers experienced in conducting meta-analyses can benefit from the proposed approach in terms of simulation setup and coding strategies. Readers without prior experience in meta-analysis can benefit both from the theoretical introduction and the simulation approach. However, for a more comprehensive overview of meta-analysis topics the reader may refer to meta-analysis textbooks (Borenstein et al., 2009; e.g., Harrer, Cuijpers, & Ebert, 2019)

We assume the reader is familiar with the basic concepts of R, but core functions will be explained. Code and materials are available on the OSF repository (https://osf.io/54djn/). A theoretical introduction, simulation examples for multivariate and multilevel models, and more details about the coding approach are available in the supplementary materials.

1.1 Meta-analysis introduction

The meta-analysis is a statistical procedure to combine multiple studies (i.e., *primary studies*) into a single statistical analysis (Borenstein et al., 2009). The idea is that combining numerous preliminary studies improve the estimation of a particular phenomenon more efficiently compared to conducting a single study. In statistical terms, the concept of the meta-analysis is to switch the statistical unit from the single participant or observation (i.e., *level 1*) to the study (i.e., *level 2*). Given that some studies will give more information because their estimation variability is smaller (e.g., higher sample size), the meta-analysis combines the studies assigning more weight as a function of the precision (i.e., the inverse of the variance).

As an example that will be used throughout the paper, we consider the efficacy of memory training in improving memory performance during a cognitive task. The typical primary study will collect data from a group of participants receiving the memory training (*experimental group*) and another group receiving a control treatment (*control group*). The focus of the meta-analysis is collecting multiple studies with similar aims and methods and estimating the average effect of memory training. Despite differences in the type of cognitive task or experimental setup, each primary study collects an *experimental group* (n_T) and *control group* (n_C) and computes the average performance (*experimental group* $\overline{T}$ and *control group* $\overline{C}$) and standard deviations (*experimental group* s_T and *control group* s_C).

1.1.1 Effect size and variance.

The first step of a meta-analysis is to extract information from included studies. This common measure should give an immediate idea of the direction (i.e., the treatment improves or reduces performance) and the size of the effect. Standardized effect size measures (see Lakens, 2013 for an overview), such as the Standardized Mean Difference (SMD), which could be estimated by Cohen's d (Cohen, 1988) or the Pearson correlation coefficient ρ which could be calculated using the

associated sample estimator r are commonly used to compare heterogeneous outcome variables[1]. If all studies used the same raw measure, such as reaction times, it is possible to directly meta-analyze the studies without standardizing e.g. using the unstandardized mean difference (UMD, Borenstein et al., 2009, pp. 21–24).

As reported in the previous section, beyond the effect size of each study, we need to assign a weight according to the precision. For this reason, we need to calculate the sampling variability of the effect size or the raw measure that will represent the estimation precision. Each raw or standardized effect size measure (e.g., raw mean difference, Cohen's d, and Pearson's correlation) has a different formula to calculate the sampling variability. The idea is to choose the appropriate measure considering the study design (e.g., between vs. within-subjects) and available information and find the proper formula to compute the sampling variability. In addition, there are formulas and approaches to convert from one effect size measure to another (Borenstein et al., 2009; Lakens, 2013; Lipsey & Wilson, 2001).[2] Usually, the effect size sampling variability depends mainly on the sample size that determines the weight assigned during the meta-analytic estimation.

1.1.2 Equal vs fixed vs random-effects model. The core of a meta-analysis is combining the results of multiple studies giving more weight to studies that provide a more precise effect estimation. We can find essentially three meta-analysis models: the equal-effects, the fixed-effects, and the random-effects model (Hedges & Vevea, 1998; Laird & Mosteller, 1990). The *equal-effects* model assumes that each study included in the meta-analysis is a more or less precise estimation of the true underlying effect (θ). In other terms, we are assuming that there is no variability (i.e., heterogeneity) among true effect sizes. On the other side, suppose some study-level characteristics (e.g., participants' age, sex, or socioeconomic status) or the experimental paradigm (e.g., type of memory task or

[1] When using bounded effect sizes such as the Pearson correlation, Fisher Z-transformation is commonly used (Borenstein et al., 2009, pp. 41–43)

[2] See also the blog post by James Pustejovsky (www.jepusto.com/alternative-formulas-for-the-smd) for a general method to compute the sampling variability for standardized mean difference measures.

difficulty) could impact the treatment effect. In this case, we have true variability (i.e., heterogeneity) among studies. The *random-effects* model assumes a distribution of real effects with mean μ_θ and variance τ^2 thus estimating the heterogeneity. Finally, the *fixed-effects* model estimate the average effect of the included pool of studies ignoring the presence of heterogeneity. While the distinction between the equal and fixed-effects model is theoretical, the estimated model is the exactly the same[3]. The *random-effects* model assumes and estimates heterogeneity among effect sizes leading to different model parameters. From an inferential point of view, the *random-effects* model provides unconditional inference to the population of effect sizes while the *fixed-effects* model estimates the average effect of the selected studies and not population-level parameters. The *equal-effects* model assumes a single true population effect to be estimated. Figure 1 depicts the theoretical distinction between the equal and random-effects model. In the presence of heterogeneity, we can include moderators (e.g., the type of experimental setup) into a meta-regression model to explain the effect size heterogeneity. An interesting proposal to combine evidence from these different models into a single analysis is called Bayesian model-averaged meta-analysis (Berkhout, Haaf, Gronau, Heck, & Wagenmakers, 2023; Gronau, Heck, Berkhout, Haaf, & Wagenmakers, 2021). However, this model is beyond the scope of the tutorial.

[3] See wviechtb.github.io/metafor/reference/misc-models.html for a detailed explanation

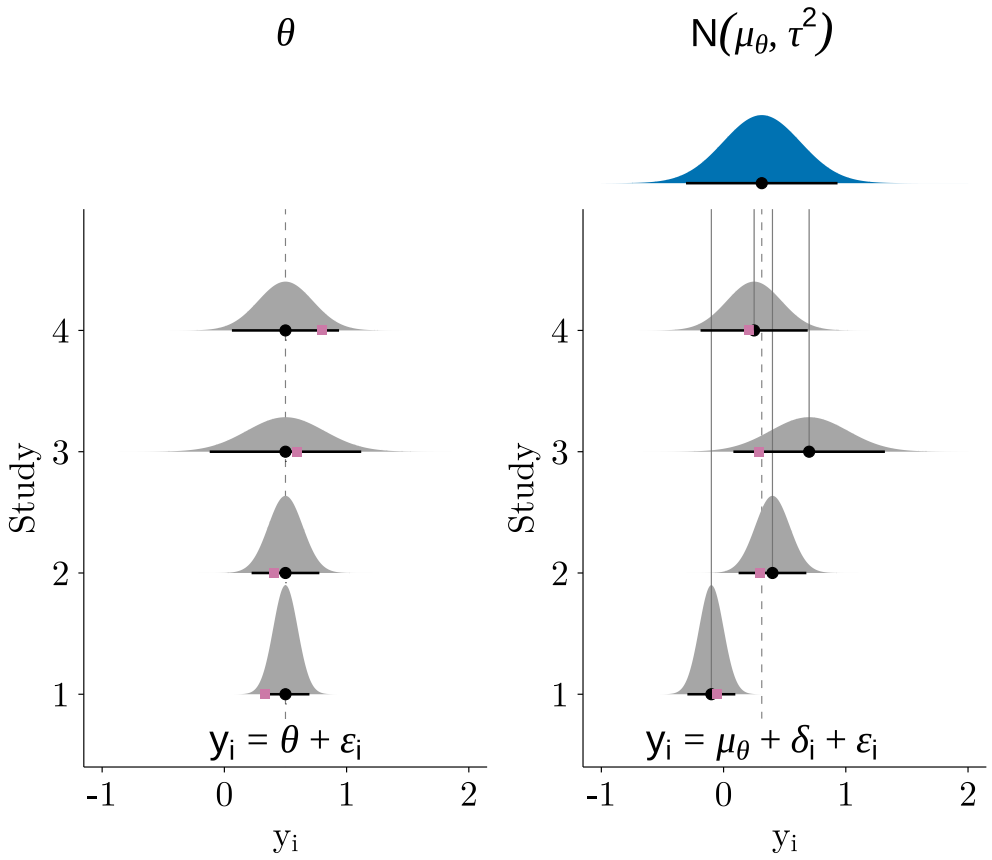

Figure 1. The difference between the assumptions of the *equal* and *random-effects model*. Each distribution depicts the sampling distribution of $k = 4$ hypothetical studies ($i = 1, 2, ..., 4$) with a certain observed effect size y_i (pink squares), sampling variability $\sigma^2_{\epsilon_i}$ and the 95% confidence interval (black segment). The *equal-effects* plot on the left suggests that each observed effect size has the same underlying true effect θ (each distribution has the same mean) with a different degree of precision (e.g., Study 1 is more precise than Study 3). In practice, each study has a different observed effect size where studies with high precision (i.e., narrow sampling distributions) will be close to the real effect (θ). The *random-effects model* on the right suggests that beyond the error term (ϵ_i), each real effect size is composed of a fixed part (now μ_θ) and a random part (δ_i) sampled from a normal distribution with mean zero and variance τ^2. When τ^2 is zero, the random-effects reduces to a equal-effects model.

2 Simulation

2.1 Monte Carlo simulations

The Monte Carlo methods are controlled experiments (Gentle, 2009). Given a set of fixed parameters, probability distributions, and the possibility of generating random

numbers, it is possible to simulate the behavior of an empirical system. Monte Carlo simulations are used for statistical and mathematical problems that cannot be solved analytically.

A straightforward example regards estimating the sampling variability of the mean difference. When calculating the mean difference between two samples, we estimate the true mean difference at the population level with a certain degree of error (i.e., the standard error of the mean difference). The central limit theorem states that the difference between the means of two random samples ($\overline{X}$ and $\overline{Y}$) is approximately normally distributed with mean $\mu_x - \mu_y$ and standard error $\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$. The same results can be obtained using Monte Carlo simulations using the following procedure:

1. Generating two random samples from two normal distributions with a fixed mean difference
2. Calculating the mean difference
3. Repeating the same process, many times
4. Calculate the standard deviation of the simulated values

Using the method, we are estimating via simulation the standard deviation of the sampling distribution of the mean difference (i.e., the standard error of the mean difference). Increasing the number of simulations will produce more stable results.

```r
set.seed(123) # seed for simulation to ensure the reproducibility of results
es <- 0.5 # real mean difference in the populations
sigma <- 1 # real standard deviations in the population
n <- 30 # sample size for both groups
nsims <- 1e5 # number of simulations = 100000

# simulates the sampling distribution of the mean difference
d_mc <- replicate(nsims, expr = {
  g1 <- rnorm(n = n, mean = es, sd = sigma)
  g2 <- rnorm(n = n, mean = 0, sd = sigma)
  mean(g1) - mean(g2) # calculates the sample mean difference and returns the value
```

```
})

# analytical standard error of the mean difference
se_an <- sqrt(sigma^2/n + sigma^2/n)

# monte-carlo standard error of the mean difference
se_mc <- sd(d_mc)
```

The standard error estimated solving analytically is 0.26 and using the Monte Carlo simulation we arrive at the same result (i.e., 0.26).

2.2 Simulation setup

This tutorial uses several R packages for the simulations, meta-analysis fitting, and figures/tables. For the data manipulation, we used the *tidyverse* (Wickham, 2023) package. For the models fitting we used the *metafor* (Viechtbauer, 2010) package. For figures and tables the *ggplot2* (Wickham et al., 2023) or the `metafor::forest()` function, *kableExtra* (Zhu, 2021) and *papaja* (Aust & Barth, 2022) packages. We set the `seed` for reproducibility of the simulation environment.

```
library(dplyr) # for data manipulation (within tidyverse)
library(tidyr) # for data manipulation (within tidyverse)
library(metafor) # for meta-analysis models fitting
library(purrr) # for *apply like functions (within tidyverse)
seed <- 2023 # general seed for all simulations
```

Before diving into the specific simulations, in this section, we define the common aspects of all simulations in the following sections. The current paper focuses on the *two-level*, *equal,* and *random-effects* models. All the examples refer to primary studies that assess the efficacy of a treatment by comparing a control and an experimental group. In simulation studies where the purpose is not evaluating effect size estimators, it is convenient to simulate unstandardized effect size measures (see Viechtbauer, 2005, 2007). The estimator is unbiased thus not requiring small sample correction (Hedges, 1981; e.g.,

Hedges, 1989). Furthermore, the effect size and the sampling variance are independent. Similar to the simulation approach by Viechtbauer (2005, 2007) the *experimental* group (T) and *control* groups C are sampled from normal distributions respectively $T_i \sim \mathcal{N}(\Delta, 1)$ and $C_i \sim \mathcal{N}(0, 1)$ where Δ is the unstandardized mean difference (UMD). The UMD and the sampling variance are calculated using Equations (1) and (2) where y is the estimated value of Δ.

$$D = \overline{T} - \overline{C} \tag{1}$$

$$\sigma^2_{\epsilon D} = \frac{s^2_T}{n_T} + \frac{s^2_C}{n_C} \tag{2}$$

We can use the following algorithm implemented in the `sim_study()` function to simulate a single study. Before using the `sim_study()` function, we can create a data frame for the simulation using the `make_data()` function[4]. Table 1 depicts an example of the `make_data()` output.

1. Choose a Δ, n_T, and n_C value.
2. Simulate n_T observations from a Gaussian distribution with $\mu = \Delta$ and $\sigma^2 = 1$ and n_T observations from a Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$. In this way, the expected difference between groups will be Δ and the expected variance for each group is 1.
3. Calculate the observed effect size y and the sampling variance σ^2_ϵ.

[4] This function is a simple wrapper of `data.frame()` that given the number of studies and other variables create the data structure for the simulation.

The simulation approach can be easily extended calculating a standardized effect size measures (e.g., Cohens' d Cohen, 1988) and the corresponding sampling variance. For example, after generating data for the two groups the mean difference can be standardized using the pooled standard deviation and applying the appropriate correction (Hedges, 1981; e.g., Hedges, 1989).

```
make_data <- function(k, nt = NULL, nc = NULL, ...){
  params <- c(as.list(environment()), list(...))
  cols <- params[!sapply(params, is.null) & names(params) != "k"]
  dat <- data.frame(
    id = 1:params$k
  )
  if(length(cols) != 0){
    dat <- cbind(dat, cols)
  }
  return(dat)
}
```

```
sim_study <- function(es, nt, nc = NULL, aggregate = TRUE){
  if(is.null(nc)) nc <- nt
  # generate from normal distribution
  yc <- rnorm(nc, 0, 1)
  yt <- rnorm(nt, es, 1)

  # effect size
  yi <- (mean(yt) - mean(yc))

  # sampling variance
  vi <- var(yt)/nt + var(yt)/nc

  if(!aggregate){
    # return raw data
    data.frame(id = 1:(nc + nt),
               group = rep(c("c", "t"), c(nc, nt)),
               y = c(yc, yt))
  }else{
    # compute effect size
    data.frame(yi, vi)
```

```
  }
}

sim_studies <- function(..., data = NULL){
  # ... (dots) are the arguments passed to mapply, with the order required by sim_study
  res <- mapply(sim_study, ..., SIMPLIFY = FALSE)
  # everything to a dataframe
  res <- dplyr::bind_rows(res)

  if(!is.null(data)){
    # attach to the original dataset
    cbind(data, res)
  }else{
    res
  }
}
```

The `es` is the true effect size (μ_θ for the random-effects model and θ for the equal-effects model), `nc` and `nt` are the sample size for the control and experimental group, `aggregate` controls if returning the effect size and the corresponding sampling variance or the participant-level data. We can generate a single study with the desired parameters with this function. A suggestion for each simulation step is to generate a large n to reduce the sampling error and check the recovery of simulated parameters. This is a general strategy that can be applied to every simulation. The `sim_studies()` function will iterate through variables in the `...` argument creating the meta-analysis data frame. The `mapply()` function is clearly explained in the supplementary materials.

The following code simulates a single study with $n = 10000$ and check the estimated mean and standard deviation.

```
set.seed(seed)
sim_study(es = 0.3, nc = 10000, nt = 10000, aggregate = FALSE)
```

The control group has a mean of -0.008 (SD = 0.996) and the experimental group has a mean of 0.306 (SD = 1.013) which are remarkably close to the simulated values.

Using the `sim_study()` function multiple times (with the appropriate adjustments) we can generate a series of studies simulating a data set for a meta-analysis (using the `sim_studies()` function). After each example, we will compute the appropriate model (e.g., *equal* or *random-effects*) using the `metafor` package (Viechtbauer, 2010) to check the recovery of simulated parameters[5]. Table 2 summarizes the notation used in simulations and code in equations and code.

```r
k <- 30 # number of studies
n <- 20 # number of participants per group, per study
es <- 0.3 # real effect size
sim <- make_data(k = k, nt = n, nc = n, es = es)
```

Table 1
Example of data generated with the ***make_data()*** *function. The* ***id*** *column is the identifier for each study.*

id	nt	nc	es
1	20	20	0.3
2	20	20	0.3
...	...	...	...
29	20	20	0.3
30	20	20	0.3

[5] By default `metafor` use the Wald z test on model parameters but other inference methods are available e.g. the Knapp and Hartung (2003) method by setting `test="knha"` within the `rma()` function. This method is useful especially when the number of studies is low (see wviechtb.github.io/metafor/reference/misc-recs.html for more details)

Table 2
Variables in the data-generating model and associated R code.

Equation	Code	Description
θ	`theta`	Equal-effects model true effect size
μ_θ	`mu_theta`	Random-effects model true effect size
τ^2	`tau2`	Effect sizes heterogeneity
τ_r^2	`tau2r`	Residual effect sizes heterogeneity
$\overline{T}, \overline{C}$		Mean of the experimental/control group
$s_{T,C}$		Standard deviation of the experimental/control group
$n_{T,C}$	`nt, nc`	Sample size of the experimental/control group
y_i	`yi`	Observed effect size
$\sigma^2_{\epsilon_i}$	`vi`	Observed sampling variance
δ_i	`deltai`	Random effect for the study i
β_0	`b0`	Meta-regression intercept
β_1	`b1`	Meta-regression slope
k	`k`	Number of studies
ϵ_i		Sampling error for the study i

Note. The `yi` and `vi` notation for the observed effect size (y_i) and sampling variance ($\sigma^2_{\epsilon_i}$) has been used to be consistent with the `metafor` notation.

2.3 Equal-effects model

The most basic model to simulate is the *equal-effects model* (EE). As reported in the previous sections, the *equal-effects model* assumes the presence of a single true effect (θ) and the observed variability is caused by each effect size being a more or less imprecise estimation of the true effect. In other words, the only source of variability is the sampling variability that depends on the variance of the primary studies (i.e., the sample size). Equations (3) and (4) formalize the *equal-effects* model.

$$y_i = \theta + \epsilon_i \tag{3}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon_i}) \tag{4}$$

Each observed effect size (y_i) is composed of the real effect size θ plus an error term (ϵ_i) that is sampled from a normal distribution with $\mu = 0$ and $\sigma^2 = \sigma^2_{\epsilon_i}$ (i.e., the known sampling variance of the study i). As demonstrated in Equation (1), the increase in sample size will decrease the sampling variability. A study with a extremely large sample size will essentially have θ as the observed effect size. Since we are sampling participants' level data, the error component is already included in the `sim_study()` function. To simulate this model in R we can just call the `sim_study()` multiple times according to the number of desired studies (k). In addition, we simulate that each primary study will have a sample size of $n = 20$ for both groups. We will discuss later the appropriateness of this assumption.

```r
set.seed(seed)
k <- 30 # number of studies
n <- 20 # number of participants per group, per study
theta <- 0.3 # real effect size

sim <- make_data(k = k, nt = n, nc = n, es = theta)
sim <- sim_studies(sim$es, sim$nc, sim$nt, data = sim) # simulated data
res <- rma(yi = yi, vi = vi, method = "EE", data = sim)
```

We are using the θ parameter to generate random data using the `sim_study()` function that will introduce the random error component ϵ_i. Then we can fit the *equal-effects* meta-analysis model with the `rma` function and `method = "EE"` of the `metafor` package. The parameter we are estimating is θ which is close to our simulation value (see Table 3). Increasing the number of studies (k) and/or the number of participants in each study (n) will improve estimation reducing the standard error.

Table 3
Summary of the simulated equal-effects model. The only estimated parameter is the average effect θ (β) with the standard error, 95% confidence interval, and the Wald z-test. The z-test evaluates the null hypothesis that the real effect equals zero.

	β	95% CI	z	p
overall	0.347 (SE = 0.056)	$[0.237, 0.456]$	6.223	< 0.001
$k = 30$				

Increasing the sample size for each study will increase the estimation precision thus the variability among studies will be reduced. This can be easily demonstrated by simulating studies with a high sample size as reported in Figure 2. As the sample size increases the only source of variability (i.e., the error component ϵ_i) is close to zero so each study is closer to the true simulated value.

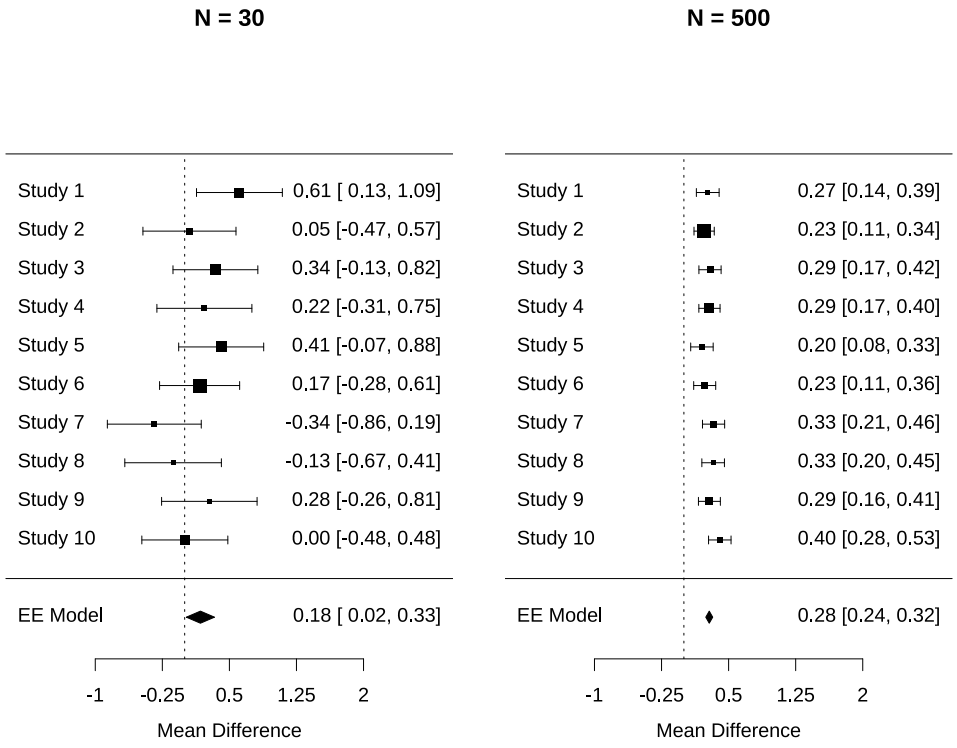

Figure 2. Forest plots of two simulated *equal-effects models*. On the left, the simulated model has $nt, nc = 30$ for each included study while on the right the sample size for each study is $nt, nc = 500$. Given that the data were generated under a *equal-effects model*, when the sample size is high (on the right) each study is aligned on the real effect size because the error component (ϵ_i) is close to zero. The average effect size is similar (depending on the random numbers generation) between the two scenarios while the estimation precision (the width of the black diamond) is narrower on the right.

2.4 Random-effects model

The *random-effects model* (RE) can be considered an extension of the *equal-effects model*. The *equal-effects model* assumes that the real effect is a single value. The random-effects model relaxes this assumption allowing the true effect size to vary across studies. For example, the difference between groups we are simulating could be influenced by the type of experiment or the participants' age. Now θ is no longer a single value but a distribution of values. Due to the effect size being a distribution, we need to estimate both the mean and the variance. The parameter μ_θ is mean of the distribution, interpreted as the average effect size across different true effect sizes and τ^2 is the variance of the distribution interpreted as variability or heterogeneity of effect sizes. In practical terms, we now have two sources of variability: τ^2, which express the real difference among effect sizes, and $\sigma^2_{\epsilon_i}$ which is the known sampling variance of each study as in the *equal-effects model*. We can easily extend Equation (3) with Equations (5), (6) and (7).

$$y_i = \mu_\theta + \delta_i + \epsilon_i \tag{5}$$

$$\delta_i \sim \mathcal{N}(0, \tau^2) \tag{6}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon_i}) \tag{7}$$

Compared to the *equal-effects model* we need to generate another adjustment to the overall effect μ_θ from a normal distribution with mean 0 and variance τ^2. The real effect

size for the study i will be $\mu_\theta + \delta_i$ where δ_i are the random-effects regulated by the τ^2 parameter.

```r
set.seed(seed)
k <- 30 # number of studies
n <- 20 # number of participants per group, per study
mu_theta <- 0.3 # real effect size
tau2 <- 0.2 # the effect size heterogeneity

sim <- make_data(k = k, nc = n, nt = n, mu_theta = mu_theta)

# simulate the random-effects adjustment
sim$deltai <- rnorm(k, 0, sqrt(tau2))

# adding the by-study adjustment to the average effect
sim$es <- sim$mu_theta + sim$deltai

# now we are using mu_deltai and no longer theta
sim <- sim_studies(sim$es, sim$nc, sim$nt, data = sim)
res <- rma(yi = yi, vi = vi, method = "REML", data = sim)
```

Then we can fit the *random-effects* meta-analysis model with the `rma` function and `method = "REML"`. Table 4 depicts the model results[6] of the `metafor` package. We are now estimating two parameters: μ_θ and τ^2. As for the *equal-effects model* , increasing the number of studies and/or the number of participants in each study will improve estimation reducing the standard error.

[6] The `method = "REML"` is not the only method to estimate a random-effects model. See the `rma` documentation (https://wviechtb.github.io/metafor/reference/rma.uni.html#specifying-the-model) for an overview of other estimation methods

Table 4
Summary of the simulated random-effects model: Compared to the equal-effects model, there are more parameters. The β is the average effect (μ_θ) with the standard error, 95% confidence interval, and the Wald z-test. The τ^2 is the estimated heterogeneity and the I^2 (explained in Section 2.4) represents the percentage of total variability due to between-study heterogeneity.

	β	95% CI	z	p
overall	0.394 (SE = 0.110)	$[0.179, 0.609]$	3.596	< 0.001

$k = 30$
$\tau^2 = 0.262$ ($SE = 0.095$)
$I^2 = 73.480\%$

An important aspect of the *random-effects model* is the interplay between the heterogeneity (τ^2) and the sampling variability (σ_i^2). As the number of studies k increases, the estimation of μ_θ and τ^2 becomes more precise (Blázquez-Rincón, Sánchez-Meca, Botella, & Suero, 2023; Rubio-Aparicio et al., 2018) . Additionally, as the sample size of each study decreases, each δ_i will be estimated with higher precision but as long as $\tau^2 \neq 0$ there will be variability among effect sizes. In other words, increasing the sample size of each study or the number of studies will not affect the value of τ^2 but only the estimation precision (see Borenstein et al., 2009, Chapter 16, Figure 16.6 for a clear explanation). This can be easily demonstrated using the previous simulation, increasing each study sample size. Figure 3 depicts the same meta-analysis but with different precision in estimating the study y_i. Compared to Figure 2, increasing the sample size of primary studies improves the estimation of each study without reducing the between-studies heterogeneity.

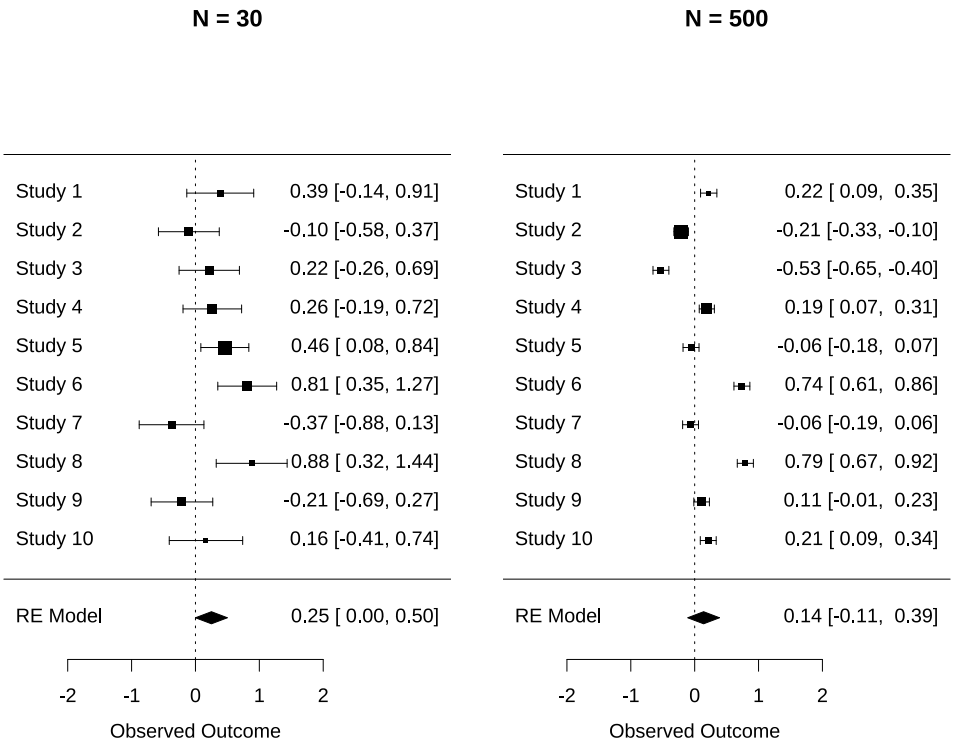

Figure 3. Forest plots of two simulated *random-effects models*. On the left, the simulated model has $nt, nc = 30$ for each included study, while on the right, the sample size for each study is $nt, nc = 500$. Compared to Figure 2, data were generated under a *random-effects model*. The estimated average effect is similar between the two scenarios regarding average effect and precision. However, compared to the *equal-effects* simulation, increasing the sample size of primary studies only affects the precision without reducing the true heterogeneity (i.e., $\tau^2 \neq 0$).

The relationship between the sampling error and the heterogeneity can be expressed using the I^2 statistics (e.g., Higgins & Thompson, 2002) that is the percentage of the total variability $\tau^2 + \tilde{v}$ that is attributable to real heterogeneity between studies (τ^2). In Equation (8), $\tilde{v}$ is the *typical* within-study sampling variability (i.e., a summary statistics of sampling variances of included studies) as proposed by Higgins and Thompson (2002)[7] and implemented in Equation (9) with $w_i = 1/\sigma^2_{\epsilon_i}$ and k is the number of studies.

[7] See https://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate#fn___1 for an overview of the I^2 statistics also for complex models. There are also other proposals to compute the $\tilde{v}$ statistics (see Takkouche, Cadarso-Suárez, & Spiegelman, 1999; Takkouche, Khudyakov, Costa-Bouzas, & Spiegelman, 2013)

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \tilde{v}} \tag{8}$$

$$\tilde{v} = \frac{(k-1)\sum w_i}{(\sum w_i)^2 - \sum w_i^2} \tag{9}$$

From Equation (8) and Figure 3 is clear that if each included study has a considerable sample size the sampling variability ($\tilde{v}$) will be reduced, and the total variability will be mainly driven by real heterogeneity (τ^2) . This is the crucial difference between the equal-effects and the random-effects model (see also Borenstein et al., 2009, pp. 117–122).

Given the interpretation of I^2 it is possible to simulate a meta-analysis fixing a certain I^2 value. The only caveat is fixing the $\tilde{v}$. When the sample size of each study is the same[8], $\tilde{v} = \sigma^2_{\epsilon_i}$. In the other case, $\tilde{v}$ needs to be calculated from sampling variances as reported in Higgins and Thompson (2002) and cannot be easily fixed a priori. With the assumption of homogeneous sample size across studies[9], we can solve Equation (8) for τ^2 obtaining the heterogeneity value associated with a certain I^2 as reported in Equation (10).

Table 5 depicts the results of the random-effects model fixing the I^2 value. Choosing a meaningful τ^2 for the simulation can be sometimes difficult. Beyond fixing the I^2 value, a possibility is choosing plausible τ^2 value from the literature. For example van Erp and colleagues (2017) estimate an empirical τ^2 distribution across several published meta-analyses that can be used to simulate a plausible scenario.

---

[8] The assumption of studies with homogeneous sample sizes is commonly used to calculate the statistical power (see Borenstein et al., 2009, Chapter 29)

[9] In case of heterogeneous sample sizes generated from a probability distribution (e.g., Gaussian) the *typical* sampling variance calculated using the expected value of the distribution can be used to somehow fix the (average) I^2 (see the supplementary materials for a simulated example)

$$\tau^2 = -\frac{I^2\tilde{v}}{I^2 - 1} \tag{10}$$

```r
set.seed(seed)
k <- 30 # number of studies
mu_theta <- 0.3 # real effect size
n <- 30 # sample size per group, per study
I2 <- 0.6 # desired I2 value

v <- (n + n)/(n * n) + mu_theta^2/(2 * (n + n - 2)) # typical within study variance
tau2 <- -((I2*v)/(I2 - 1))

sim <- make_data(k = k, nc = n, nt = n, mu_theta = mu_theta)

sim$deltai <- rnorm(k, 0, sqrt(tau2))
sim$es <- sim$mu_theta + sim$deltai
sim <- sim_studies(sim$es, sim$nc, sim$nt, data = sim)
res <- rma(yi, vi, method = "REML", data = sim)
```

Table 5
Summary of the random-effects model fixing the I^2 value. Parameters are the same as described in Table 4.

	β	95% CI	z	p
overall	0.300 (SE = 0.071)	[0.160, 0.440]	4.201	< 0.001

$k = 30$
$\tau^2 = 0.087$ $(SE = 0.040)$
$I^2 = 57.904\%$

2.5 Meta-regression

From a linear regression perspective, both the EE and the RE models can be seen as *intercept-only* models where only the mean (i.e., the linear regression intercept) is estimated. As reported in the meta-analysis introduction, the between-study heterogeneity usually represents the true variability of the effect due to differences among primary studies. A natural extension of the *intercept-only* analysis is a model that includes

variables (i.e., *moderators*) that could explain the observed heterogeneity among effect sizes. For example, a group of studies could use a particular memory task where the expected effect is higher than another. In this way, considering the type of task will explain part of the observed heterogeneity as in standard regression models. Figures 4 and 5 depict a random-effects meta-regression model for a categorical and numerical predictor.

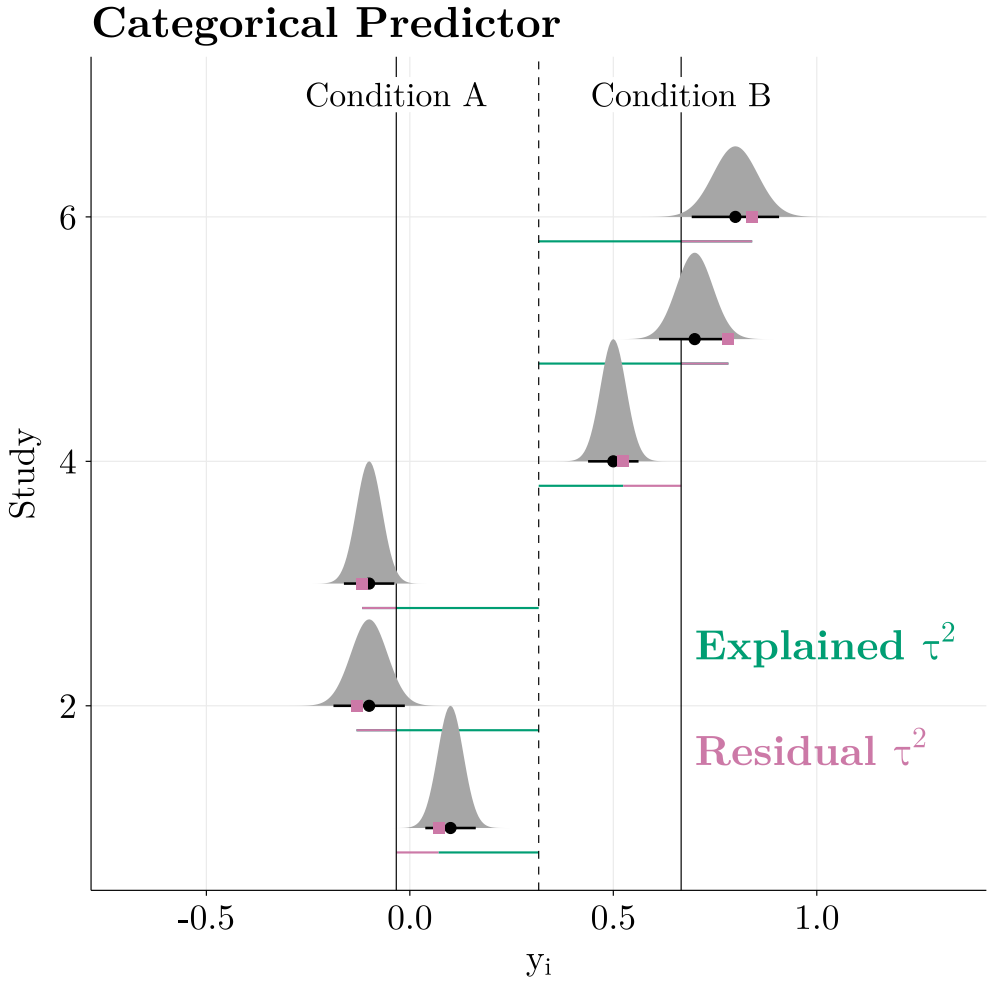

Figure 4. Graphical representation of a *random-effects meta-regression* model with a categorical predictor (Condition A and Condition B). Each grey distribution represents the sampling distribution of included studies. The dotted line is the average effect (i.e., random-effects model without moderators). The effect size differs between conditions A and B, including the *condition* moderator explaining part of the total heterogeneity (pink plus green segments). The green segments depict the explained heterogeneity, and the pink segments the residual (unexplained) heterogeneity. The pink squares are simulated observed effect sizes from the sampling distributions.

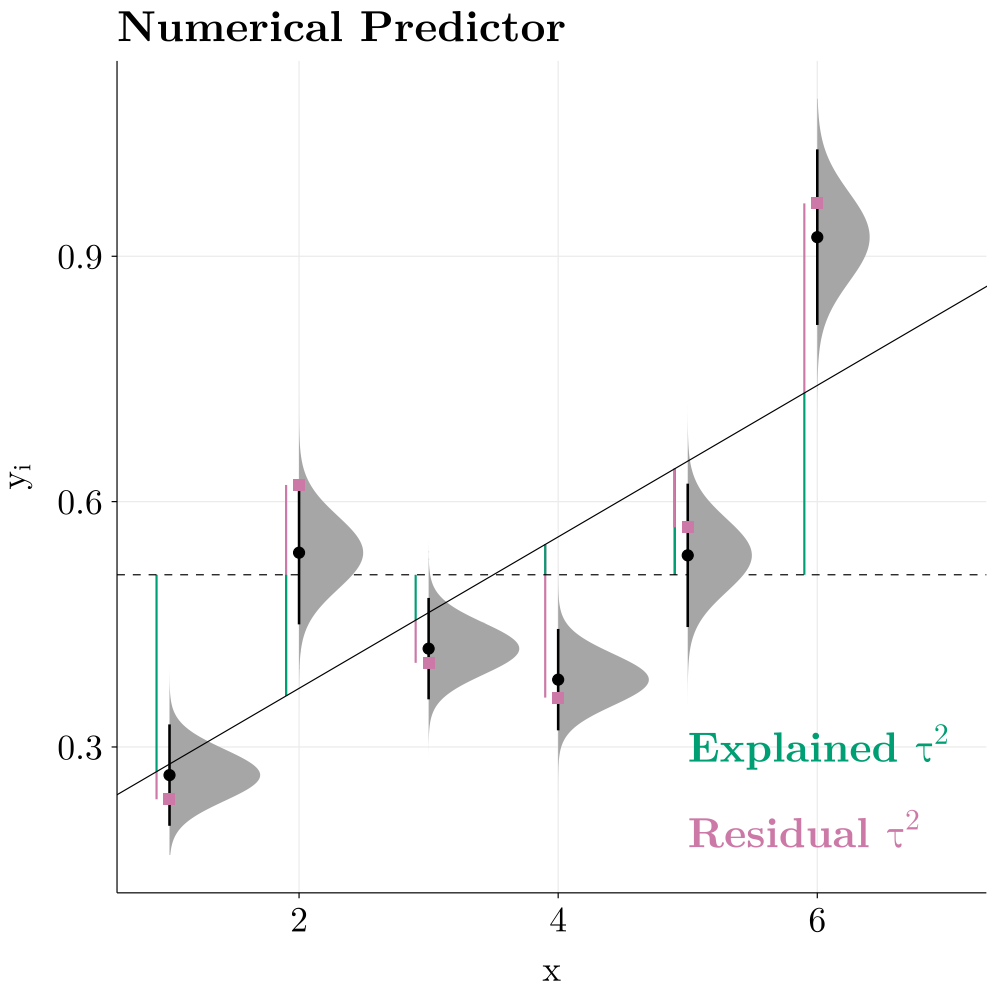

Figure 5. Graphical representation of a *random-effects meta-regression* model with a numerical predictor (x). Each grey distribution represents the sampling distribution of included studies. The dotted line is the average effect (i.e., random-effects model without moderators). The effect size increases as a function of the x variable. Therefore, including x as a predictor explains the heterogeneity. The rest of the Figure follows the same logic as Figure 4.

2.5.1 Meta-regression with a categorical moderator. A common example of meta-regression is by including a categorical predictor, including information about study-level features. In our example, a group of studies uses an online memory task while others use a standard lab-based task. The Equation (5) can be easily extended for a meta-regression model by including a variable encoding the type of task (online vs. lab-based) and the expected difference between the two levels of the moderator (i.e., the *lab vs. online effect*). In regression terms (see equation (11)), we could use a dummy

variable (X_1) that takes the value of 0 for the *lab-based task* (L) and a value of 1 for an *online task* (O)[10] . Now we fix β_1 to be the *lab vs. online effect* (i.e., the expected mean difference between the two groups of studies) and the product between $\beta_1 X_{1_i}$ will consider the *lab vs. online effect*. Crucially, despite τ^2 is still the heterogeneity between effect sizes, now we need to fix τ^2 considering that we included a moderator. In other terms, $\delta_i \sim \mathcal{N}(0, \tau_r^2)$ where τ_r^2 is the residual heterogeneity after including the moderator. We can describe our model using equation (12) according to the value of X_{1_i} (L = lab-based experiments, O = online experiments).

$$y_i = \beta_0 + \delta_i + \beta_1 X_{1_i} + \epsilon_i \tag{11}$$

$$\begin{aligned} y_{L_i} &= \beta_0 + \delta_i + \beta_1 \times 0 + \epsilon_i \\ y_{O_i} &= \beta_0 + \delta_i + \beta_1 \times 1 + \epsilon_i \end{aligned} \tag{12}$$

We can simulate the same scenario of the random-effects model with $k_O = 15$ (*online tasks*) and $k_L = 15$ (*lab-based tasks*). Then we fix the $\beta_1 = 0.2$ and $\tau_r^2 = 0.1$ (r for residual).

```
set.seed(seed)
k <- 30 # the total number of studies
b0 <- 0.1 # intercept, the effect size of the lab-based studies
b1 <- 0.2 # the difference between the two levels of the moderator
tau2r <- 0.1 # the residual heterogeneity
n <- 30 # the sample size per group, per study
```

[10] The *dummy coding* (also known as treatment coding) is the default in R and `metafor` but other coding schemes could be used (see Schad, Vasishth, Hohenstein, & Kliegl, 2020 for an overview of contrast coding schemes). As in standard linear regression, the contrast coding of categorical factors or the presence/absence of the intercept influences the interpretation of estimated model parameters and omnibus meta-regression tests (see https://www.metafor-project.org/doku.php/tips:models_with_or_without_intercept for a detailed discussion)

```r
sim <- make_data(k = k, nc = n, nt = n, exp = rep(c("lab", "online"), each = k/2))

sim$deltai <- rnorm(k, 0, sqrt(tau2r)) # the by-study residual adjustment
sim$es <- b0 + sim$deltai + b1*ifelse(sim$exp == "lab", 0, 1)

sim <- sim_studies(sim$es, sim$nc, sim$nt, data = sim)
res <- rma(yi, vi, mods = ~exp, method = "REML", data = sim)
```

Now we can fit the meta-regression model with the `rma` function as for the random-effects model with the addition of `mods = ~ exp` that indicates which variables to consider as moderators. The results are presented in Table 6. Now the model will estimate an *intercept* parameter (i.e., β_0) that is the value of y when X_1 is zero (i.e., for lab-based studies) or in other terms the expected value for lab-based studies. Then the β_1 parameter represents the estimated difference in y between the values of X_1 (i.e., lab-based vs online experiments). As said before, τ^2_r is now the residual heterogeneity that is interpreted as the variability between effect sizes after controlling for the moderator X_1.

Table 6
Summary of the random-effect *model with a categorical predictor (lab vs online experiments). The* ***intercept*** *is the average effect for lab-based experiments and* ***exponline*** *is the difference between lab-based and online experiments. The R^2 is the percentage of explained heterogeneity and τ^2_r is the estimated residual heterogeneity. Other parameters are the same as the standard* random-effect *model (see Table 4).*

	β	95% CI	z	p
intercept	0.038 (SE = 0.101)	$[-0.161, 0.236]$	0.372	0.710
exponline	0.323 (SE = 0.143)	$[0.043, 0.602]$	2.263	0.024

$k = 30$
$\tau^2_r = 0.087$ $(SE = 0.041)$
$R^2 = 20.659\%$
$I^2 = 57.741\%$

2.5.2 Meta-regression with a numerical moderator. The same approach can be used for a continuous predictor. For example, we can simulate that the average participant's age within each study could explain part of the observed heterogeneity. Now the X_1 is a continuous predictor representing the average age for each study and β_1 is the

effect size increase for a unit increase (i.e., 1 year) of average age. Sometimes guessing a plausible β_1 value with a continuous predictor is not straightforward. A first strategy could be to use values estimated from the literature. Another approach consists in setting up the model and simulating several expected y_i and calculating the range of simulated values. A third possibility is fixing the proportion of explained heterogeneity calculating the β_1 value accordingly. As in standard regression analysis, we can use the R^2 statistic to describe the amount of heterogeneity explained by the included moderators. Equation (13) reports how to calculate the R^2 for a meta-regression model. The τ_r^2 is the residual heterogeneity after considering the moderators and τ_f^2 is the heterogeneity estimated without considering the moderators. In the next sections, we will present an example of the simulation-based and the R^2 based approaches.

$$R^2 = 1 - \frac{\tau_r^2}{\tau_f^2} \tag{13}$$

In terms of regression parameters now the intercept (β_0) is no longer the overall effect or the average of one category as in the previous example but the estimated value for a specific X_1 thus for a specific age. If X_1 is a variable representing the average age for each study, then the β_0 is the average effect size when the age is zero. Depending on the moderator, the intercept is interpreted in different ways. For example, with the age, the intercept has no empirical meaning given that no studies could have a participant average age of zero. A strategy could be to mean-center the age (i.e., subtracting from each study age the average age across the study). Now the intercept is still the average effect size when the age is zero but now zero is the average age. Importantly, the contrast coding for categorical predictors or centering numerical variables does not affect the overall model but only parameters values and interpretation.

2.5.2.1 Assessing the impact of β_1.

As reported in the previous section, a strategy to guess plausible values for β_1 is by simulating several expected y_i given the meta-regression equation and summarizing or plotting the effect size range. The range of simulated y_i values are also affected by the simulated age values across studies. However, it is probably more intuitive to guess a plausible range of moderator values compared to the β_1 value. In this specific example, if all studies target a specific population (e.g., adults below 50 years), the expected average age range can be easily simulated. In our case, we simulated k average age values from a uniform distribution $\text{age}_i \sim U(20, 40)$. Then we can plot the distribution of y_i values to check the plausibility of simulated values. As shown in Figure 6, with the same range for the moderator, a $\beta_1 = 0.1$ gives a plausible effect sizes range while a $\beta_1 = 0.7$ predicts very extreme values.

```r
set.seed(seed)
k <- 1000 # number of studies
b0 <- 0.3 # the intercept i.e., average yi when x is 0
b1 <- c(0.1, 0.7) # the beta1 i.e., the increase in yi for an increase in 1 year
tau2r <- 0.1 # the residual heterogeneity after including x1
n <- 30 # number of participants per study, per group
x1 <- runif(k, 20, 40) # random mean-age for each study
x10 <- x1 - mean(x1) # centering the age

sim <- tidyr::expand_grid(id = 1:k, nc = n, nt = n, b1 = b1)

sim$age <- rep(x1, each = 2)
sim$age0 <- rep(x10, each = 2)

deltai <- rnorm(k, 0, sqrt(tau2r))
sim$deltai <- rep(deltai, each = 2)

sim$es <- b0 + sim$deltai + sim$b1*sim$age0

sim <- sim_studies(sim$es, sim$nc, sim$nt, data = sim)
```

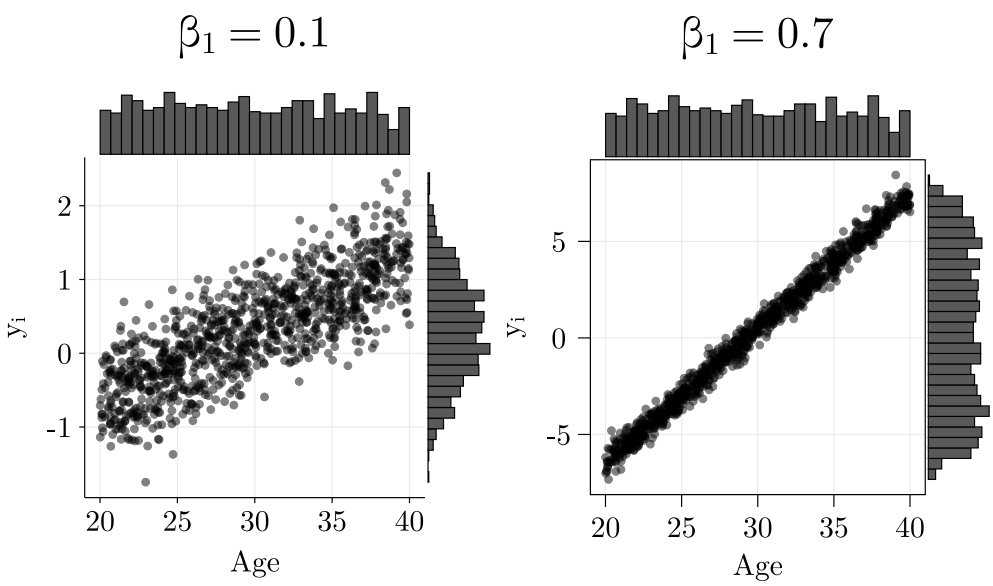

Figure 6. Scatter plots with marginal histograms for the range of simulated y_i with two β_1 values. The x-axis depicts the average age of simulated studies and the y-axis depicts the simulated effect size. On the left, the majority of simulated values range between -1.5 and 1.5 thus $\beta_1 = 0.1$ can be considered a plausible value. On the right, ($\beta = 0.7$) values range between -10 to 10 that despite theoretically possible can be considered highly implausible values in real meta-analyses for psychological data.

2.5.2.2 Simulating using R^2.

A more intuitive way to simulate a continuous predictor is fixing the desired R^2 value and finding the coefficient that produces the desired value. This approach has been implemented by Lopez-Lopez and colleagues (2014). We can use Equation (14) and (15) to find the β_1 value that is associated with a certain R^2.

$$\beta_1^2 = \tau^2 R^2 \tag{14}$$

$$\tau_r^2 = \tau^2 - \beta_1^2 \tag{15}$$

Now we can simulate the regression model using $\sqrt{\beta_1^2}$ as coefficient and τ_r^2 as residual

heterogeneity. Results from the fitted model fixing the R^2 values are presented in Table 7 and Figure 7. As Lopez-Lopez and colleagues (2014) demonstrated, to reliably estimate R^2 the number of studies needs to be large[11]. Lopez-Lopez and colleagues (2014) generated the moderator (X_1) values from a standard normal distribution. In the following example, we standardized the moderator (`scale()`) after simulating values on the *age* scale (e.g., `runif(k, 20, 40)`).

```
set.seed(seed)
k <- 100 # the number of studies
r2 <- 0.2 # the desired r2 value
tau2 <- 0.3 # the overall tau2
b0 <- 0.3 # the intercept i.e., average yi when x1 is 0
b1_2 <- tau2 * r2 # the beta1^2 i.e., the increase in yi for an increase in 1 year
b1 <- sqrt(b1_2) # b1_2 is squared, back to the original scale
tau2r <- tau2 - b1_2 # the residual heterogeneity after including x1
n <- 30 # number of participants per study, per group
x1 <- runif(k, 20, 40) # random mean-age for each study

sim <- make_data(k = k, nt = n, nc = n, age = x1)

sim$deltai <- rnorm(k, 0, sqrt(tau2r))
sim$age0 <- scale(sim$age, center = TRUE, scale = TRUE) # standardize the moderator
sim$es <- b0 + sim$deltai + b1*sim$age0

sim <- sim_studies(sim$es, sim$nt, sim$nc, data = sim)
res <- rma(yi, vi, mods = ~age0, method = "REML", data = sim)
```

[11] see also https://www.metafor-project.org/doku.php/tips:ci_for_r2 for a discussion about confidence intervals for the R^2 statistic

Table 7
Summary of the random-effects *model fixing the* R^2 *value. The* `intercept` *is the effect size for the average* `age` *(given that age is mean-centered). The* `age0` *parameter is the slope between* `age` *and the effect size interpreted as an increase in effect size for a unit increase in the average age. Other parameters are the same as described in Table 4 and 6.*

	β	95% CI	z	p
intercept	0.339 (SE = 0.051)	[0.239, 0.439]	6.643	< 0.001
age0	0.168 (SE = 0.051)	[0.067, 0.268]	3.272	0.001

$k = 100$
$\tau^2_r = 0.197$ $(SE = 0.037)$
$R^2 = 11.029\%$
$I^2 = 76.611\%$

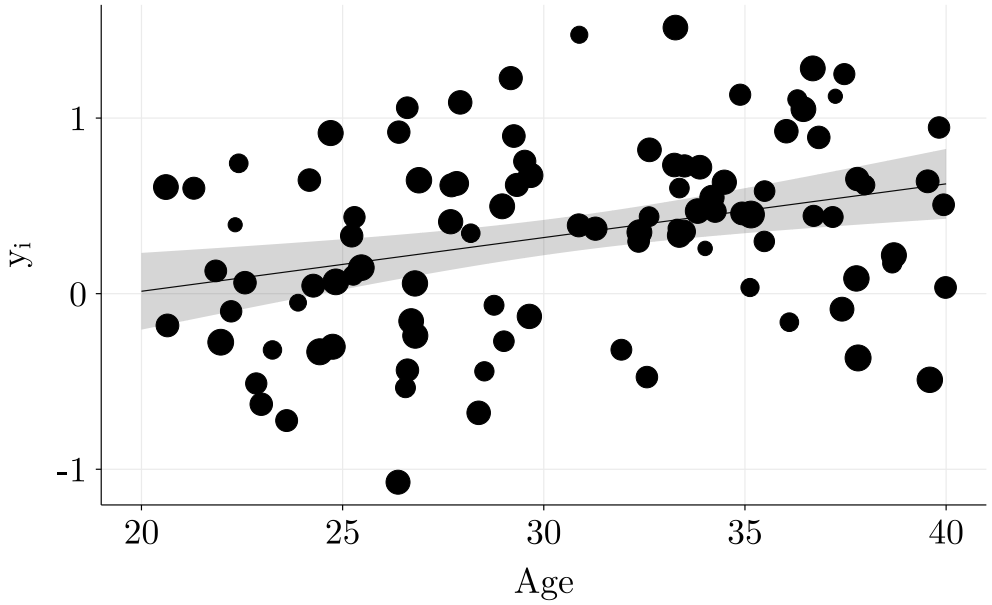

Figure 7. Meta-regression results for the *random-effects model* with a numerical moderator. Each effect size is represented with a black dot where the dimension represents the weight according to the inverse of the variance. The line represents the estimated meta-regression slope with the 95% confidence interval (grey bands).

3 Power Analysis

The previous simulation examples can be easily implemented for multiple purposes. For example, we can use different effect sizes and variance estimators when using the `sim_study()` function to check the impact on the fitted meta-analysis model. However, one of the most critical applications is estimating the power of a specific statistical model.

One of the purposes of power analysis by simulations is to estimate the required number of studies to detect an hypothetical effect size when planning a meta-analysis. At the same time, the meta-analysis that we presented (*two-level* equal or random-effects model) can be considered as a multi-lab (e.g., Klein et al., 2018) study where experiments are planned and not collected from the literature. We can use the same simulation approach to optimize the number of participants and studies when planning a multi-lab project.

As explained in the introduction, there are several approaches and tools to estimate the power of *equal* and *random-effects models* (see Borenstein et al., 2009, Chapter 29; Harrer et al., 2019, Chapter 14). These methods are easy to implement but made strong assumptions, such as the homogeneity of sample size, and did not consider the uncertainty in estimating τ^2. Jackson and colleagues (2017) partially solve the issue by developing an interesting method that takes into account the uncertainty in estimating τ^2 without using simulations. However, complex simulation scenarios can be handled only using Monte Carlo methods.

A general Monte Carlo simulation for the power analysis can be implemented with the following steps:

1. Choose the model that generates the data (e.g., equal or random-effects model).
2. Fix the relevant parameters (e.g., τ^2 and θ).
3. Simulate a dataset.
4. Fit the appropriate model.
5. Store the p-value associated with the parameter of interest
6. Repeat 3-5 a large number of times (e.g., 10000).
7. Calculate the power as the proportion of p-values below the α level.

For example, we can estimate the power of a *random-effects* model by repeating the simulation presented in Section 2.4 many times. We simulated heterogeneity of sample sizes sampling n_T and n_C values from a Poisson distribution with $\lambda = 20$. In this way, on

average, the sample size is 20 for primary studies with a certain amount of heterogeneity. Usually, it is more informative to simulate different scenarios according to the relevant parameters, such as sample sizes, number of studies, or heterogeneity. For example, we can estimate the power with a different number of studies k. We define the `do_sim()` function that, according to the input parameter, repeats the simulation a certain number of times (i.e., `nsim`)[12]. Increasing the number of simulations will increase the power analysis estimation precision. Then the `summary_sim()` function analyzes each simulation returning the relevant values. We repeat the simulation of the *random-effects* model several times with different parameters.

```
do_sim <- function(k, mu, tau2, navg, nmin, nsim, alpha = 0.05, summary = TRUE){
  # preallocate for computation speed
  p <- vector(mode = "numeric", length = nsim)

  # start the simulation loop
  for(i in 1:nsim){
    deltai <- rnorm(k, 0, sqrt(tau2))
    es <- mu + deltai
    # simulate sample size, it is possible to use other distributions e.g., Gaussian
    n <- nmin + rpois(k, navg - nmin)
    # simulate the studies
    sim <- sim_studies(es, n, n, data = NULL)
    res <- rma(yi, vi, method = "REML", data = sim)
    p[i] <- res$pval # store the p value
  }
  if(summary){
    # return directly the power
    summary_sim(p, alpha)
  }else{
    # return the list of pvalues
    data.frame(p)
  }
}
```

[12] We are using the `purrr::pmap()` function that can be considered very similar to `mapply()` but less verbose. Compared to `mapply()`, the `sim_grid` columns are directly passed to the `do_sim()` function without specifying the order.

```r
summary_sim <- function(p, alpha){
  power <- mean(p <= alpha) # compute power
  data.frame(power)
}
```

Simulation results are presented in Figure 8 and Table 8 showing that to reach 80% power (usually considered an appropriate level) with $\alpha = 0.05$ we need ~35 studies. The same approach could be used to estimate the power of a meta-regression by simply modifying the `do_sim()` function simulating the effect of a moderator and extracting the relevant p-value.

```r
set.seed(seed)
nsim <- 5000 # number of simulations per condition (5000, higher is better)
k <- c(5, 15, 25, 35, 50) # number of studies
delta <- 0.3 # the average effect size
tau2 <- 0.3 # the heterogeneity
navg <- 20 # average sample size per study
nmin <- 10 # minimum sample size per study
alpha <- 0.05 # the alpha level

# creating all combinations
sim_grid <- tidyr::expand_grid(k, mu = delta, tau2, navg, nmin, nsim)

# apply the simulation to all combinations
res <- purrr::pmap(sim_grid, do_sim)

# combine the results
res <- dplyr::bind_rows(res)
sim_grid <- cbind(sim_grid, res)
```

Table 8
The results from the power analysis simulation. The table depicts the simulation parameters, the estimated power using the `summary_sim()` *function, and the average sample size (n) across the simulations.*

k	Δ	τ^2	n_{avg}	n_{min}	nsim	Power
5	0.30	0.30	20	10	5000	0.26
15	0.30	0.30	20	10	5000	0.47
25	0.30	0.30	20	10	5000	0.65
35	0.30	0.30	20	10	5000	0.79
50	0.30	0.30	20	10	5000	0.92

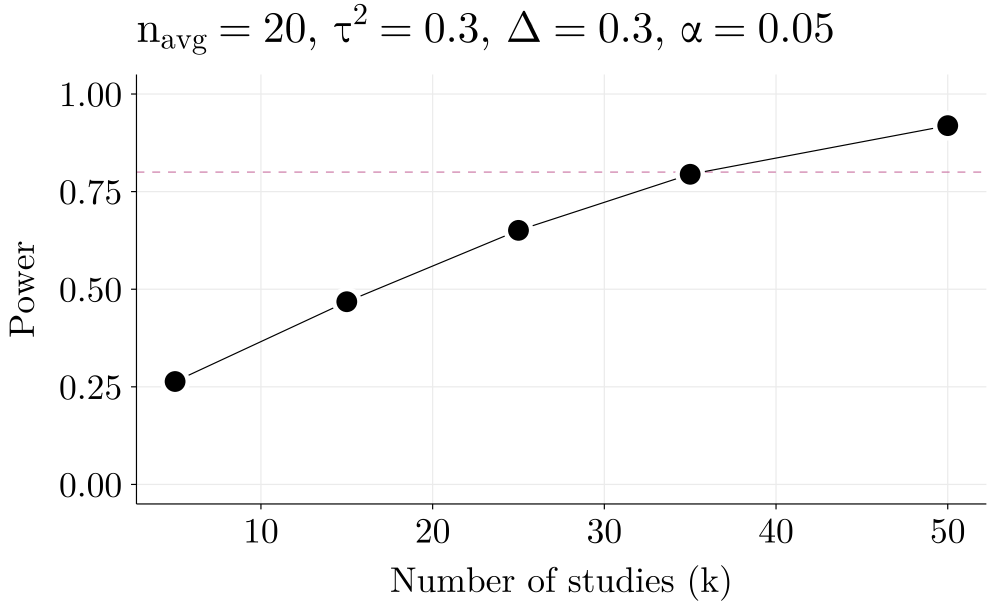

Figure 8. Results from the *random-effects model* power analysis. The x-axis depicts the number of studies (k), and the y-axis the estimated power. The pink dotted line is the 80% power level, usually considered a good value for power analysis.

4 Conclusions

The present work introduced the basic concepts of the meta-analysis regarding equal-effects, random-effects and meta-regression models with a simulation-based approach. We believe the presented examples are useful to implement alternative or more complex

models. For example, the `sim_study()` function can be easily modified to simulate another effect sizes index, such as correlations or odds ratios. In addition, more complex models, such as *multivariate* or *multilevel* models, can be simulated following a similar approach (see the supplementary materials). The *multilevel* (e.g., *three-level*) model estimates another heterogeneity component representing the variability of multiple independent effect sizes within the same study. Similarly, the *multivariate* model includes the correlation between multiple outcomes and the correlation between sampling errors.

The present work did have a few limitations. Firstly, we only introduced basic concepts about meta-analysis and Monte Carlo simulations, while setting up complex simulations requires more knowledge and complexity of the simulation setup. We decided to give the foundations to understand meta-analyses with a simulation approach because more complex models are still based on the same principles. Second, there are limitations concerning simulating participant-level data. We decided to simulate the meta-analysis data starting from the participants' level to maximize the flexibility and clearness of each step. The downside concerns the efficiency and scalability of the simulation setup. For large-scale simulations (e.g., many conditions, iterations, or complex models), simulating from aggregated statistics is probably more efficient (see Heuvel, Almalik, & Zhan, 2020 for an example) to improve the simulation efficency[13].

In conclusion, data simulation is a very powerful tool for each step of a data analysis process. Starting from the learning phase, where simulating data can be used to understand the statistical model in terms of assumptions and the data generation process, to the estimation of statistical power. Moreover, we believe that data simulation as part of a standard research workflow could improve the overall research quality. Data simulation requires understanding the statistical model, setting appropriate and reasoned parameters, and realizing how the chosen analysis method behaves across different scenarios.

---

[13] See https://www.jepusto.com/simulating-correlated-smds/ for a very clear example of the participant-level vs aggregated data simulation approaches

Acknowledgments. We thank the *R-sig-meta-analysis* mailing-list. Their suggestions and clarifications significantly improved the simulations approach and the R code.

Conflicts of Interest. The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.

Data, materials, and online resources. The code to reproduce simulations, figures and tables can be found on Open Science Framework (https://osf.io/54djn/) and Github (https://github.com/shared-research/simulating-meta-analysis).

Supplemental Material. The supplementary materials can be also found on the Open Science Framework repository (https://osf.io/54djn/) and Github (https://github.com/shared-research/simulating-meta-analysis).

Prior versions. The manuscript preprint has been uploaded on PsyArXiv https://psyarxiv.com/br6vy/

References

Aust, F., & Barth, M. (2022). *Papaja: Prepare american psychological association journal articles with r markdown*. Retrieved from https://github.com/crsh/papaja

Berkhout, S. W., Haaf, J. M., Gronau, Q. F., Heck, D. W., & Wagenmakers, E.-J. (2023). A tutorial on bayesian model-averaged meta-analysis in JASP. *Behav. Res. Methods*. https://doi.org/10.3758/s13428-023-02093-6

Blázquez-Rincón, D., Sánchez-Meca, J., Botella, J., & Suero, M. (2023). Heterogeneity estimation in meta-analysis of standardized mean differences when the distribution of random effects departs from normal: A monte carlo simulation study. *BMC Med. Res. Methodol.*, *23*(1), 19. https://doi.org/10.1186/s12874-022-01809-0

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. https://doi.org/10.1002/9780470743386

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119

Gelman, A., & Hill, J. (2006). *Data analysis using regression and Multilevel/Hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and other stories*. Cambridge University Press. https://doi.org/10.1017/9781139161879

Gentle, J. E. (2009). Monte carlo methods for statistical inference. In J. E. Gentle (Ed.), *Computational statistics* (pp. 417–433). New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-98144-4/_11

Gronau, Q. F., Heck, D. W., Berkhout, S. W., Haaf, J. M., & Wagenmakers, E.-J. (2021). A primer on bayesian model-averaged meta-analysis. *Adv. Methods Pract. Psychol. Sci.*, *4*(3), 251524592110312. https://doi.org/10.1177/25152459211031256

Harrer, M., Cuijpers, P., & Ebert, D. (2019). *Doing Meta-Analysis in R*. https://doi.org/10.5281/zenodo.2551803

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis with r: A Hands-On guide*. CRC Press.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *J. Educ. Behav. Stat.*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V. (1989). An unbiased correction for sampling error in validity generalization studies. *J. Appl. Psychol.*, *74*(3), 469–477. https://doi.org/10.1037/0021-9010.74.3.469

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

Heuvel, E. R. van den, Almalik, O., & Zhan, Z. (2020). *Simulation models for aggregated data meta-analysis: Evaluation of pooling effect sizes and publication biases*. Retrieved from https://arxiv.org/abs/2009.06305

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Ingalls, R. G. (2011). Introduction to simulation. *Proceedings of the 2011 Winter Simulation Conference (WSC)*, 1374–1388. ieeexplore.ieee.org. https://doi.org/10.1109/WSC.2011.6147858

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Res. Synth. Methods*, *8*(3), 290–302. https://doi.org/10.1002/jrsm.1240

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr, Alper, S., … Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Adv. Methods Pract. Psychol. Sci.*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Laird, N. M., & Mosteller, F. (1990). Some statistical methods for combining experimental results. *Int. J. Technol. Assess. Health Care*, *6*(1), 5–30. https://doi.org/10.1017/s0266462300008916

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front. Psychol.*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. *Applied Social Research Methods Series; Vol 49.*, *247*.

López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *Br. J. Math. Stat. Psychol.*, *67*(1), 30–48. https://doi.org/10.1111/bmsp.12002

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubio-Aparicio, M., López-López, J. A., Sánchez-Meca, J., Marín-Martínez, F., Viechtbauer, W., & Van den Noortgate, W. (2018). Estimation of an overall standardized mean difference in random-effects meta-analysis if the distribution of random effects departs from normal. *Res. Synth. Methods*, *9*(3), 489–503. https://doi.org/10.1002/jrsm.1312

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *J. Mem. Lang.*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schmid, C. H., Stijnen, T., & White, I. R. (2022). *Handbook of Meta-Analysis*. Taylor & Francis Limited.

Takkouche, B., Cadarso-Suárez, C., & Spiegelman, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Am. J. Epidemiol.*, *150*(2), 206–215. https://doi.org/10.1093/oxfordjournals.aje.a009981

Takkouche, B., Khudyakov, P., Costa-Bouzas, J., & Spiegelman, D. (2013). Confidence intervals for heterogeneity measures in meta-analysis. *Am. J. Epidemiol.*, *178*(6), 993–1004. https://doi.org/10.1093/aje/kwt060

Van Erp, S., Verhagen, J., Grasman, R. P. P. P., & Wagenmakers, E.-J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in *psychological bulletin* from 1990–2013. *Journal of Open Psychology Data*, *5*(1), 4. https://doi.org/10.5334/jopd.33

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Stat.*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2007). Approximate confidence intervals for standardized effect sizes in the two-independent and two-dependent samples design. *J. Educ. Behav. Stat.*, *32*(1), 39–60. https://doi.org/10.3102/1076998606298034

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & López-López, J. A. (2022). Location-scale models for meta-analysis. *Res. Synth. Methods*, *13*(6), 697–715. https://doi.org/10.1002/jrsm.1562

Wickham, H. (2023). *Tidyverse: Easily install and load the tidyverse*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., … Dunnington, D. (2023). *ggplot2: Create elegant data visualisations using the grammar of graphics*. Retrieved from https://CRAN.R-project.org/package=ggplot2

Zhu, H. (2021). *kableExtra: Construct complex table with kable and pipe syntax*. Retrieved from https://CRAN.R-project.org/package=kableExtra