Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze

Relazione Finale

Meta-analysis and meta-regression: application and discussion in case of small sample size

Relatore Prof.ssa Annamaria Guolo
Dipartimento di Scienze Statistiche

Laureando Enrico Roma
Matricola 1217336

Anno Accademico 2022/2023

Contents

0.1 Acronym List

OSF The Open Science Framework 37

RT Resistance Training 39

CE Common Effect 3

RE Random Effect 3

REML Restricted Maximum Likelihood iii

Introduction

This paper addresses the topic of meta-analysis, a statistical tool often used in systematic literature reviews to obtain a numerical estimate of the effect of interest. In addition to a general introduction to the statistical methodology, we have chosen to focus on one particular aspect, namely the use of meta-analysis in the context of small samples. Because the observations in meta-analysis are summary statistics obtained from previous analyses, small samples can be understood as a small number of studies with a sufficient number of participants in each study, a sufficient number of studies with a small number of participants in each of them, or both conditions simultaneously. Unfortunately, there are no available statistical methods for the last case.

The first chapter presents the theoretical foundations of this fascinating statistical technique. Classical models such as the Common Effect (CE) and the Random Effect (RE) models are discussed. In addition, their main extensions to include covariates, such as meta-regression models and mixed effects models, are examined. The key assumptions underlying these models are emphasized, as some of them are very problematic in the context of small samples. In addition, advanced methods that can be used to obtain reliable estimates in the context of small numbers of studies are presented. These methods include the use of maximum likelihood theory, the application of second-order asymptotic theory, and some nonparametric techniques such as permutation testing and bootstrapping. The final part of the chapter presents the basic and best-known methods for empirically evaluating the assumptions underlying the model and assessing the presence of publication bias.

Chapter 2 presents a practical application of the methods discussed in Chapter 1 using an example data set from 2021 published by Jones et collegues (Jones et al., 2021). This application focuses on sex differences in adaptation to resistance training. The first part of the chapter describes the methodology of the systematic literature review and explains the procedures used by the authors to derive the effect measures. The second

part presents the results of six different analyses in which the effect measure expresses the difference between males and females in post-pre changes following a resistance training program. The authors focused on upper limb muscle strength, lower limb muscle strength, and muscle size. This division was suggested primarily by the number of studies reporting each specific effect measure. For each "category", results were analyzed using the raw difference score. In addition, to reduce the effect of greater males body size, the effect measure was normalized to its baseline value before the administration of the training protocol.

Chapter 3 focuses on the integrated likelihood. A general introduction to the method and a comprehensive overview of its implementation in the context of meta-analysis is followed by a practical application using the same data set as in Chapter 2. This method proves particularly useful when working with a meta-analysis where the number of studies is sufficient but the sample size within each study is small. In fact, the integrated likelihood eliminates any nuisance parameter through integration, but accounting for the uncertainty linked to the eliminated parameters. It is also computationally convenient and numerically more stable because the elimination is done by averaging rather than maximization.

Overall, this thesis provides an exploration of meta-analysis, including its theoretical foundations, advanced techniques for small sample sizes, and empirical evaluation of assumptions and biases. Chapter 4 serves to summarize the findings and compare the results obtained with the selected dataset to the original study. A discipline-specific interpretation of the results is also provided.

Chapter 1

Introduction to meta-analysis

1.1 A brief introduction to meta-analysis through its history

Meta-analysis is a quantitative method of summarizing information from different sources, usually presented in aggregate form. In England, for example, it was used to model local authority data on tonsillectomy rates in children under 15 years of age from 2009 to 2011 (Suleman et al., 2010; Senn, 2016). However, it is much more commonly used in research practice, particularly in the social sciences, psychology, and medicine. According to historical research, the roots of meta-analysis can be traced to the 18th and 19th centuries, when Laplace and Gauss began to distinguish between observations within a given study and summary results between different studies (O'rourke, 2007). However, the first statistical approach to the subject is attributed to Pearson in 1904, who was asked to compare infection and mortality rates among soldiers who did or did not volunteer for typhoid vaccination in different regions of the British Empire (Pearson, 1904). In the original manuscript, Pearson provided two tables in which he calculated the correlation coefficient between vaccination and mortality or immunity as a measure of effect, as well as an estimate of within-study uncertainty. The pooled estimate presented was then the average of the within-study correlation coefficient, but was presented only as a point estimate. One could argue that the first real application concerns Josef Goldberg (Winkelstein, 1998) in 1907, who addressed the risk of urinary tract infection in typhoid fever (Goldberger, 1907). The dispute is based on the definition to be given to the term meta-analysis. In fact, Goldberg was the first to use four criteria that characterize modern meta-analysis. First, he conducted a comprehensive review of the literature to

identify 44 relevant studies. Then he selected the studies to be included in his work based on a specific criterion: the use of a newly developed technique called the serum agglutination test. In the third step, he summarized the information from the selected papers in a table and then performed the statistical analysis, calculating the average bacteriuria rate. It cannot be overlooked that the criteria for selecting the studies to be included in the final synthesis had a major impact on reducing what is now called heterogeneity. Indeed, the 44 selected studies showed a large variability, with infection rates ranging from 0 to 1 (with the extreme values included). One might suspect that the type of test used in the original investigation is a moderator variable. In their historical review, Chalmers, Hedges, and Cooper suggested that Goldberg's work highlights the need to distinguish two different methodological challenges in research synthesis, namely, taking measures to reduce bias and considering whether meta-analysis can be used to reduce statistical imprecision (Chalmers et al., 2002). This historical debate inevitably leads us to clarify some definitions that we will use from here on. According to the Dictionary of Epidemiology, meta-analysis is defined as "a statistical analysis of results from separate studies, examining sources of differences in results among studies, and leading to a quantitative summary of the results if the results are judged sufficiently similar or consistent to support such synthesis"(Porta, 2014). It should not be confused with systematic review, which is "a review of the scientific evidence which applies strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on the specific topic..." (Porta, 2014). In essence, we will use the term meta-analysis to identify a set of statistical procedures, whereas systematic review might involve meta-analysis but is not a statistical technique. The contribution of the eugenicist Ronald Fisher is more than notable as we continue our story. He first suggested that "although few or none [studies] can be claimed individually as significant, yet the aggregate [estimate] gives an impression that the probabilities are on the whole lower than would often have been obtained by chance"(Fisher, 1932, page 99) and then developed a technique for combining the p-value of independent tests of the same hypothesis (Fisher, 1932). He also worked in agriculture and reported an analysis of multiple trials that showed that the effect of fertilizers varied by year and location (Fisher, 1935). In the history of meta-analysis, there are at least a few other events that should be mentioned. In this thesis we will mainly deal with two types of models, the CE and the RE models. Both will be formally introduced in the following sections, asking the reader for an act of patience to finish the historical narrative. The latter was first proposed by Cochran in 1937 (Cochran, 1937) and its first application dates from 1938 (Yates & Cochran, 1938). Cochran developed this model to combine individual

experiments with different degrees of precision, relying on Fisher's work on experiments with the same degree of precision (Fisher, 1935). As a reminder, Fisher's work was in turn based on Airy's book (Airy, 1861). As we shall see, the RE model became popular with the publication of an approximation method for the Cochran model by Der Simonian and Laird (DerSimonian & Laird, 1986). The first application of meta-analysis in medicine dates from 1980 and was used to evaluate the effect of aspirin on the risk of recurrence of myocardial infarction after the first (Elwood, 2006). This application is particularly well known because it demonstrates the efficacy of a drug, although all of the randomized controlled trials included did not yield statistically significant results. Peto was the statistician responsible for the analysis and a few years later proposed using the (fixed) weighted mean to pool studies when the treatment effect varied (Peto, 1987), taking a position in favor of the CE model. Another milestone worth mentioning is the first article dealing with publication bias, i.e., the decision to publish or reject an article based on the direction of the results. in 1959, Sterling noted that psychological journals had an offbeat "publication policy" that led to a strange dichotomy:

- "Experimental results will be printed with a greater probability if the relevant test of significance rejects H_0 for the major hypothesis with $Pr(E|H_0) \leq .05$ than if they fail to reject H_0 at that level" or
- "The probability that an experimental design will be replicated becomes very small once such an experiment appears in print" (Sterling, 1959).

To the end of the paragraph, Gene Glass must be mentioned as the person who first proposed the term meta-analysis as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976). However, to be unbiased and to give voice to the critique, it is right to recall Feinstein who coined one of the favorite terms of the editor of this paper, "statistical alchemy for the 21st century" (Feinstein, 1995).

1.2 Set-up of meta-analysis

In modern meta-analyses, a collection of studies is analyzed, and the analyst work with a data set that includes at least one effect measure and one uncertainty measure for each study. The approach assumes that each study, indexed by $i = 1, \ldots, K$, provides an estimate y_i of the effect measure and a standard error s_i representing the uncertainty associated with the estimate. The effect measure y_i estimates the true effect measure

in the i-th study, denoted β_i. Note that pairwise univariate meta-analysis allows only one effect measure per study, although there are extensions that allow the introduction of more than one effect measure per study, provided some assumptions are made about the within-trial variance matrix. In addition, the included studies are considered independent. Sometimes this assumption may seem an exaggeration, especially if the meta-analysis includes studies from the same research team or if, for example, a correlation is suspected between effect sizes collected in the same nation.

1.3 The within-studies model

The first step in pairwise meta-analysis is the definition of a within-study model. Typically, we have

$$Y_i \sim \mathcal{N}(\beta_i, s_i^2). \tag{1.1}$$

With this model we more or less explicitly assume that (Jackson & White, 2018):

1. Y_i is an unbiased estimate of β_i;
2. the variance of the within-study model is known and equal to s_i^2;
3. the shape of the within-study model distribution is normal.

However, we are in the realm of approximations. Even in the most idyllic scenario, in which the meta-analysis we are working on is not subject to publication bias or other types of bias, it is possible that the effect measure under consideration is subject to transformation bias. For example, the logit transformation of an unbiased incidence measure would be biased because it is a nonlinear transformation of an unbiased quantity (Jensen's inequality). To some extent, these considerations invalidate assumption number 1. Moreover, assumption number 2 implies that the standard methodology ignores the uncertainty of the within-study variances. The approximation is based on the central limit and the Slutsky theorem (Jackson, 2009), and it is generally accepted if the sample of the i-th study is large enough. However, it should be noted that the within-study variance formula is often only an approximation, depending on the effect measure considered. In addition, assumption 2 hides a second assumption, namely that y_i and s_i^2 are uncorrelated. This can be a problem, consider for example the case where $Y_i = \log(\frac{A_i/B_i}{C_i/D_i})$ and $s_i^2 = \frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C_i} + \frac{1}{D_i}$, where a particularly small value of B_i implies $\mathrm{Cor}(Y_i, S_i^2) > 0$. Another notable example concerns meta-analyses in which the definition of the population of interest is particularly broad and subgroups with varying

degrees of treatment efficacy can be identified. In this case, if the authors of the original studies perform a sample size calculation, it is likely that in studies where higher treatment efficacy is suspected, the sample size will be smaller. Since s_i^2 is an inverse function with respect to sample size, any larger y_i would be associated with larger s_i^2 and vice versa, effectively creating a correlation. Assumption number 3 is never technically appropriate. Partly because it depends on the nature of the outcome. For example, a variety of alternative models have been proposed for event outcome (Stijnen et al., 2010). But in general, $Y_i \mid \mu_i \sim \mathcal{N}(\mu_i, s_i^2)$ is false if we assume $Y_i \mid \mu_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, where σ_i^2 is the true variance of the distribution. In fact, the pivotal distribution for the confidence interval on μ_i should be a Student's t (Jackson & White, 2018).

1.4 The common-effect model

Once a within-studies model has been assumed, a between studies model must be introduced, so that a summary effect measure can be provided. The most well known between study model is based on the assumption of a constant (or common) general expected value, that is $\beta_i = \beta,\ \forall i = 1, \ldots, K$. This assumption gives the CE model

$$\begin{aligned} Y_i &\sim \mathcal{N}(\beta, s_i^2) \\ \mathrm{Cov}\,(Y_i, Y_j) &= 0, \quad \forall i, j = 1, ..., K, \end{aligned} \tag{1.2}$$

or equivalently,

$$\begin{aligned} Y_i &= \beta + \epsilon_i \quad \text{with} \quad \epsilon_i \sim \mathcal{N}(0, s_i^2) \\ \mathrm{Cov}\,(\epsilon_i, \epsilon_j) &= 0 \quad \forall i, j = 1, ..., K. \end{aligned} \tag{1.3}$$

1.4.1 Estimation

The only parameter that needs to be estimated in model (1.2) is β. The weighted least squares method can be used for this purpose, and the resulting estimator is equivalent to the maximum likelihood estimator. It follows:

$$\hat{\beta}_{CE} = \frac{\sum_{i=1}^{K} Y_i \cdot w_{i,CE}}{\sum_{i=1}^{K} w_{i,CE}}, \tag{1.4}$$

then

$$w_{i,CE} = \frac{1}{\sqrt{\sum_{i=1}^{K} s_i^2}}. \tag{1.5}$$

The standard error of the estimator $\hat{\beta}_{CE}$ is equal to

$$se(\hat{\beta}_{CE}) = \frac{1}{\sqrt{\sum_{i=1}^{K} w_{i,CE}}}, \tag{1.6}$$

and its asymptotic distribution for $K \to \infty$ is

$$\hat{\beta}_{CE} \sim \mathcal{N}(\beta, se(\hat{\beta}_{CE})). \tag{1.7}$$

The distributional assumption can be justified as a linear combination of normal random variables with constant coefficients. However, this assumption is primarily made for convenience, as the weights in the equation (1.5) are known only because s_i^2 is assumed to be known (Jackson & White, 2018). Additionally, as discussed in section 1.3, the weighted least squares estimation may not be unbiased if y_i and s_i^2 are correlated (Jackson & White, 2018). From a different perspective, the justification for normality is based on the first-order asymptotic theory using the likelihood approach.

1.4.2 Confidence intervals and hypothesis test

Using the asymptotic result presented in equation (1.7), a Wald-type confidence interval of level α can be derived as follows

$$IC_{\text{Wald}}(\alpha) : \left\{ \hat{\beta}_{CE} - z_{1-\alpha/2} \cdot se(\hat{\beta}_{CE}), \hat{\beta}_{CE} + z_{1-\alpha/2} \cdot se(\hat{\beta}_{CE}) \right\},$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha)/2$ quantile of a standard normal distribution. Moreover, given a hypotheses set $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$, the statistic $Z_{CE} = \hat{\beta}_{CE}/se(\hat{\beta}_{CE}) \sim \mathcal{N}(0,1)$, hence the null hypothesis will be rejected for $\left|Z_{CE}^{obs}\right| \geq Z_{1-\alpha/2}$, where Z_{CE}^{obs} is the observed value of the statistic.

1.4.3 Heterogeneity

In the CE model, the term heterogeneity is used when the effect measures included in the meta-analysis are not all estimates of a common parameter. More formally, the interest is to evaluate the null hypothesis $H_0 : \beta_i = \beta, \quad \forall i = 1, \ldots, K$, which represents the main assumption of the CE model. The Cochran's Q-statistic could be used for the purpose (Cochran, 1937, 1954) and it is defined as

$$Q = \sum_{i=1}^{K} w_{i,CE}(Y_i - \hat{\beta}_{CE})^2 \overset{H_0}{\sim} \chi^2_{K-1}. \tag{1.8}$$

However, the distributional results only holds approximately when $s_i^2, \quad \forall i = 1, \ldots, K$ are assumed to be known. It should be noted that the test has a low power.

1.4.4 Criticisms and limitations of the model

The main criticism against the use of the CE model concerns the assumption of homogeneity of effects between the different studies. If this were not checked, the estimator would be rather a weighted average with weights proportional to the inverse of the study variance, which turns out to be a function of sample size and could have consequences depending on how the sample size was determined in the original studies. Indeed, it can be chosen arbitrarily or depend on the effect expected by the researcher (and therefore determined by an appropriate calculation). In the first case, it would no longer be possible to interpret the weights and therefore the estimator. In the second case, the correlation between y_i and the weights implies that the weighted least squares estimator would be biased. The second and main criticism against the use of the CE model concerns the validity of the within-study model (1.1). Even assuming that the within-study distribution is normal, the model holds as long as the s_i are good approximations of the true standard errors and, in turn, this is guaranteed somehow as long as the sample size of the study is sufficiently large. Finally, it should be noted that all the inference procedures for the model (1.2) hold accurately only if the number of study K is large enough.

1.5 The random-effect model

Although historically the RE model has seen an independent development from the CE model (Cochran, 1937), to date it is proposed as the main alternative if heterogeneity is suspected among the included effect measures. In fact, the model introduces an additional parameter, namely τ^2, with the aim of quantifying the amount of between studies heterogeneity. In this section, the model will be presented according to the standard formulation proposed by (DerSimonian & Laird, 1986). Particularly, the RE model is defined according to the following two-stage sampling procedure

$$\begin{aligned} \beta_i &\sim \mathcal{N}(\beta, \tau^2), \\ Y_i|\beta_i &\sim \mathcal{N}(\beta_i, s_i^2), \\ \text{Cov}\,(Y_i, Y_j) &= 0, \quad \forall i, j = 1, ..., K. \end{aligned} \tag{1.9}$$

Hence, marginally it holds

$$Y_i \sim \mathcal{N}(\beta, \tau^2 + s_i^2). \tag{1.10}$$

In fact, denoting the probability density function with $p(\cdot)$, the previous result follows from the marginalization on β,

$$p(y_i) = \int_{\beta_i \in S_\beta} p(y_i|\beta_i)p(\beta_i)d\beta_i,$$

where S_β stands for the support of β.
In addition, as a result of the Normal distribution being a scale-location family, it results equivalently:

$$\begin{aligned} Y_i &= \beta + \delta_i + \epsilon_i, \\ &\text{with } \delta_i \sim \mathcal{N}(0, \tau^2), \epsilon \sim \mathcal{N}(0, s_i^2); \\ \text{Cov}\,(\delta_i, \delta_j) &= 0,\ Cov(\epsilon_i, \epsilon_j) = 0. \quad \forall i \neq j, i, j = 1, ..., K, \\ \text{Cov}\,(\delta_i, \epsilon_j) &= 0 \quad \forall i, j = 1, ..., K. \end{aligned} \tag{1.11}$$

Note that if $\tau^2 = 0$, the model reduces to the CE model. However, the main difference from the CE model is that the true effect of the single study is not constant among all studies but it is a realization of a normal random variable centered in β. Therefore, the objective is to estimate the center of the effect distribution, represented by β, allowing an intrinsic variability of the phenomenon, described by τ^2 and named heterogeneity.

1.5.1 Estimation

In the RE model the unknown to-be-estimated parameter is the bidimensional

$$(\beta, \tau)^T \in \mathbb{R} \times \mathbb{R}^+.$$

A review of the most commonly used methods to estimate the parameter of interest will be proposed next. Note that generally the component of main interest is β and τ^2 represents a nuisance parameter. However, it is undeniable that τ^2 plays an important role in the interpretation of meta-analysis results and especially in the process of technology transfer, where prediction intervals are of primary importance. Almost all estimation methods proceed with separate components, first estimating $\hat{\tau^2}$ and only then $\hat{\beta}_{RE}$ using $\hat{\tau^2}$ to compute the weights.

1.5.1.1 Standard approach

The best-known approach to estimate the two-dimensional $(\beta, \tau^2)^T$ parameter is a two-step procedure in which an estimator for τ^2 is first obtained using the method of moments and then an estimator for β is derived, assuming τ^2 to be known and by plugging-in the estimated $\hat{\tau}^2$ into the RE weights. To obtain the τ^2 estimator, (DerSimonian & Laird, 1986) began rewriting Cochran's Q-statistic (1.8) to apply the method of moments. Specifically,

$$\begin{aligned} Q_{CE} &= \sum_{i=1}^{K} w_{i,CE}(Y_i - \hat{\beta}_{CE})^2 \\ &= \sum_{i=1}^{K} w_{i,CE}\left\{(Y_i - \beta) - (\hat{\beta}_{CE} - \beta)\right\}^2 \\ &= \sum_{i=1}^{K} w_{i,CE}(Y_i - \beta)^2 - (\hat{\beta}_{CE} - \beta)^2 \sum_{i=1}^{K} w_{i,CE}. \end{aligned} \tag{1.12}$$

The last equivalence can be obtained simply by developing the square of the binomial and using the fact that

$$\hat{\beta}_{CE} = \frac{\sum_{i=1}^{K} Y_i w_{i,CE}}{\sum_{i=1}^{K} w_{i,CE}} \iff \sum_{i=1}^{K} w_{i,CE} Y_i = \hat{\beta}_{CE} \sum_{i=1}^{K} w_{i,CE}.$$

The expected value of Q_{CE} with respect to Y_i under the RE model is

$$\begin{aligned} E\{Q_{CE}\} = & \sum_{i=1}^{K} w_{i,CE} E\{(Y_i - \beta)^2\} - E\{(\hat{\beta}_{CE} - \beta)^2\} \sum_{i=1}^{K} w_{i,CE} = \\ = & \sum_{i=1}^{K} w_{i,CE} \operatorname{Var}(Y_i) - Var\left(\hat{\beta}_{CE}\right) \sum_{i=1}^{K} w_{i,CE} = \\ = & \sum_{i=1}^{K} w_{i,CE}(w_{i,CE}^{-1} + \tau^2) - \frac{\sum_{i=1}^{K} w_{i,CE}^2 (w_{i,CE}^{-1} + \tau^2)}{(\sum_{i=1}^{K} w_{i,CE})^2} \sum_{i=1}^{K} w_{i,CE} = \\ = & K - 1 + \left\{ \sum_{i=1}^{K} w_{i,CE} - \frac{\sum_{i=1}^{K} w_{i,CE})^2}{\sum_{i=1}^{K} w_{i,CE})} \right\} \tau^2. \end{aligned} \tag{1.13}$$

Note that under the CE model the weights are given by $w_{i,CE} = \frac{1}{s_i^2}$, and that the variances are calculated under the RE model as

$$\operatorname{Var}(Y_i) = s_i^2 + \tau^2, \quad \operatorname{Var}\left(\hat{\beta}_{CE}\right) = \frac{\sum_{i=1}^{K} \operatorname{Var}(Y_i)\, w_{i,CE}^2}{(\sum_{i=1}^{K} w_{i,CE})^2}.$$

By equating the right hand side of the (1.13) with Q_{CE} in (1.8), and solving for τ^2, it results

$$\hat{\tau^2_{DL}} = \max\left\{0, \frac{\sum_{i=1}^{K} w_{i,CE}(Y_i - \hat{\beta}_{CE})^2 - (K-1)}{\sum_{i=1}^{K} w_{i,CE} - \sum_{i=1}^{K} w_{i,CE}^2 / \sum_{i=1}^{K} w_{i,CE}}\right\}. \tag{1.14}$$

The $\hat{\tau}^2_{DL}$ estimator is a special case of the general method of moments (Kacker, 2004). Considering a set of weights a_i with the following properties,

$$a_i \geq 0, \quad 0 \leq \frac{a_i}{\sum_{i=1}^{K} a_i} \leq 1;$$

the Q statistic can be defined as

$$Q_a = \sum_{i=1}^{K} a_i (Y_i - \hat{\beta}_a)^2,$$

where $\beta_a = \sum_{i=1}^{K} a_i Y_i / \sum_{i=1}^{K} a_i$. Assuming that the underlying generative model is the RE model, we have

$$\mathrm{E}\{Q_a\} = \sum_{i=1}^{K} a_i (s_i^2 + \tau^2) - \sum_{i=1}^{K} a_i^2 (s_i^2 + \tau^2) / \sum_{i=1}^{K} a_i, \tag{1.15}$$

hence by equating Q_a to its expectation in equation (1.15), the estimator of τ^2 becomes

$$\hat{\tau}_a^2 = \max\left\{0, \frac{\sum_{i=1}^K a_i(Y_i - \hat{\beta}_a)^2 - (\sum_{i=1}^K a_i s_i^2 - \sum_{i=1}^K a_i^2 s_i^2 / \sum_{i=1}^K a_i)}{\sum_{i=1}^K a_i - \sum_{i=1}^K a_i^2 / \sum_{i=1}^K a_i}\right\}. \tag{1.16}$$

The calculation follows those in (1.12) and (1.13). Interestingly, eight out of the sixteen methods reviewed by Veroniki et al. (2016) to estimate τ^2 fall under the framework presented in this paragraph. Moreover, it is straightforward to verify that the CE model is retrieved using $a_i = s_i^{-2}$.

One last interesting note concerns the Paule-Mandel estimator (Paule & Mandel, 1982). In fact, if $a_i = 1/(\tau^2 + s_i^2)$, which are the optimal but unknown weight, the (1.15) becomes:

$$\begin{aligned} \mathrm{E}\{Q_a\} &= \sum_{i=1}^K \frac{(s_i^2 + \tau^2)}{s_i^2 + \tau^2} - \sum_{i=1}^K \frac{(s_i^2 + \tau^2)}{(s_i^2 + \tau^2)^2} / \sum_{i=1}^K (s_i^2 + \tau^2)^{-1} \\ &= K - 1. \end{aligned} \tag{1.17}$$

An estimating equations can be derived as

$$Q(\tau^2) - (K-1) = \sum_{i=1}^K \left\{ \frac{Y_i - \beta(\hat{\tau^2})}{\tau^2 + s_i^2}^2 \right\} = 0, \tag{1.18}$$

and the latter can be solved in τ^2 via an iterative algorithm.

1.5.1.2 Maximum likelihood estimation

Alternatively, Hardy & Thompson (1996) proposed an estimator based on likelihood theory. It is particularly important to note that since s_i^2 are assumed to be known, the likelihood function depends only on the two-dimensional parameter $(\beta, \tau^2)^T$. To obtain the estimator, we consider the likelihood function of a multivariate normal distribution:

$$L(\beta, \tau^2 | y_1, \dots y_K) = \prod_{i=1}^K \frac{1}{\sqrt{2\pi(\tau^2 + s_i^2)}} e^{-\frac{(y_i - \beta)^2}{2(\tau^2 + s_i^2)}}. \tag{1.19}$$

The log-likelihood is defined as

$$\begin{aligned} \ell(\beta, \tau^2 | y_1, \dots y_K) &= \log L(\beta, \tau | y_1, \dots y_K) = \\ &\sum_{i=1}^K \frac{1}{2} \log 2\pi(\tau^2 + s_i^2) - \sum_{i=1}^K \frac{(y_i - \beta)^2}{2(\tau^2 + s_i^2)}. \end{aligned} \tag{1.20}$$

Calculating the partial derivatives with respect to the parameters, we obtain the score function

$$\begin{cases} \ell'(\beta) = \frac{\partial}{\partial\beta} l(\beta, \tau^2) = -\sum_{i=1}^{K} \frac{(y_i - \beta)}{(\tau^2 + s_i^2)}, \\ \ell'(\tau^2) = \frac{\partial}{\partial\tau^2} l(\beta, \tau^2) = \sum_{i=1}^{K} \frac{1}{2(\tau^2 + s_i^2)} + \sum_{i=1}^{K} \frac{(y_i - \beta_i)^2}{2(\tau^2 + s_i^2)^2}. \end{cases} \tag{1.21}$$

Therefore, equating the first component of (1.21) to zero, the estimator for β is given by

$$\hat{\beta}_{RE}(\tau^2) = \frac{\sum_{i=1}^{K} y_i \cdot w_{i,RE}}{\sum_{i=1}^{K} w_{i,RE}}, \tag{1.22}$$

where

$$w_{i,RE} = \frac{1}{\sqrt{\sum_{i=1}^{K} (s_i^2 + \tau^2)}}.$$

Note that the estimator $\hat{\beta}_{RE}(\tau^2)$ can be considered to be in a closed form only if $\tau^2 = \hat{\tau}^2$ is plugged-in. Conversely, it is not possible to obtain a closed form for the estimator of τ^2 and numerical procedure must be used (e.g. the Newton-Rapson algorithm). For the purpose, a convenient rearrangement is (Hardy & Thompson, 1996):

$$\hat{\tau^2} = \frac{\sum_{i=1}^{K} \frac{\{\hat{y}_i - \beta(\tau^2)\}^2 - s_i^2}{(s_i^2 + \hat{\tau}^2)^2}}{\sum_{i=1}^{K} \frac{1}{(s_i^2 + \hat{\tau}^2)^2}}. \tag{1.23}$$

The mean and the variance of a Normal distribution are well known to be orthogonal (i.e. the diagonal element of the Fisher information matrix are zeros). Hence, the standard error of the estimator $\beta(\hat{\tau^2})$ is obtained by the square root of the first diagonal element of the (observed) information matrix reciprocal. Let

$$-\frac{\partial^2}{\partial\beta^2} \ell(\beta) = \sum_{i=1}^{K} \frac{1}{s_i^2 + \tau^2}, \tag{1.24}$$

the standard error of the estimator $\beta(\tau^2)$ is equal to

$$se(\hat{\beta}_{RE}(\tau^2)) = \frac{1}{\sqrt{\sum_{i=1}^{K} w_{i,RE}}} \tag{1.25}$$

So, although the parameter of interest is β and an estimator in a closed form is available, it is mandatory to obtain at least one estimate of τ^2 for both, the calculation of the

point estimate and its standard error.

1.5.1.3 Profile maximum likelihood

Given a two-dimensional parameter $(\beta, \tau^2)^T$, where β represents the component of interest and τ is a nuisance parameter, the profile likelihood can be defined as (Murphy & Van der Vaart, 2000):

$$\ell_p(\beta) = \max_{\tau^2} \ell(\beta, \tau^2), \tag{1.26}$$

where $l(\beta, \tau^2)$ is the log likelihood function for the RE model, formula (1.20). If tau were the parameter of interest, the profile maximum likelihood estimator could be obtained as

$$\ell_p(\tau^2) = \max_{\beta} \ell(\beta, \tau^2). \tag{1.27}$$

However, the definition of τ^2 as the parameter of interest and β as the nuisance could be at least debatable, although it is used as a mere technique to obtain a confidence interval for heterogeneity. Unfortunately, a closed form in both the cases is not available and the optimization problem has to be solved iteratively (Hardy & Thompson, 1996).

1.5.1.4 Restricted maximum likelihood

The maximum likelihood estimate of variance components has been found to be negatively biased (Harville, 1977) and an unbiased estimate can be obtained with an approach based on the REML (Patterson & Thompson, 1971). However, it should be noted that the maximum likelihood estimate typically has lower mean squared error. In order to present the approach as generally as possible, it is preferred to introduce the vector notation typically used for mixed-effect models. In fact, the RE model can be considered a form of mixed effects model with the peculiarity that the $sigma_i$ are known.
Consider following mixed effect model:

$$\mathbf{Y} \sim \mathcal{N}_K(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}) \quad \text{where} \quad \boldsymbol{\Sigma} = \sum_{i=1}^{m} \sigma_i^2 \mathbf{Z}_i \mathbf{X}_i^T + \sigma^2 \mathbf{K}_n, \tag{1.28}$$

and $\mathbf{X}$ is a $K \times p$ matrix of rank $r \leq p$, and $\boldsymbol{\Sigma}$ is positive define a $K \times K$ matrix. The REML maximize the likelihood of $\mathbf{Ky}$ with $\mathbf{K}$ chosen so that $\mathbf{KX} = \mathbf{0}$. It follows that $\mathrm{E}[\mathbf{Ky}] = \boldsymbol{KX\beta} = \mathbf{0}$ and $\mathbf{K}$ is required to be full rank and with the maximal number of rows possible.

A full-rank matrix $\mathbf{K}$ with maximal number of rows such that $\boldsymbol{KX\beta} = \mathbf{0}$, is an $(K - r) \times K$. Furthermore, $\mathbf{K}$ must be of the form $\mathbf{K} = \mathbf{C}(\mathbf{K} - \mathbf{H})$ where $\mathbf{C}$ specifies a full-rank transformation of the rows of $\mathbf{K} - \mathbf{H}$ and $\mathbf{H} = \mathbf{X}\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}$. The proof can be found in (Rencher & Schaalje, 2008, page 487). Please note that there are an infinite $\mathbf{K}$ and the derived likelihood would be equivalent.
The method is well known as residual maximum likelihood. In fact, the vector $(\mathbf{K} - \mathbf{H})\mathbf{Y}$ is the vector of model residuals and $\mathbf{KY} = \mathbf{C}(\mathbf{K} - \mathbf{H})\mathbf{Y}$ one of its linear combination. The quantity $\mathbf{k_j X}$ with $\mathbf{k_j}$ elements of $\mathbf{K}$ are known as error contrasts (Harville, 1977). The distribution of $\mathbf{KY}$ is

$$\mathbf{KY} \sim \mathcal{N}_{K-r}\left(\mathbf{0}, \boldsymbol{K\Sigma K^T}\right). \tag{1.29}$$

$\mathbf{K}$ is a matrix of known constant (as it is a function of $\mathbf{X}$ which is known) and so (1.29) can be derived from the well-known property of the multivariate normal distribution.
The REML conists in the maximization of the likelihood of $\mathbf{Ky}$ which depends only from the variance component parameter. The classical likelihood method can be used to derive an estimator for each $m + 1$ variance component. Iterative technique could be necessary to solve the derived equation.
With respect to the RE model, the REML is equivalent to the marginal log-likelihood function for the residuals (Guolo, 2012) and can be written as

$$\ell_{REML}(\tau^2) = -\frac{1}{2}\sum_i^K \log \hat{\sigma^2} + \tau^2 - \frac{1}{2}\sum_i^K \frac{1}{\hat{\sigma^2} + \tau^2} - \frac{1}{2}\sum_i^K \frac{(y_i - \hat{\beta})^2}{\hat{\sigma^2} + \tau^2}. \tag{1.30}$$

giving the estimate for τ^2 of

$$\hat{\tau}^2_{REML} = \frac{\{(y_i - \hat{\beta})^2 - \hat{\sigma}_i^2\}/(\hat{\sigma_i}^2 + \tau^2)^2}{\sum_i^K 1/(\hat{\sigma^2} + \tau^2)^2} + \frac{1}{\sum_i^K 1/(\hat{\sigma_i}^2 + \tau^2)^2}, \tag{1.31}$$

often approximated as (Guolo, 2012)

$$\hat{\tau}^2_{REML} \approx \frac{\sum_i^K \{K(K-1)^{-1}(y_i - \hat{\beta})^2 - \hat{\sigma_i^2}\}/(\hat{\sigma_i^2} + \tau^2)^2}{\sum_i^K 1/(\hat{\sigma_i}^2 + \tau^2)^2}. \tag{1.32}$$

1.5.2 Confidence intervals and hypothesis tests on β

The rationale behind standard method for obtaining a interval estimate on β is built as follows:

$$\hat{\beta}_{RE}(\tau^2) \sim \mathcal{N}(\beta, se(\hat{\beta}_{RE}(\tau^2))), \tag{1.33}$$

where

$$se(\hat{\beta}(\tau^2)) = \left(\sum_{i=1}^{K} w_{i,RE}\right)^{-1/2}$$

and

$$w_{i,RE} = (\tau^2 + s_i^2)^{-1}.$$

Therefore, it is sufficient to plug-in an estimate of τ, for instance using the $\hat{\tau}^2_{DL}$ proposed in (1.14). Then, a Wald-type confidence interval of level α can be obtained as

$$IC_{\text{Wald}}(\alpha) : \left\{\hat{\beta}_{RE}(\hat{\tau}^2) - z_{1-\alpha/2} \cdot se(\hat{\beta}_{RE}(\hat{\tau}^2)), \hat{\beta}_{RE}(\hat{\tau}^2) + z_{1-\alpha/2} \cdot se(\hat{\beta}_{RE}(\hat{\tau}^2)\right\},$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha)/2$ quantile of a standard normal distribution. For a bilateral hypotheses test with an hypotheses test of the form $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$, the statistic $Z_{RE} = \hat{\beta}_{RE}(\hat{\tau}^2/se(\hat{\beta}_{RE}(\hat{\tau}^2)$ can be used. Under the null hypothesis, it follows a standard normal distribution. The null hypothesis is rejected at level α when $\left|Z^{obs}_{RE}\right| > Z_{1-\alpha/2}$, where Z^{obs}_{RE} is the observed value of the statistic.
However, in the RE model part of the error (the between-study heterogeneity) is not known and has to be estimated. Through the inferential procedure described above, the uncertainty behind the estimate of τ^2 is not taken into account. As a result, this additional source of error would potentially affect the coverage of the confidence intervals for β. Moreover, noticing that standard errors depend non-linearly on τ^2, even if the estimate of τ^2 is unbiased, this exact property cannot be translated to hold for the standard error themselves.
The profile likelihood (1.26) offers an immediate solution to the problem eliminating the nuisance parameter τ^2 via maximization. A confidence interval of level α can be obtained by solving

$$pl(\beta) > pl(\hat{\beta})) - \chi^2_{1,1-\alpha}, \tag{1.34}$$

while the hypothesis $H_0 : \beta = 0$ can be tested using the profile log-likelihood ratio test

$$-2(pl(0) - pl(\hat{\beta})) \sim \chi^2_1, \tag{1.35}$$

where 0 is the value of β under the null hypothesis. Obviously, researchers who wanted to proceed in this way, would have to give up making inferences about τ^2. A third method, named Hartung-Knapp-Sidik-Jonkman correction, could be also considered. Briefly, let

$$\begin{aligned} Z &= \frac{\hat{\beta} - \beta}{1/\sqrt{\sum_{i=1}^{K} w_i}} \sim \mathcal{N}(0,1), \\ Q &= \sum_{i=1}^{K} w_i (y_i - \hat{\beta})^2 \sim \chi^2_{K-1}, \\ Z &\perp Q, \text{then} \\ \frac{Z}{Q} &= \frac{\hat{\beta} - \beta}{\sqrt{\sum_{i=1}^{K} w_i (y_i - \hat{\beta})^2 / ((K-1)\sum_{i=1}^{K} w_i)}} \sim t_{K-1}. \end{aligned} \tag{1.36}$$

where w_i are the true weights using the true τ^2. At this point is sufficient to plug-in the estimated weights. Note that the denominator of the ratio Z/Q is an unbiased estimate of the standard error of $\hat{\beta}$ if the weights are known and that the confidence interval are wider using the t-distribution. Therefore, a confidence interval of level α can be obtained as

$$\begin{aligned} IC(\alpha) : \Big\{ &\hat{\beta}_{RE}(\hat{\tau}^2) - t_{K-1,1-\alpha/2} \cdot se_{HKSJ}(\hat{\beta}_{RE}(\hat{\tau}^2)), \\ &\hat{\beta}_{RE}(\hat{\tau}^2) + t_{K-1,1-\alpha/2} \cdot se_{HKSJ}(\hat{\beta}_{RE}(\hat{\tau}^2)) \Big\} \end{aligned}$$

where $se_{HKSJ}(\hat{\beta})_{RE} = \sqrt{\sum_{i=1}^{K} w_i (y_i - \hat{\beta})^2 / \{(K-1)\sum_{i=1}^{K} w_i\}}$ and $t_{K-1,1-\alpha/2}$ is the quantile $1-\alpha/2$ of a Studen's t distribution with $K-1$ degrees of freedom. The hypothesis $H_0 : \beta = 0$ can be tested using the statistics $T = Z/Q$ defined in (1.36) and the null hypothesis is rejected at level α when $\left|T^{obs}\right| > T_{K-1,1-\alpha/2}$, where T^{obs} is the observed value of the statistic.

1.5.3 Confidence intervals and hypothesis tests on τ^2

Concerning τ^2, a systematic review found nine different methods to provide confidence intervals (Veroniki et al., 2016). In this paragraph we will consider three of the most commonly used. Interestingly, in their original publication, DerSimonian & Laird (1986) did not provide the confidence interval for τ^2.

Based on likelihood theory, if the number of studies K is large, a Wald-type confidence interval of level α can be derived as

$$IC_{\text{Wald}}(\alpha) : \left\{\hat{\tau^2} - z_{1-\alpha/2} \cdot se(\tau^2), \hat{\tau^2} + z_{1-\alpha/2} \cdot se(\tau^2)\right\},$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha)/2$ quantile of a standard normal distribution. The normality assumption of the limiting distribution of τ^2 however, is highly debatable with small sample size, particularly for the assimetric shape of the variance distribution. To address the issue Hardy & Thompson (1996) proposed a confidence interval based on the profile likelihood ratio statistic and an interval of level α can be obtained by solving iteratively the following equation:

$$\ell_p(\hat{\beta}(\tau^2), \tau^2) > \ell_p(\hat{\beta}(\hat{\tau}^2), \hat{\tau}^2) - \chi^2_{1,1-\alpha}, \tag{1.37}$$

where $\ell_p(\cdot)$ has been defined in (1.27). The third method has been proposed by Viechtbauer (2007) and it is based on a generalization of the Q-statistic to the RE model. Particularly, denoting with $Q(\tau^2)$ the Q-statistic under the RE, we have

$$Q(\tau^2) = \sum_i^K \frac{Y_i - \hat{\beta}_{RE}}{\tau^2 + s_i^2} \sim \chi^2_{K-1}. \tag{1.38}$$

Indicating with $\chi^2_{K-1,\alpha/2}$ and $\chi^2_{K-1,1-\alpha/2}$ the $\alpha/2$ and the $1-\alpha/2$ of a χ^2_{K-1} distribution, it follows that $P(\chi^2_{K-1,\alpha/2} \leq Q(\tau^2) \leq \chi^2_{K-1,1-\alpha/2}) = 1-\alpha$. Using the inversion principle (Casella & Berger, 2021, Page 420)) it follows that the lower and upper bound of the confidence intervals are

$$\left\{Q(\tilde{\tau}^2_{lower}) = \chi^2_{K-1,1-\alpha/2}, \quad Q(\tilde{\tau}^2_{upper} = \chi^2_{K-1,\alpha/2}\right\}.$$

To calculate $\tilde{\tau}^2_{lower}$ and $\tilde{\tau}^2_{upper}$ it is sufficient to increase iteratively τ^2 until the value is reached. However, to respect the τ^2 parameter space, only non negative values can be provided. If $Q(0) < \chi^2_{K-1,\alpha/2}$, the interval is set to the null set. Please note that it is possible that the confidence interval does not actually contain the estimate of the between-study variance, except if the Paule-Mandel estimator is used (Viechtbauer, 2010b).

1.5.4 Heterogeneity

The studies considered in a meta-analysis exhibit several elements of diversity. Indeed, the authors may have employed different designs, sampled from different reference

populations, or implemented interventions of varying duration or effect. Higgins & Thompson (2002) refer to these characteristics as methodological or clinical heterogeneity, distinct from statistical heterogeneity. The latter exists when the true effect being evaluated differs between studies. In the RE model, the statistical heterogeneity is represented by the parameter τ^2, which is the variance of the distribution of the true effects. However, given that a variance is not a pure number, it does not allow for heterogeneity comparisons across meta-analyses based on different effect measures. Furthermore, lacking an upper limit, it is difficult to interpret its extent. Higgins & Thompson (2002) assumed the RE model with the within study variances $\sigma_i^2 = \sigma^2, \quad \forall i = \ldots, K$ and proposed three different statistics of the type $f(\beta, \tau^2, \sigma^2, K)$, based on the following principles:

1. $f(\beta, \tau_1^2, \sigma^2, K) > f(\beta, \tau_2^2, \sigma^2, K)$ if $\tau_1^2 > \tau_2^2$;
2. $f(a + b\beta, \tau_1^2, \sigma^2, K) = f(\beta, \tau_2^2, \sigma^2, K)$ for any a, b (scale invariance);
3. $f(\beta, \tau_1^2, \sigma^2, K_1) = f(\beta, \tau_2^2, \sigma^2, K_2)$ for any K_1 and $K - 2$ (size invariance).

It follows that the measure of heterogeneity should not involve β (criterion 2) and K(criterion 3) and it should increase monotonically with τ^2. Moreover, the second criteria suggests the use of a function of the ratio τ^2/σ^2. Note that they do not require the measure to be independent from the within study variability. With the assumption made in so far, it can be written:

$$\begin{aligned}
\hat{\tau}_{DL}^2 &= \sigma^2 \left(\frac{Q}{K-1} - 1 \right), \\
v_{CE} &= \frac{\sigma^2}{K}, \\
v_{RE} &\approx \frac{\sigma^2 + \tau^2}{K},
\end{aligned}$$

where Q is the value of the common effect Cochran's statistic defined in (1.8). Moreover, let $\rho = \tau^2/\sigma^2$, it follows

$$\rho + 1 = \frac{\tau^2 + \sigma^2}{\sigma^2}.$$

And substituting τ^2 with $\hat{\tau}_{DL}^2$ we obtain

$$H^2 = \frac{Q}{K-1}.$$

A second measure proposed by the authors is R^2 and particularly

$$\begin{aligned}\sigma^2 &= K \cdot v_{CE},\\ (\sigma^2 + \tau^2) &= K \cdot v_{RE},\\ R^2 = \rho + 1 &= \frac{\tau^2 + \sigma^2}{\sigma^2} = \frac{v_{RE}}{v_{CE}}.\end{aligned}$$

Finally, the third measure considered was defined as follows:

$$\frac{\rho}{1+\rho} = \frac{\tau^2}{\tau^2 + \sigma^2}.$$

Doing the plug-in with the estimated quantities, we obtain:

$$I^2 = \frac{\hat{\tau^2}}{\hat{\tau^2} + \hat{\sigma}^2}.$$

A generalization to meta-analysis with different precision is straightforward for H^2 and R^2 as they do not involve explicitly σ^2. However, Higgins & Thompson (2002) proposed the use of the "typical" within-study variance s^2 to generalized I^2. Using the result in (1.13), it follows that

$$E[H^2] = \frac{\tau^2 + s^2}{s^2},$$

where:

$$s^2 = \frac{\sum_i^K w_{i,CE}(K-1)}{(\sum_i^K w_{i,CE})^2 - \sum_i^K w_{i,CE}^2}.$$

The I^2 statistic can then be written as

$$I^2 = \frac{H^2 - 1}{H^2}.$$

The interpretation of these statistics is straightforward. H^2 represents the relative excess of Q over its degree of freedom. Since $E\{Q\} = k - 1$ in absence of heterogeneity, $H = 1$ indicates homogeneity of the treatment effect. R can be interpret as the inflation of the confidence interval for a single summary estimate under a random effect model compared with a fixed effect model. A value of 1 indicates the same inference under the two model. Finally, we could refer to I^2 as the proportion of total variation in the estimates of the treatment effect that it is due to heterogeneity between studies, if s^2 is considered a typical value for the within study variance.

The prediction interval is another tool that helps to understand the impact of heterogeneity. It represents an interval wherein there is a probability of 0.95 that the effect of a new study will fall within. It can be derived as follows:

$$\begin{aligned} &\hat{\beta} \sim \mathcal{N}(\beta, se(\hat{\beta})^2), \\ &\beta_{new} \sim \mathcal{N}(\beta, \tau^2), \\ &\frac{(\beta_{new} - \hat{\beta})}{\sqrt{\tau^2 + se(\hat{\beta})^2}} \sim N(0,1) \end{aligned} \tag{1.39}$$

However the true τ^2 is unknown and the common practice is to use a t_{K-2} as reference distribution. The Wald-type prediction interval of level α can be written as:

$$PI_{Wald}(\alpha) = \Big\{ \hat{\beta} - t_{1-\alpha/2,K-2} \cdot \sqrt{\hat{\tau}^2 + se(\hat{\beta})^2}, \\ \hat{\beta} + t_{1-\alpha/2,K-2} \cdot \sqrt{\hat{\tau}^2 + se(\hat{\beta})^2} \Big\}$$

1.5.5 Criticism and limitation

Given that the within-study model is a component of the RE model, the same criticism discussed in the section (1.2) applied. Additionally, the assumption that the studies are independent replications drawn from a Gaussian distribution is not widely accepted (Rice et al., 2018). For example, the symmetry of the distribution of effect sizes might be questionable. However, despite this limitation, the model does enable the quantification and examination of heterogeneity.

1.6 Meta-regression and mixed-effect model

Study-level covariates are frequently accessible in meta-analyses, and their inclusion can be valuable for elucidating the statistical heterogeneity observed. An illustrative case is evident when conducting a meta-analysis on the impact of a medical treatment, where the duration of treatment is incorporated as a study-level covariate. To accommodate the modeling of the effect measure as a function of study-level covariates, the meta-analysis model can be extended using the meta-regression model. Both these models can be understood and generalized within the framework of linear mixed-effects

models, providing a comprehensive approach to investigating heterogeneity in meta-analytic studies. Although models with more complex structures can be used, the most immediate extension of the meta-analysis model is one that includes dependence on covariates only in the linear predictor of the mean. Hence, a mixed-effect model can defined as follows:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + \delta_i + \epsilon_i, \\ &\text{with } \delta_i \sim \mathcal{N}(0, \tau^2),\ \epsilon \sim \mathcal{N}(0, s_i^2); \\ Cov\left(\delta_i, \delta_j\right) &= Cov\left(\epsilon_i, \epsilon_j\right) = 0. \quad \forall i \neq j,\ i, j = 1, ..., K, \\ \text{Cov}\left(\delta_i, \epsilon_j\right) &= 0 \quad \forall i, j = 1, ..., K, \end{aligned} \tag{1.40}$$

where the x_{ij} with $j = 1, \ldots, q$ are moderator variables. The present specification allows to explain the heterogeneity by means of a predictor linear in the parameters but not necessarily in the moderator variables (study-level covariates). The meta-regression model is obtained if $\tau = 0$. The inferential procedures follow those previously seen for the RE model, adjusting the number of degrees of freedom according to the number of moderators.

1.7 Higher-order asymptotic approximation

The results presented in sections 1.5.1.2 and 1.5.1.3 rely on a first-order asymptotic approximation, which may raise concerns when dealing with small sample sizes. It is worth noting that in the context of meta-analysis, a small sample size corresponds to a limited number of included studies. Consequently, the within-study model is still assumed to be valid in such cases.

Consider the signed version of likelihood ratio statistics used in (1.34) to compute the confidence interval for β. Given the parameter $\psi = (\beta, \tau^2)$, we assume to be interested in the inference on β. Hence, τ^2 qualifies as a nuisance parameter. We have that:

$$r_p(\beta) = sign(\beta - \hat{\beta})\sqrt{l_p(\hat{\psi}) - l_p(\tilde{\psi})} \tag{1.41}$$

where $\hat{\psi} = (\hat{\beta}, \hat{\tau^2})^T$ is the unconstrained (maximum likeliood) estimate of ψ and $\tilde{\psi} = (\hat{\beta}, \hat{\tau_\beta^2})^T$ is the constrained maximum likelihood estimate of ψ fixed β. Asymptotically it has been demonstrated that $r_p(\beta) \sim \mathcal{N}(0, 1)$ and confidence interval for β can be calculated as $z_{alpha/2} < r_p(\beta) < z_{1-alpha/2}$.

To increase the accuracy of the asymptotic approximation, the following modification has been proposed by Barndorff-Nielsen (1986)

$$r_p^*(\beta) = r_p(\beta) + \frac{1}{r_p(\beta)} \log \frac{u(\beta)}{r_p(\beta)}, \tag{1.42}$$

where $u(\beta)$ is a function of the observed information evaluated at the unconstrained maximum likelihood estimate and the sample space derivatives of likelihood quantities with respect to the maxium likelihood estimates (Guolo, 2012). However, outside of the exponential model the calculation of $u(\beta)$ is impractical as it is difficult to determine an ancillary statistics.

A valid solution is the use of Skovgaard (1996) approximation of the (1.42), defined as

$$\bar{r_p}(\beta) = r_p(\beta) + \frac{1}{r_p(\beta)} \log \frac{u(\beta)}{r_p(\beta)}. \tag{1.43}$$

In fact, it does not require the specification of the ancillary statistic and the evaluation of the sample space derivatives. Moreover, the (1.42) has a standard normal approximation up to an error of order $O(n^{-3/2}$ in the moderate-deviation case and $O(n^{-1}$ in the large-deviation case, while the (1.43) up to an error of order $O(n^{-1}$ in the moderate-deviation case and $O(n^{-1/2}$ in the large-deviation case.

The approximation is exact in a full exponential family and the (1.43) becomes

$$\bar{r_p}(\beta) = r_p(\beta) + \frac{1}{r_p(\beta)} \log \frac{\bar{u}(\beta)}{r_p(\beta)}, \tag{1.44}$$

where $\bar{u}(\beta)$ is

$$\bar{u}(\beta) = \left[S^{-1}q\right]_\beta |\hat{j}|^{1/2} |\hat{i}|^{-1} |S| + |\tilde{j}_{\lambda,\lambda}|^{-1/2}, \tag{1.45}$$

and $\hat{j}$ is the expected and $\hat{i}$ the observed information, both evaluated in the maximum likelihood estimate $\hat{\psi}$. $\tilde{j}_{\lambda,\lambda}$ is a block of the expected information matrix, corresponding to λ and evaluated at the constrained maximum likelihood estimate $\tilde{\psi}$. Finally, $[S^{-1}q]_\beta$ is the component of the vector $S^{-1}q$ corresponding to β, where:

$$S = \mathrm{Cov}_{\psi_1} \left(\frac{\partial l_p(\psi_1)}{\partial \psi}, \frac{\partial l_p(\psi_2)}{\partial \psi} \right) \Bigg|_{\psi_1 = \hat{\psi}, \psi_2 = \tilde{\psi}}$$

and

$$q = \mathrm{Cov}_{\psi_1} \left[\frac{\partial l_p(\psi_1)}{\partial \psi}, l_p(\psi_1) - l_p(\psi_2) \right] \Bigg|_{\psi_1 = \hat{\psi}, \psi_2 = \tilde{\psi}}.$$

The Skovgaard's statistic is particularly useful in meta-analysis as the random effect meta-regression cannot be seen as an exponential family, except if $\hat{\sigma}_i^2 = \hat{\sigma}^2 \forall i = 1, ..., K$ and as a consequence the use of the original Barndoff-Nielsen's approximation is impractical.

Given the model (1.10) and taking $\psi = (\beta, \tau^2)^T$, let $\hat{\beta}$ and $\hat{\tau}^2$ the unconstrained maximum likelihood estimates and $\tilde{\beta}$ and $\tilde{\tau}^2$ the respective constrained maximum likelihood estimate, the S matrix in the Skovgaard's statistic is

$$S = \begin{bmatrix} S_{\beta\beta} & S_{\beta\tau^2} \\ S_{\tau^2\beta} & S_{\tau^2\tau^2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^K (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} & \sum_{i=1}^K (\hat{\beta} - \tilde{\beta}) (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-2} \\ 0 & \sum_{i=1}^K (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-2} /2 \end{bmatrix}$$

and q is a vector of two components equal to

$$q = \begin{bmatrix} q_\beta \\ q_{\tau^2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^K (\hat{\beta} - \tilde{\beta}) (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} \\ -\sum_{i=1}^K \left\{ (\hat{\sigma}_i^2 + \hat{\tau}^2)^{-1} - (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} \right\} /2 \end{bmatrix}.$$

Given the meta-regression model (1.40) with p covariates, the matrix S is a $(p+1)\times(p+1)$ matrix

$$S = \begin{bmatrix} \sum_{i=1}^K X_i X_i^{\mathcal{T}} (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} & \sum_{i=1}^K X_i X_i^{\mathcal{T}} (\underline{\hat{\beta}} - \underline{\tilde{\beta}}) (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-2} \\ 0 & \sum_{i=1}^K \{2 (\hat{\sigma}_i^2 + \tilde{\tau}^2)\}^{-1} \end{bmatrix},$$

and q is a $(p+1)$ components vector:

$$q = \begin{bmatrix} \sum_{i=1}^K X_i X_i^{\mathcal{T}} (\underline{\hat{\beta}} - \underline{\tilde{\beta}}) (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} \\ -\sum_{i=1}^K \left\{ (\hat{\sigma}_i^2 + \hat{\tau}^2)^{-1} - (\hat{\sigma}_i^2 + \tilde{\tau}^2)^{-1} \right\} /2 \end{bmatrix}.$$

The exact calculation are reported in Guolo (Appendix A, 2012).

From the results of the simulations presented in the article Guolo (2012) note that the Skovengard's correction achieves nominal 95% coverage of confidence intervals for β in the meta-analysis random effect model and for β_1 in a single covatiate meta-regression model even with relatively small number of studies (less than 10) and regardless of the value of τ^2. In addition, the Skovengard's statistic seems to perform particularly better than the REML and profile likelihood even considering the empirical coverage of the confidence intervals for the parameter τ^2. In fact, the empirical coverage of 95% is guaranteed with a number of studies less or equal to 15, regardless of the value of τ^2 (Guolo, 2012).

1.8 Other small sample corrections

1.8.1 Bartlett's correction

The asymptotic distribution of the likelihood ratio statistic $r_p(\beta)^2$ is χ^2_1 distribution. To construct a $1-\alpha$ confidence interval for β, one can solve the equation $r_p(\beta)^2 \leq \chi^2_{1,1-\alpha}$. Furthermore, hypothesis testing can be performed by comparing $r_p(\beta)^2$ to the quantile $\chi^2_{1,1-\alpha}$. It is important to note that this test relies on first-order asymptotics, and the accuracy of the χ^2_1 approximation may be affected by small sample sizes. To address this issue, the Bartlett's correction is introduced as $(1 + A)^{-1} r_p(\beta)^2$, where A represents a function of both between and within study variances. The development of a generalized formula for the Bartlett's correction in the mixed linear model was carried out by Zucker et al. (2000), and the specific formula for the meta-regression mixed effect model can be found in the Appendix of Huizenga et al. (2011).

1.9 Non-parametric approaches

The non-parametric approach have been proposed to circumvent the distributional assumption and the criticism associated with it (Jackson & White, 2018). These approaches offer advantages in terms of robustness, particularly in the face of model misspecification. However, it is important to note that non-parametric methods can be computationally intensive and may have reduced statistical power, due to the less amount of information introduced in the model.

1.9.1 Permutation test

The permutation method can be used to test a specific hypothesis. In fact, the test statistics is calculated for each of the M permutation and the process is possible because data are exchangeable under the hypothesis of no effect. The two sided p-value is obtained by doubling the proportion of permutation statistics that exceed the value of the observed statistic. The application in meta-analysis proposed by Follmann & Proschan (1999), is quite simple and it consist of permuting the signs of effect sizes as positive and negative value are considered equally likely under the null hypothesis. In meta-regression, the approach can be applied permuting the row of the design matrix as under the null hypothesis there is no relationship between the covariate and the measure of effect (Higgins & Thompson, 2004) suggest permuting the rows of the design matrix. The permutation method has several limitation as it may not achieve conventional levels

of statistical significance such as 0.05 when the number of studies is small (Viechtbauer, 2010a). For instance, when n = 5, the smallest possible P-value is 0.0625, and when $K = 6$, it is 0.031. When the number of studies included is too large, the method can be computationally intensive and require a considerable amount of time to evaluate all the possible permutation. The main advantage however, is that it can be used with any statistics of interest.

1.9.2 Resampling

The resampling method aim to estimate the distribution of the test statistic from the data. Huizenga et al. (2011) consider a form of residual resampling for testing the significance of β. Particularly, the following steps have been applied:

1. the mixed effect meta-regression model (1.40) is fit on the data and the desired statistics is computed. (Huizenga et al., 2011) test the significance coefficient of beta using $Z = \hat{\beta}_j / se(\hat{\beta}_j)$, where $se(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$ obtained with the maximum likelihood method.

2. set $\hat{\beta}_j = 0$ and compute the reduced model residuals as $r_i = y_i - \hat{y}_i$, where $\hat{y}_i$ is the predicted y_i under the reduced model where $\beta_j = 0$.

3. for h from 1 to B:

    a. resample K times without replacement the residuals r_i,

    b. add each r_i to the model for y_i,

    c. name Δ the matrix containing the between study variances, i.e. the element δ_i in the model (1.40). The diagonal element of the estimated Δ, i.e $\hat{\delta}_i$ has to be reordered matching the order of the resampled residual r_i.

    d. fit the full model (with β_j not equal to zero) on the data and compute the $Z_h^* = \hat{\beta}_j / se(\hat{\beta}_j)$ statistic.

4. the p-value can be calculated a the proportion of $Z_h^* > Z$.

The resampling procedure does not require any distributional assumption and so it is robust against many types of misspecification. However, the maximum amount of unique permutation is K and so the p-values are multiple of $1/K!$. Hence, the minimum p-value attainable is $1/K!$ (Huizenga et al., 2011). In other words, the number of study K could affect the precision of the conclusions. Other resampling methods are available in the literature, for example Van Den Noortgate & Onghena (2005) proposed four

different types of bootstrap. Two of them were non-parametric, i.e. the case and the error bootstrap, while the other two were parametric, i.e. the effect size and raw data bootstrap. Given that Van Den Noortgate & Onghena (2005) suggested not to use the case bootstrap in meta-analysis and that the other three types of bootstrap gave similar results, we limit ourself to describe one of the parametric approaches. The effect size bootstrap is based on the model (1.40) and it consist of repeating B times the following steps:

1. estimate the model (1.40) on the data;
2. draw a set of K δ_i^* from $\mathcal{N}(0, \hat{\tau^2})$, which is the distribution of the random intercept of the mixed effect meta-regression model;
3. generate the new true effect measure as $\hat{\beta}_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_q x_{iq} + \delta_i^*$;
4. draw a set of K residuals as $R_i^* \sim \mathcal{N}(0, \hat{\sigma}_i^2)$, where $\hat{\sigma}_i^2$ is estimated as $\hat{\sigma}_i^2 = \frac{N_i}{n_{i,E} n_{i,C}} + \frac{\hat{\beta}_i^*}{2N_i}$ (according to (Hedges, 1981)) and $n_{i,C}$ is the sample size of the control group, $n_{i,E}$ of the experimental one and $N_i = n_{i,C} + n_{i,E}$ for the $i - th$ study.
5. calculated the $y_i^* = \hat{\beta}_i^* + r_i^*$, with r_i^* realization of R_i^*.

The parametric bootstrap assumes that the model is correctly specified as well as the random effect and error distribution. Once the boostrap distribution has been derived, different types of confidence intervals are availables. A complete discussion can be found in the paper of Carpenter & Bithell (2000).

1.10 Checking model assumptions

Because the two-stage meta-analysis model is based on several assumptions, it is recommended that an empirical analysis be conducted to evaluate the extent to which the assumptions of the model are met. This will help to ensure the validity of inferential conclusions . In this section, we focus on the between-study level assumptions and assume that the within-study level assumptions are met. In reality, violations of the within-study model are more likely to be the result of decisions made during the planning and data extraction phases of the meta-analysis than of statistical decisions. For example, when dealing with binary data, the choice of effect measure and its transformation (e.g., log odds ratio) directly affects the speed of convergence to a normal distribution.

1.10.1 Model fit and between-studies normality assumption

Considering the general case of the mixed-effects meta-regression model specified in (1.40), the internally stundentized residuals can be defined as

$$r_i = \frac{y_i - \hat{\mu}_i}{Var(y_i - \hat{\mu}_i)}$$

where $\hat{\mu}_i$ is the estimated of $E\left[Y_i|\beta_0, \beta_1, \ldots, \beta_q x_{iq}\right]$ which can be written as $\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_q x_{iq}$. The sampling variance of the residuals is equal to $Var(y_i - \hat{\mu}_i) = (1 - h_i)(s_i^2 + \tau^2)$, where h_i is the i-th diagonal element of the $\mathbf{H}$ matrix and it is well known in the linear model theory as "leverage". The $\mathbf{H}$ is defined as $\mathbf{H} = \mathbf{X}(\mathbf{X^T}\tilde{\mathbf{W}}\mathbf{^T}\mathbf{X})^{-1}\mathbf{X^T}\tilde{\mathbf{W}}$, with $\mathbf{X}$ being the $K \times q$ model matrix and $\tilde{W} = \left[diag(1/(s_i^2 + \tau^2))\right], \quad \forall i = 1, \ldots, K$. The $\tilde{W}$ is the weighted least square weight matrix. In a meta-regression model, outliers can potentially inflate the estimate of τ^2, leading to small residuals and at the same time they can pull μ_i toward themselves, reducing the numerator. For this reason, Lv (1985) proposed the external studentized residuals for the fixed effect model. They are also known as deleted residuals in fact Viechtbauer & Cheung (2010) defined them as

$$t_i = \frac{y_i - \hat{\mu}_{i(-i)}}{Var(y_i - \hat{\mu}_{i(-i)})} = \frac{y_i - \hat{\mu}_{i(-i)}}{Var(s_i^2 + \hat{\tau}^2_{(-i)} + Var(\hat{\mu}_{i(-i)}))}$$

where $\hat{\mu}_{i(-i))}$ is the predicted average for the study i from the estimated model excluding the i-th study. In the second equality, the term $\hat{\tau}^2_{(-i)}$ indicates the estimate of τ^2 from the model excluding the i-th study and $Var(\hat{\mu}_{i(-i)})$ is the estimated variance of $\hat{\mu}_{i(-i)}$ from the model that excludes the i-th study. The simplification is justified considering that y_i and $\hat{\mu}_{i(-i)}$ are not correlated. If the studies followed the assumed model, the studentized deleted residuals approximately follow a normal distribution $\mathcal{N}(0, 1)$. Therefore, in the plot (i, t_i), only 5% of the points should lie outside the bounds of $\pm\phi^{-1}(0.975)$, where $\phi^{-1}(\cdot)$ indicates the quantile function of a standard normal distribution. Similarly, the quantile-quantile plot against a standard normal distribution can be useful for visually checking the goodness of fit. A formal assessment of the normality assumption can be carried out with any test suitable for the purpose, for example the Shapiro-Wilk or the Kolmogorov-Smirnov. Moreover, as explained by Viechtbauer & Cheung (2010), the studentized deleted residuals formalize a test for outlier. Given the model (1.40), we assumed one study not following it. In particular, the outlier study $\tilde{i}$ follows a model with conditional expected value

$$E\left\{Y_i|x_{\tilde{i}1}, \ldots x_{\tilde{i}1}\right\} = \beta_0 + \beta_1 x_{\tilde{i},1} + \cdots + \beta_q x_{\tilde{i},q} + \lambda,$$

where λ is the fixed amount by which the study $\tilde{i}$ shifted away from the true model. The hypothesis of $H_0 : \lambda = 0$ can be tested adding a dummy variable to the model which takes the value of 1 only for the $\tilde{i}$-th study. The statistics to test the hypothesis would be the standardized deleted residual of the $\tilde{i}$-th study.

1.10.2 Influential studies

While a few outliers may not create undue problems, it is necessary to recognize whether a study affects the model's predictions, or the estimation of the parameters of interest. The statistics of interest to evaluate the issue fall under the name of case deletion diagnostics and in this section we will follow the discussion made in Viechtbauer & Cheung (2010). A point to start is

$$\text{DFFITS}_i = \frac{\hat{\mu}_i - \hat{\mu}_{i(-i)}}{\sqrt{h_i \left(v_i + \hat{\tau}^2_{(-i)} \right)}},$$

which is the difference between the predicted average effect for the i-th estimated including or not the study itself. The difference is standardized by the standard error $\hat{\mu}_i$, calculated replacing $\hat{\tau}^2$ with $\hat{\tau}^2_{(-i)}$. The statistic resemble the standardized mean difference, in fact it quantifies how much the two prediction differs, including and excluding the i-th study, in standard deviation unit. Going more in depth, a Cook's distance similar measure is given by

$$D_i = \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)} \right)^T \left(X^T \tilde{W} X \right) \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)} \right) = \sum_i^K \frac{\left(\hat{\mu}_i - \hat{\mu}_{i(-i)} \right)^2}{v_i + \hat{\tau}^2}$$

where $\hat{\boldsymbol{\beta}}_{(-i)}$ denotes the vector of parameter estimates from the fitted model after deletion of the i-th study. It can be interpreted as the Mahalanobis between the predicted values with and without the i-th study. Furthermore, the join confidence region of the regression coefficients is defined by

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \left(\boldsymbol{X}^T \tilde{W} \boldsymbol{X} \right) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \chi^2_{p^T, 1-\alpha}$$

with $p' = p+1$. Therefore, a value of D_i equal to $\chi^2_{p', 1-\alpha}$ indicates that the deletion of the i-th study would move the parameter estimates to the edge of a $(1-\alpha)$ joint confidence region based on the complete data. Based on Cook & Weisberg (1982) worrisome values are defined as the $D_i > \chi^2_p, 0.5$.

To examine the influence of the i-th study on the regression coefficient the following statistic has been proposed:

$$\text{DFBETAS}_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(-i)}}{\sqrt{\left(X^T\tilde{W}_{(-i)}X\right)^{-1}_{[j^Tj^T]}}}$$

where $\left(X^T\tilde{W}_{(-i)}X\right)^{-1}_{[j'j']}$ denotes the value of the $(j+1)$ th diagonal element of the matrix $\left(X^T\tilde{W}_{(-i)}X\right)^{-1}$ and $\tilde{W}_{(-i)} = \text{diag}\left[1/\left(v_1 + \hat{\tau}^2_{(-i)}\right), 1/\left(v_2 + \hat{\tau}^2_{(-i)}\right) \ldots, 1/\left(v_k + \hat{\tau}^2_{(-i)}\right)\right]$. The formula can be simplified to

$$\text{DFBETAS}_i = \left(\hat{\mu} - \hat{\mu}_{(-i)}\right)\sqrt{\sum_{l=1}^{k}\tilde{w}_{l(-i)}}$$

in the random-effects model, where $\tilde{w}_{I(-i)} = 1/\left(v_l + \hat{\tau}^2_{(-i)}\right)$.
The influence of the i-th study can also be examined by means of the change in the variance–covariance matrix of the parameter estimates when excluding the study from the model fitting. The ratio of the generalized variance is defined as

$$\text{COVRATIO}_i = \frac{\det\left\{\text{Var}\left(\hat{\boldsymbol{\beta}}_{(-i)}\right)\right\}}{\det\{\text{Var}(\hat{\boldsymbol{\beta}})\}}$$

and a value below 1 is suspicious as it indicates that the removal of the i-th study lead to a more precise estimate of the model coefficient.
If τ^2 is of interest, it may be informative to examine the influence of outlier on its estimate too. The effect of the removal of the i-th study can be evaluated using

$$R_i = 100 \times \left(\hat{\tau}^2 - \hat{\tau}^2_{(-i)}\right)/\hat{\tau}^2$$

quantifying the percent change in the estimate of τ^2 when the i-th study is excluded relative to the estimated amount of (residual) heterogeneity when all the studies are included.

1.11 Publication bias

The validity of conclusions drawn from a meta-analysis is affected by publication and outcome reporting bias. A meta-analysis is intended to be a comprehensive synthesis of all available evidence, but some studies may not be included if the authors misreported

their results or if the study was not published. As a result, the estimate of the meta-analysis may not reflect the true effect of the treatment (in a broad sense). Some statistical methods have been developed to examine the presence of these two types of bias; the following are the most widely used, although probably not the most effective.

1.11.1 Estimation of the number of unpublished studies

The fail-safe N method aims to calculate the number of studies with negative results needed to transform the meta-analysis result into non-significant. Given a set of studies $i = 1, \ldots, K$ and considering the hypothesis $H_0 : \beta_i = 0$, we can consider the statistics

$$Z_s = \frac{\sum_{i=1}^{K} z_i}{\sqrt{K}}$$
$$z_i = \Phi^{-1}(p_i)$$

and p_i, the p-value of the i-th study. Under the null hypothesis $H_0 : \beta_i = 0 \, \forall i$, $Z \sim N(0, 1)$. To not reject the null hypothesis we should have:

$$\begin{aligned} &\frac{\sum_{i=1}^{K} z_i}{\sqrt{K+N}} \leq Z_\alpha \\ &N > \left(\frac{\sum_{i=1}^{K} z_i}{Z_{alpha}} \right)^2 - K = K \left(\frac{Z_s}{Z_\alpha} \right)^2 - K, \end{aligned} \tag{1.46}$$

where N is the number of studies necessary to change the conclusion of the meta-analysis.
Note that while the denominator in the first line of the (1.46) is increased by N, the sum at the numerator remains indexed in K. The reason is that we add trial with a true effect of zero. The major limitations of this method are that it relies only on p-values and that the missing studies have an exact zero effect.

1.11.2 Funnel plots

The Funnel plot graphs pairs of $(y_i, \sqrt{(s_i^2 + \tau^2)}^{-1})$, but any kind of precision measure (such as variance or sample size) can be used in the ordinate axis. Using the standard deviation, the boundary lines have a slope of $1/\Phi^{-1}(1 - \alpha/2)$. Generally, the plot is drawn with a vertical line corresponding to the estimate of the average effect among the plotted studies. In the interpretation of the Funnel plot, it should be considered the

assumption of a common average among the plotted studies. If there were one or more important moderators, the single units would be clustered and the interpretation would not be valuable in term of asymmetry. In addition, if an effect measure is used such that the variance depends on the effect (e.g. odds ratio), there will be a natural asymmetry in the graph. To improve the reading and interpretation of the plots, the Contour-enhanced funnel plots (Peters et al., 2008) have been proposed. Under the normality assumption of the treatment effect, the significance of any effect size can be calculated from the effect size and the standard error (the two axis of the plot). Thus, contours representing conventional "milestone" levels of statistical significance (e.g., < 0.01, < 0.05, < 0.1) can be defined and regions associated with these significance levels plotted. Note that the contour enhance funnel plot is centered on zero and not on the model estimate.

1.11.3 Egger's regression test

The appearance of the funnel plot is largely determined by the effect of chance, so several statistics have been created to assess whether the association between the size of the effect measure and the chosen precision measure is greater than would be expected by pure effect of chance. According to Egger et al. (1997) the following model can be estimated

$$E\left[Z_i\right] = \beta_0 + \beta_1(1/s_i).$$

where $z_i = y_i/s_i$. In Egger's original formulation, the hypothesis to be tested was $H_0 : \beta_0 = 0$. Under the null hypothesis, the slope of Egger's regression would be equal to the joint effect and $\beta_0 = 0$ because the smallest study would have an effect close to zero and a small standard error. The test is equivalent to estimating the weighted regression of y_i on s_i, where the weighting for $1/s_i$ and the coefficient β_1 is the parameter of interest. In this sense, the regression model can also be run with s_i^2 instead of s_i. The main limitation is that the Egger approach does not test the publication bias hypothesis, but only the asymmetry of the funnel plot. It is therefore necessary to rule out other possible causes of the asymmetry before the results of the test can be used as evidence in favour or against publication bias.

1.11.4 Trim and fill

The trim-and-fill method is a non-parametric approach that try to idenfity and correct the funnel plot asymmetry. The algorithm works as follows (Chapter 13, section 13.8, page295, Schmid et al. (2020)):

1. Order the observed treatment effects by size and potentially re-code them so that the missing studies have negative effects corresponding to asymmetry on the left hand side of the funnel plot;
2. Estimate the mean treatment effect, β, using a common or random-effects model;
3. Estimate the number of unpublished studies, u;
4. Trim off or remove the u most extreme studies causing asymmetry
5. Re-estimate β and u using the trimmed data, repeating the trimming process until u stabilizes;
6. Replace the trimmed studies and impute their missing counterparts symmetrically about β;
7. Calculate a final adjusted estimate of β using both observed and imputed studies.

It results that the imputed studies are symmetric in the effect but opposite in direction to the u studies chosen to be omitted in order to make the plot symmetric. It is known that in step 3 different methods can be used, for example u is estimated as one less than the number of positive effects that are larger than the absolute value of the most negative effect. The weaknesses of the method are the strong assumption, in fact it assumes:

- that funnel plot asymmetry (small-study effects) occurs only because of publication bias, ignoring all the other mechanisms;
- the existance of a symmetric target funnel plot.

Chapter 2

Application to a case study

2.1 Introduction to the dataset

In this chapter we will use the methods described in Chapter 1 to reproduce and further investigate the statistical analysis of a paper published in 2021 by Jones and colleagues (Jones et al., 2021). The original dataset is available on the The Open Science Framework (OSF) repository (`https://osf.io/afn3y/`). The aim of the meta-analysis is to assess sex-related differences in adaptations to resistance training and specifically in older adults. The research design is globally well structured, and we will indicate the key methodological (nonstatistical-mathematical) details below. A brief paragraph describing resistance training and the key variables governing exercise prescription will also be provided before the results are presented.

2.2 Resistance training in older adults

Resistance training is defined as "a specialized method of conditioning, which involves the progressive use of a wide range of resistive loads and a variety of training modalities designed to enhance health, fitness, and sports performance" (Faigenbaum et al., 2009). It is also commonly known as "strength training" or "weight training". When prescribing resistance training exercises, several crucial variables must be taken into account, including relative and absolute intensity, proximity to failure, frequency, volume, velocity, and exercise selection (Scott et al., 2016). Relative and absolute intensity refer to the magnitude of the resistance moved, either in absolute terms (e.g., weight on the bar) or relative to a benchmark (e.g., one-repetition maximum). Proximity to

failure represents the number of repetitions that can be performed before reaching failure. The literature has presented multiple definitions of failure (Steele et al., 2017), but due to the heterogeneity in its definition (and the inaccurancy in reporting it), Jones and collegues (Jones et al., 2021) could not extract this variable from all the study. Frequency relates to the number of training sessions conducted per week, while volume pertains to the total number of sets and repetitions performed for each exercise. Exercise selection involves choosing specific exercises that target the desired muscle group as the prime mover. Finally, extensive research has been conducted on periodization and models of progression in resistance training (of Sports Medicine et al., 2009). Sarcopenia, characterized by a gradual loss of muscle mass, strength, and muscle performance accompanied by alterations in muscle fibers, poses a significant concern for older adults (Frontera et al., 2000). This condition represents a prominent public health issue due to its association with adverse events and subsequent illness in later stages of life (Du et al., 2019), disability (Fielding et al., 2011), and an increased risk of falls (Yeung et al., 2019). Resistance training is the preferred exercise modality (and the gold standard) for accrual of skeletal muscle (Schoenfeld, 2010), with adaptations possible throughout the lifespan, even in nonagenarians (Fiatarone et al., 1990). Moreover, this exercise modality significantly contributes to healthy aging, positively impacting diverse health-oriented and performance-related outcomes. Consequently, it is endorsed and recommended by the World Health Organization and the National Strength and Conditioning Association for older adults (Bull et al., 2020; Fragala et al., 2019). Nevertheless, investigating the physiological differences in response to resistance training programs between sexes remains of utmost interest. Gender-specific dissimilarities have been observed in inflammatory responses subsequent to muscle-damaging eccentric exercise (Stupka et al., 2000), the recovery time course following resistance training (Flores et al., 2011), and muscle fiber size and composition (Haizlip et al., 2015). Given the extensive body of literature accumulated in recent years on the subject, Jones et al. made the decision to undertake a meta-analysis, which was subsequently published as Jones et al. (2021)

2.3 Methodological details in the dataset construction phase

2.3.1 Literature search, inclusion and exclusion criteria

The literature search was conducted using five well-recognized and well-established electronic databases:

- Web of Science,
- Science Direct,
- SPORTDiscus,
- CINAHL,
- MEDLINE.

The authors identified 7560 abstracts but the total amount of full text included was 30 (2 of them by manual search). The retrieved articles were included if they were prospective Resistance Training (RT) trial, published in English and including participants older than 50 years old. In physiology a well-recognize cut-off age to differentiate between adults and older adults is 65 years old, however the choice was justify assuming that hormonal alteration due to menopause might limit the effect of RT before the usual target age. Finally, only dynamic RT intervention, where men and women performed the same program, were considered. The exclusion criteria mostly encompass medical conditions, injury or failing in reporting the outcomes.

2.3.2 Outcomes selection

The adaptations to the training stimulus were represented by three constructs:

- maximal upper-body strength;
- maximal lower-body strength;
- muscle size.

Scientific investigations often report multiple realizations of these outcomes because muscle strength is assessed on different movements (and possibly with different techniques) and "muscle size" can be evaluated at least at local or whole body level. To maintain the stochastic independence of the effect sizes across different studies, outcome extraction hierarchies have been adopted. For maximum strength the authors proceed from less-skilled bi-articular movement, to bi-articular skilled movement, to single joint exercise. The direct assessment was always the preferred test modality. For "muscle size" the technique with the highest construct validity was chosen and the whole body assessment was preferred to the local one. However, when the training protocol was specific to a body region and both whole and local assessment were available, the more protocol-specific one would be chosen. Hierarchies are deterministic rules to guarantee

the stochastic independence between the measure of effect from different studies. However, this solution did not address the the presence of studies with multiple outcome due to multiple experimental groups. Regarding the latter detail, its handling by the authors has not been clarified. However, two effect measures for the same study, are sometimes present in the dataset. In this paper, the possible dependence of these effects has been ignored, well aware that it constitutes a limitation. However, it allows to reproduce the original results and the comparison with the exposed methods.

2.3.3 Data extraction

Since all the studies included adopted a longitudinal design, at least one measure before and one after the administration of the intervention protocol was available. For each outcome of interest, the mean and standard deviation of the absolute and relative changes from baseline for male and female were extracted. When the standard deviation was not available, it was calculated as (Higgins & Deeks, 2011)

$$\hat{\sigma}_{d,ij} = \sqrt{\hat{\sigma}^2_{pre,ij} + \hat{\sigma}^2_{post,ij} - 2 \cdot (\hat{\rho}_{pp,ij} \cdot \hat{\sigma}_{pre,ij} \cdot \hat{\sigma}^2_{post,ij})},$$

where $\hat{\sigma}_{d,ij}$ is the estimated standard deviation of the $post - pre$ differences of the outcome i in the study j, $\hat{\sigma}_{pre,ij}$ is the estimated standard deviation of the pre score, $\hat{\sigma}_{post,ij}$ is the estimated standard deviation of the post score and $\hat{\rho}_{pp,ij}$ is the estimated correlation between the pre and post values. In case where $\hat{\rho}_{pp,ij}$ was not available, the authors plugged-in a value derived from the studies with enough information to obtain a correlation estimate. Firstly, they estimated $\hat{\rho}_{pp,ij}$ as (Higgins & Deeks, 2011)

$$\hat{\rho}_{pp,ij} = \frac{\hat{\sigma}^2_{pre,ij} + \hat{\sigma}^2_{post,ij} - \hat{\sigma}^2_{pp,ij}}{2 \cdot \hat{\sigma}_{pre,ij} \cdot \hat{\sigma}_{pre,ij}}, \tag{2.1}$$

where $\hat{\sigma}_{pp,ij}$ is the estimated standard deviation of the change scores. Due to the paucity of study reporting all the information necessary to compute the within group post-pre correlation (equation 2.1), males and females data were pooled. Secondly, they used the following estimator:

$$\hat{\rho}^*_{pp,ij} = \text{med}(\hat{\rho}_{pp,ij}) - 0.1$$

where $\text{med}(\cdot)$ indicates the sample median. A sensitivity analysis were performed forcing $\hat{\rho}^*_{pp,ij} = 0.5$. The Table 2.1 reports the estimated $\hat{\rho}^*_{pp,ij}$ values used in the original publication.

TABLE 2.1: Estimated pre-post correlation values

Outcome	$\hat{\rho}^*_{pp,ij}$
Upper-body strength	0.78
Lower-body strength	0.70
Muscle size	0.87

Unfortunately, the authors do not offer theoretical justification of the estimator used to $\hat{\sigma}_{pp,ij}$.

2.3.4 The measure of effect

Once the group-level data were obtained, the Hedge's g (Hedges, 1981) was computed as the study-level effect measure. The statistic is a standardized mean difference defined as

$$y_{ij} = c \cdot \frac{\hat{\mu}_{male,ij} - \hat{\mu}_{female,ij}}{\hat{\sigma}_p},$$

with

$$\hat{\sigma}_p = \sqrt{\frac{(n_{male,ij} - 1)\hat{\sigma}_{d,male,ij} + (n_{male,ij} - 1)\hat{\sigma}_{d,female,ij}}{n_{male,ij} + n_{female,ij} - 2}},$$

and

$$c = 1 - \frac{3}{4m - 1} \cdot y_{ij},$$

where $m = n_{male,ij} + n_{female,ij} - 2$, $\hat{\mu}_{male,ij}$ is the estimated average difference $post - pre$ in the outocome j, $n_{male,ij}$ is the sample size of the group and $hat\sigma_{d,female,ij}$ is the estimated standard deviation of the $post - pre$ differences in the male group and the same applies to female. The constant c is a correction for small sample size to obtained an unbiased estimate of the effect measure. The variance of the effect measure can be computed as (Hedges, 1981)

$$s^2_{ij} = \frac{2}{m_i} \cdot \left(1 + \frac{y^2_{ij}}{4}\right).$$

From now on, the study-level data will be indicated as (y_i, s^2_i) because the number of outcomes j is different for each study and therefore a univariate model for each research question was proposed.

2.4 Auxiliary variables

The working dataset is encoded in an excel file named "Data_190620.xslx" and contains multiple variables. In addition to effect measures and labels useful in identifying the studies included in the meta-analysis, the variables considered are:

- `training`, which represents the type of training performed during the study. It is divided into "upper body," "lower body," and "full body."
- `duration`, which indicates the duration of the study in weeks;
- `metareg_intensity`, which indicates the number of repetitions per week by adding up each exercise;
- `metareg_intensity`, which indicates the intensity expressed as a percentage of the 1RM (the maximum weight lifted for a single repetition in a given exercise);
- `metareg_frequency`, which indicates the weekly training frequency;
- `metareg_exercise`, which indicates the number of exercises;
- `metareg_sets`, which indicates the average number of sets per exercise.

2.5 Analysis of the relative changes in upper body muscle strength

The RE model estimated on the relative change in upper body muscle strength is based on the results of 7 studies and 160 individuals, of whom 80 are female. The point estimate of the parameter of interest is $\hat{\beta} = -0.29$, with an associate 95% confidence interval equal to $(-0.62, 0.04)$. Accordingly, the p-value for the test of the null hypothesis of no effect is $p = 0.09$. Therefore, the null hypothesis that there is no difference between males and females in upper body strength gain cannot be rejected when the change is expressed relative to baseline. The Cochran's Q test is not significant ($Q = 6.60, \quad df = 6, \quad p = 0.36$) and thus the hypothesis of a common true effect is not rejected. However, as explained earlier, this could be due to a lack of statistical power, and given the limited number of included studies, it is better to discuss the results of the RE model than to estimate the CE one. Nevertheless, the statistic indicating the percentage of variance explained by heterogeneity is $I^2 = 4.90\%$, indicating that the main source of variance is due to within-study variability. The results obtained so far

are based on the standard method for estimating RE models introduced by DerSimonian & Laird (1986). Moreover, the point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0.10$, with an associate 95% confidence interval of $(< 10^{-4}, 1.13)$, calculated using the method of Viechtbauer (2007). The results presented so far were shown in the forest plot in Figure 2.1. Given the small percentage of remaining heterogeneity and the small number of included studies, it does not seem reasonable to estimate a meta-regression model.

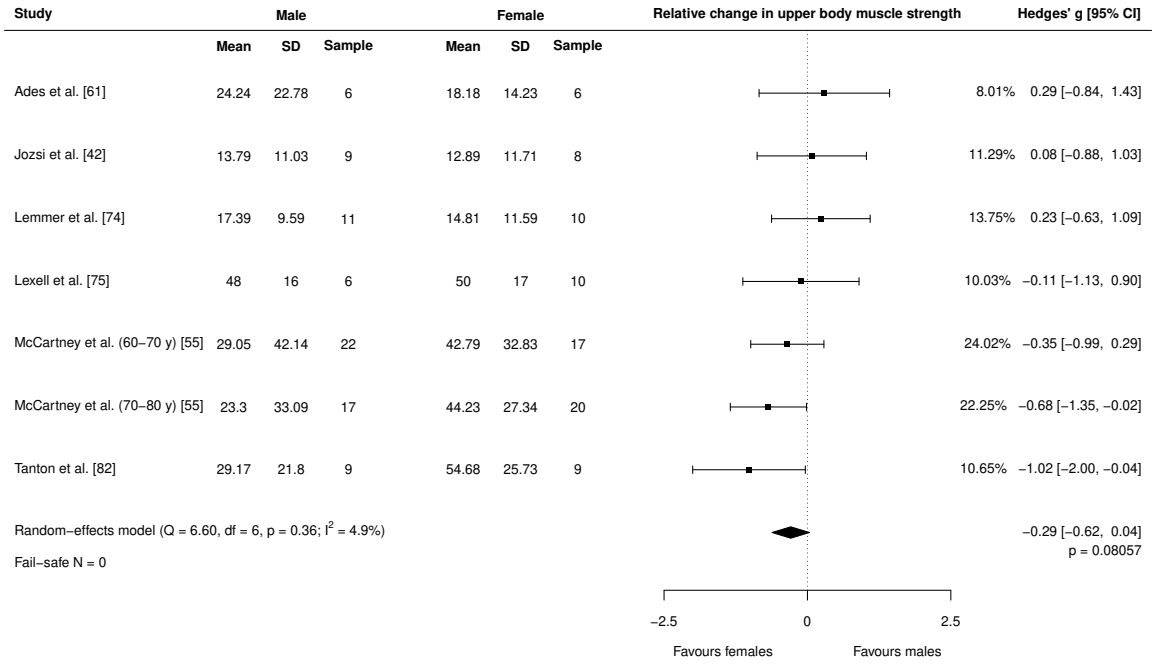
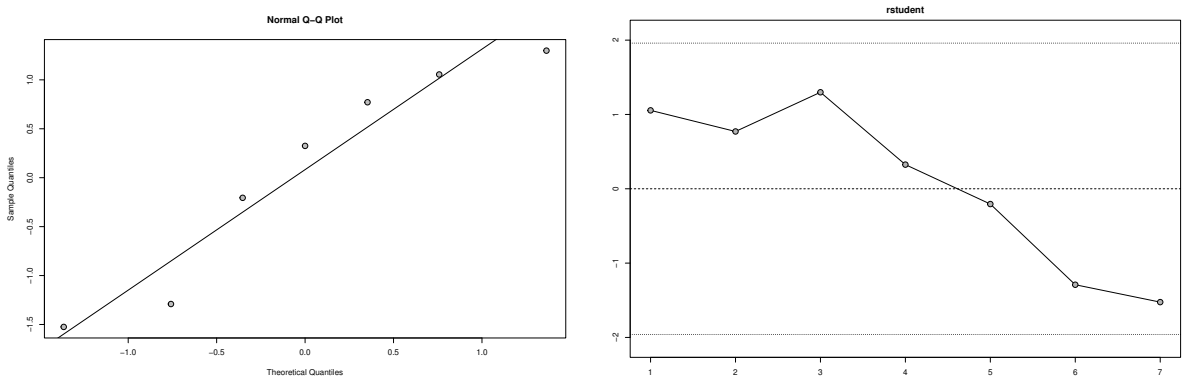


FIGURE 2.1: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom

2.5.1 Evaluation of model assumption and influential values

According to Figure 2.2a, the standardized deleted residuals of the model are normally distributed. The Shapiro-Wilk test to support the graphical analysis, being not significant ($W = 0.91, p = 0.37$). Moreover, the graph in Figure 2.2b shows that none of the standardized deleted residuals is particularly high, as they all fall within the range of $(-1.96, 1.96)$.



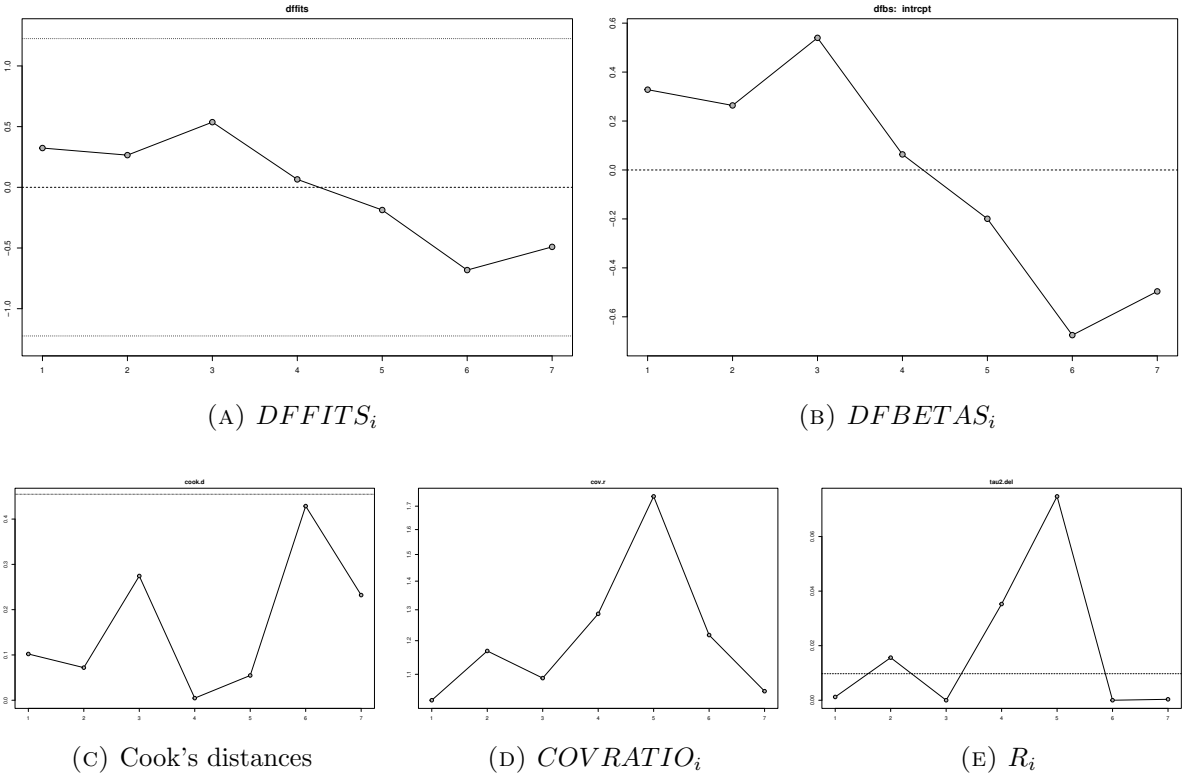
(A) Residuals Q-Q plot (B) Residuals vs study numbers

FIGURE 2.2: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention.

None of the studies appear to be influential. Indeed, all $DFFITS_i$ values (Figure 2.3a) are below their limit 1.22, Cook's distances (Figure 2.3c) are all below $\chi^2_{0.5,p+1} = 1.39$, and all $dfbs_i < |1|$ (Figure 2.3b). It is worth noting that the covariance ratio (Figure 2.3a) is always greater than 1, indicating that the uncertainty underlying the fixed effects β estimates increases when one of the included studies is excluded. The increase is particularly relevant when Study 5, corresponding to "McCartney et al (60-70 y)", is excluded because it is the largest among those included. Figure 2.3e shows the change in the estimate of heterogeneity (τ^2) when the i-th study is excluded. It can be observed that the removal of "McCartney et al. (60-70 y)" leads to a substantial inflation of τ^2."

2.5.2 Evaluation of risk of publication bias

Looking at the contour-enhanced funnel plot in Figure 2.4a, there seems to be a slight indication of publication bias, as one would expect a larger number of studies with negative and significant effect sizes. However, Egger's test for asymmetry of the funnel plot is not significant ($Z = 1.06, p = 0.29$). Figure 2.4b shows the Egger regression line calculated using standard error and sample variance as predictors.



(A) $DFFITS_i$ (B) $DFBETAS_i$

(C) Cook's distances (D) $COVRATIO_i$ (E) R_i

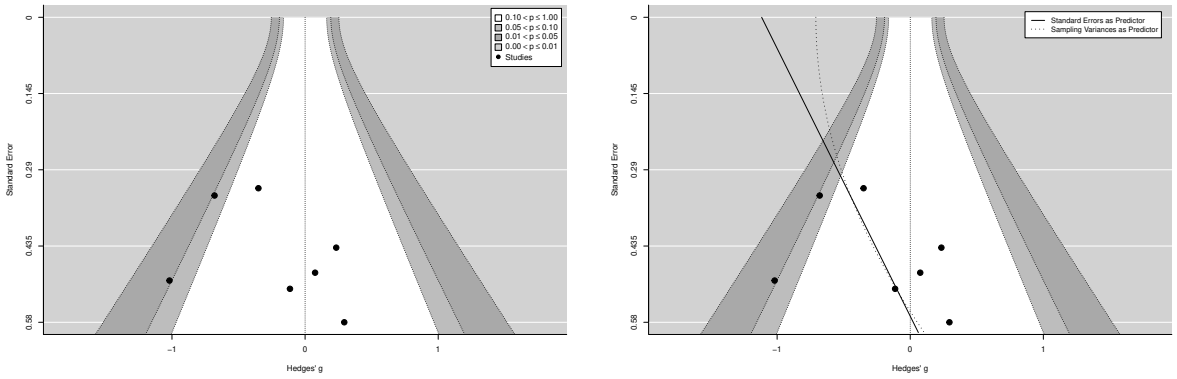
FIGURE 2.3: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention. The statistics are plotted against a study progressive number.

Figure 2.4c shows the contour-enhanced funnel plot where the number of missing trials was added to achieve symmetry using the trim and fill method. The results support the conclusion that there is a lack of studies with negative effect sizes and relatively high standard errors. It is also noteworthy that the sensitivity analysis for asymmetry performed with the trim and fill method yields $\hat{\beta} = -0.34$ with a 95% confidence interval of $(-0.65, -0.03)$, which slightly changes the conclusions of the meta-analysis. Finally, Rosenthal's fail-safe N calculation is equal to zero because the estimated effect is not statistically significant (Figure 2.1).

2.5.3 Comparison with more advanced methods

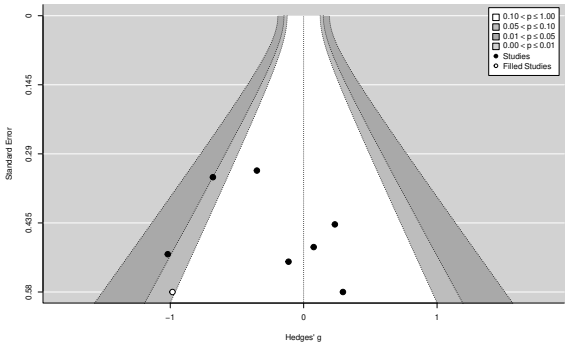
A comparison with the other methods proposed in Chapter 1 is also presented, both to understand the sensitivity of the β and τ^2 estimates and to assess the robustness of the inference performed.

As can be seen in Table 2.2, the point estimate of the intercept of the model is particularly robust to the choice of method used. In fact, the method proposed by DerSimonian & Laird (1986), the maximum likelihood estimation, the restricted maximum likelihood



(A) Contour enhanced funnel plot

(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.4: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention. The Egger regression 'functions' are reported for s_i and s_i^2.

estimation (REML), and Paule-Mandel's method yield similar results. This is not surprising since the estimator $\hat{\beta}$ is the same with exception of the weights used, which vary between methods based on τ^2. The percentage of heterogeneity in the model is very small, as indicated by previous analyses, and therefore has little effect on the β estimate. Similarly, the standard errors of $\hat{\beta}$ in the different models differ based on τ^2 and therefore also appear very similar.

The confidence intervals for β also yield the same conclusion, with similar endpoints and comparable width, except for the Hartung-Knapp-Sidik-Jonkman method (see Table 2.3).

It should also be noted that the p-values to assess the significance of the intercept test in the model are mostly consistent with the confidence intervals. Table 2.4 shows that the Hartung-Knapp-Sidik-Jonkman method indeed provides the highest p-value. Furthermore, it is interesting to note that the p-value calculated using the permutation test

TABLE 2.2: Point estimate and standard error of β of the RE model fitted to estimate the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	-0.29	0.17
ML	-0.30	0.16
REML	-0.29	0.17
PM	-0.29	0.17

TABLE 2.3: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β of the RE model fitted to estimate the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	LL	UL	Width
DL_Wald	-0.62	0.05	0.67
ML_Wald	-0.61	0.02	0.63
REML_Wald	-0.62	0.04	0.65
ML_prof_Wilks	-0.61	0.02	0.63
REML_prof_Wilks	-0.62	0.04	0.65
Hartung-Knapp-Sidik-Jonkman	-0.68	0.15	0.82
prof_Wilks_Skovgaard	-0.62	0.09	0.72

is similar to those calculated using maximum likelihood based methods, indicating that the conclusions are robust to deviations from the assumption of normality of the true study effects. Finally, consider that the p-value calculated using the Skovgaard's statistic is particularly high, indicating that the first-order approximation of the likelihood ratio test is not reliable due to the small number of included studies.

The point estimate of τ^2 is also very similar among the methods used, with comparable standard errors, see Table 2.5. This is likely due to the negligible amount of heterogeneity. However, it should be noted that although the estimate is close to the lower limit of the parameter space, the standard error is considerable.

TABLE 2.4: Significance test on the parameter β of the RE model fitted to estimate the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	p
DL	0.08
Hartung-Knapp-Sidik-Jonkman	0.21
ML_Wald	0.07
REML_Wald	0.08
ML_prof_Wilks	0.11
Skovgaard	0.27
Bartlett	0.17
Permutations	0.11

TABLE 2.5: Point estimate and standard error of τ^2 of the RE model fitted to estimate the difference between males and females in relative changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.02	0.12
ML	$< 10^{-4}$	0.09
REML	0.01	0.11
PM	0.02	0.12

2.6 Analysis of the absolute changes in upper body muscle strength

The RE model estimated to test the sex effect in the absolute change in upper body muscle strength is based on the same number of studies and participants as the model on the relative differences. The point estimate of the parameter of interest is $\hat{\beta} = 0.48$, with an associate 95% confidence interval of $(< 10^{-4}, 1.82)$, using the Viechtbauer (2007) method. Accordingly, the p-value for the test of the null hypothesis of no effect is $p = 0.02$. Therefore, the null hypothesis of no difference between males and females can be rejected.

The Cochran's Q test is not significant ($Q = 8.92$, $df = 6$, $p = 0.18$) and thus the hypothesis of a common true effect is not rejected. The percentage of variance explained by heterogeneity is $I^2 = 29.6\%$, indicating that the main source of variance is

due to within-study variability but that heterogeneity is considerably higher than in the previous model. Hence, considering the slightly higher point estimate of the I^2 statistics and the discussion about the Cochran's Q test, we prefer not to estimate the CE model. Moreover, the point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0.08$, with an associate 95% confidence interval of$(-0.18, 0.23)$, with the negative endpoint typically set equal to 0. The results presented so far were shown in the forest plot in Figure 2.5.

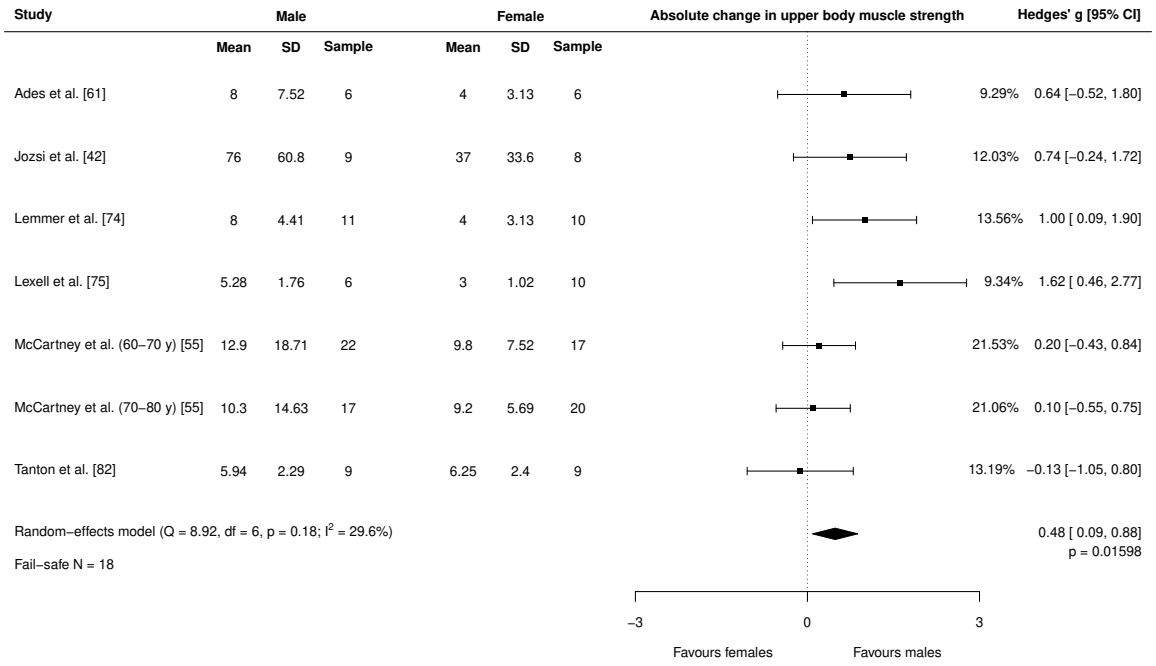
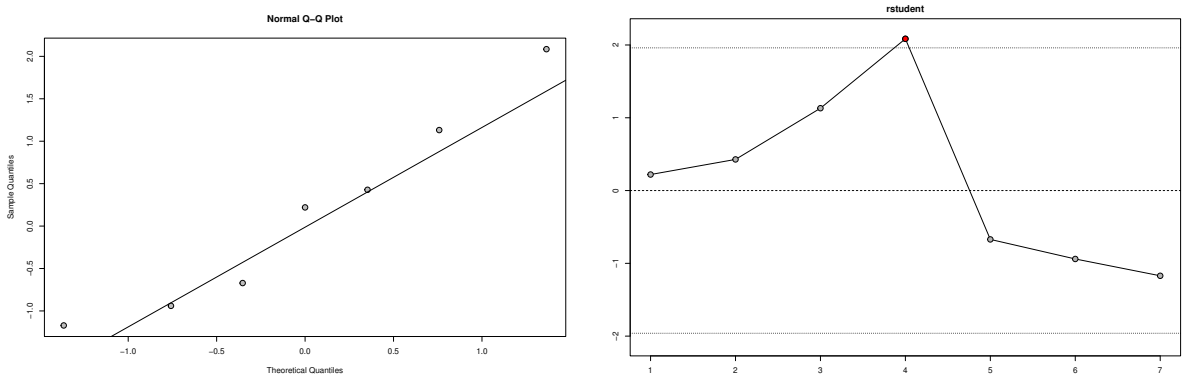


FIGURE 2.5: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom



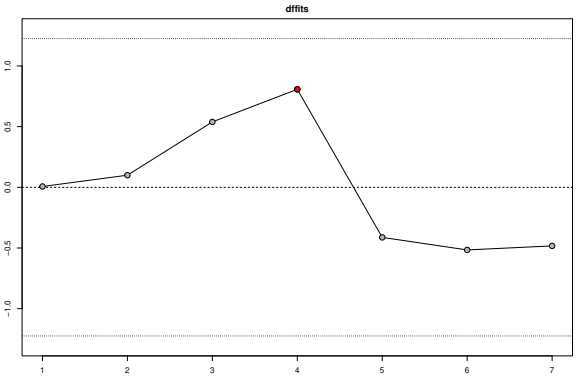
(A) Residuals Q-Q plot (B) residuals vs study numbers

FIGURE 2.6: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention.

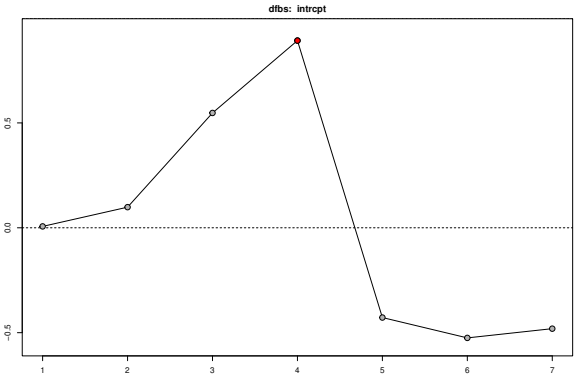
2.6.1 Evaluation of model assumption and influential values

According to Figure 2.6a, the standardized deleted residuals of the model are normally distributed. The Shapiro-Wilk test to support the graphical analysis, being not significant ($W = 0.94, p = 0.65$). Moreover, the graph in Figure 2.6b shows that only one standardized deleted residuals is above the limit 1.96. Under the normality hypothesis, the 5% of the residuals are expected to exceed those limits, and thus what is observed could be compatible with what is expected. In fact the 5% of 7 is 0.35 and if we had multiple sample of this size from the same normal distribution, we would expect some of them to have any value above 1.96 or below -1.96 and some of them with one value outsome of the threeshold.

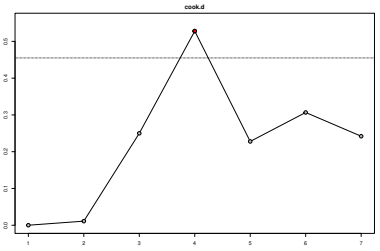
The study of "Lexell et al. [75]" appears to be an influential value. In fact, in addition to the standardized deleted residual above the limit threshold, also its Cook's distance (Figure 2.7c) seems particularly elevated. The influence analysis showed that its removal decrease the estimate of β (Figure 2.7b) and the amount of heterogeneity (Figures 2.7d and 2.7e). However, even removing the influential study, the model intercept remains significantly different from zero ($\hat{\beta} = 0.34, \quad se(\hat{\beta}) = 0.17, \quad p = 0.05$), while the point estimate of τ^2 by the method of moments becomes exactly zero. For this reason and given the limited sample size, it is not considered interesting to perform meta-regression analyses in order to explain the heterogeneity of the effect, as it is mainly attributable to unique study characteristics of "Lexell et al. [75]".



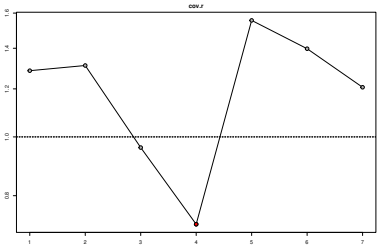
(A) Difference in fits value ($DFFITS_i$)



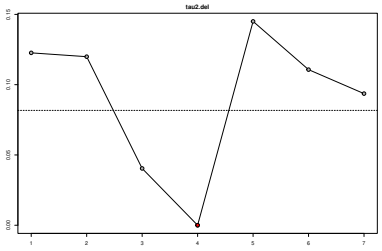
(B) $DFBETAS_i$



(C) Cook's distances



(D) $COVRATIO_i$



(E) R_i

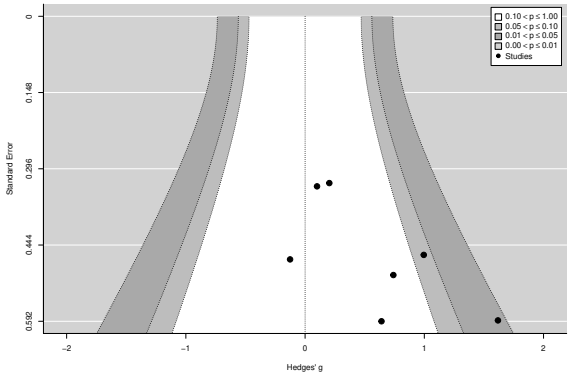
FIGURE 2.7: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention. The statistics are plotted against a study progressive number.

2.6.2 Evaluation of risk of publication bias

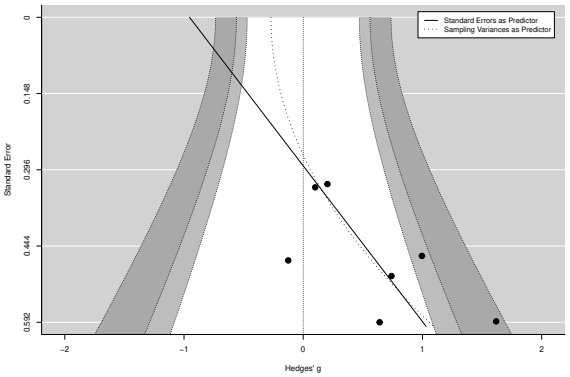
Looking at the contour-enhanced funnel plot in Figure 2.8a, there seems to be a slight indication of publication bias, as one would expect a larger number of studies with negative not significant effect sizes and the Egger's test for asymmetry is significant both using variance ($Z = 2.1, p = 0.04$) or the standard error ($Z = 2.0, p = 0.04$) as a predictor(Figure 2.8b). However, the trim and fill methods suggest that the estimated number of missing study on the left is zero (Figure 2.8c). Finally,the Rosenthal's fail-safe N calculation suggested that 18 studies with nil effects are needed to have a non-significan estimate of β (Figure 2.5).

2.6.3 Comparison with more advanced methods

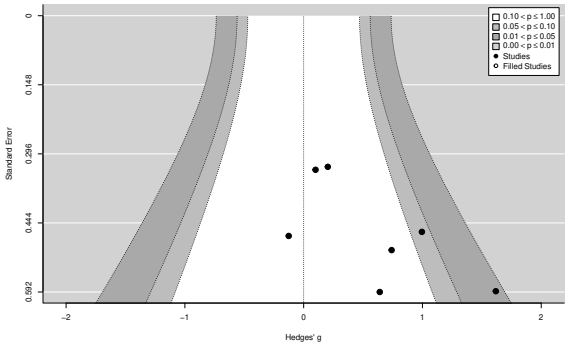
The discussion of the robustness of the point estimate of β and its standard error with respect to the estimation method is similar to that in Section 2.5.3. The values are given in Table 2.6.



(A) Contour enhanced funnel plot



(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.8: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention.

TABLE 2.6: Point estimate and standard error of β of the RE model fitted to estimate the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML) and the restricted maximum likelihood method (REML) and the Paule-Mandel (PM) est.

Estimator	Estimate	Standard Error
DL	0.49	0.21
ML	0.45	0.18
REML	0.48	0.20
PM	0.49	0.21

It should also be noted that the p values used to assess the significance of the intercept test are all consistent in the model (Table 2.8), and the only one that is not significant is obtained using the Skovgaards's statistic, suggesting that the first-order approximation of the likelihood ratio test is not reliable because of the small number of studies included. However, the discrepancy between Skovgaard's p-value and the other p-values is much smaller compared to the previous example.

TABLE 2.7: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β of the RE model fitted to estimate the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	LL	UL	Width
DL_Wald	0.09	0.89	0.81
ML_Wald	0.11	0.80	0.69
REML_Wald	0.09	0.88	0.79
ML_prof_Wilks	0.11	0.80	0.69
REML_prof_Wilks	0.09	0.88	0.79
Hartung-Knapp-Sidik-Jonkman	0.04	0.99	0.94
prof_Wilks_Skovgaard	0.09	0.95	0.86

TABLE 2.8: Significance test on the parameter β of the RE model fitted to estimate the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	p
DL	0.02
Hartung-Knapp-Sidik-Jonkman	0.03
ML_Wald	0.01
REML_Wald	0.02
ML_prof_Wilks	0.02
Skovgaard	0.05
Bartlett	0.04
Permutations	0.01

Also, the discussion of the robustness of the estimator of τ^2 is analogous to Section 2.5.3, but for none of the methods is the point estimator exactly equal to zero, see Table 2.9.

TABLE 2.9: Point estimate and standard error of τ^2 of the RE model fitted to estimate the difference between males and females in absolute changes in upper body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.09	0.17
ML	0.02	0.11
REML	0.08	0.16
PM	0.11	0.18

2.7 Analysis of the relative changes in lower body muscle strength

The RE model estimated on the relative effect measures of differences between males and females in the relative change in lower body muscle strength is based on the results of 34 studies and 1196 individuals, of whom 630 are female. The point estimate of the parameter of interest is $\hat{\beta} = -0.21$, with an associate 95% confidence interval of$(-0.33, -0.10)$. Accordingly, the p-value for the test of the null hypothesis of no effect is close to 0, $p < 0.001$. Therefore, the null hypothesis that there is no difference between males and females in lower body strength gain can be rejected when the change is expressed relative to baseline.

The Cochran's Q test is not significant ($Q = 24.59, \quad df = 34, \quad p = 0.88$) and thus the hypothesis of a common true effect is not rejected. The I^2 statistic is equal to zero, confirming the absence of heterogeneity. The point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0$ and the associated standard error (obtained with likelihood theory) is 0.02, indicating that the uncertainty behind the estimate of the heterogeneity is small. The forest plot is presented in the Figure 2.9.

2.7.1 Evaluation of model assumption and influential values

According to the Figure 2.10a, the standardized deleted residuals of the model appears to follow a normal distribution. The Shapiro-Wilk test to support the graphical analysis, being not significant ($W = 0.95, p = 0.14$). Moreover, the graph in Figure 2.10b shows that none of the standardized deleted residuals is particularly high, as they all fall within the range of $(-1.96, 1.96)$.

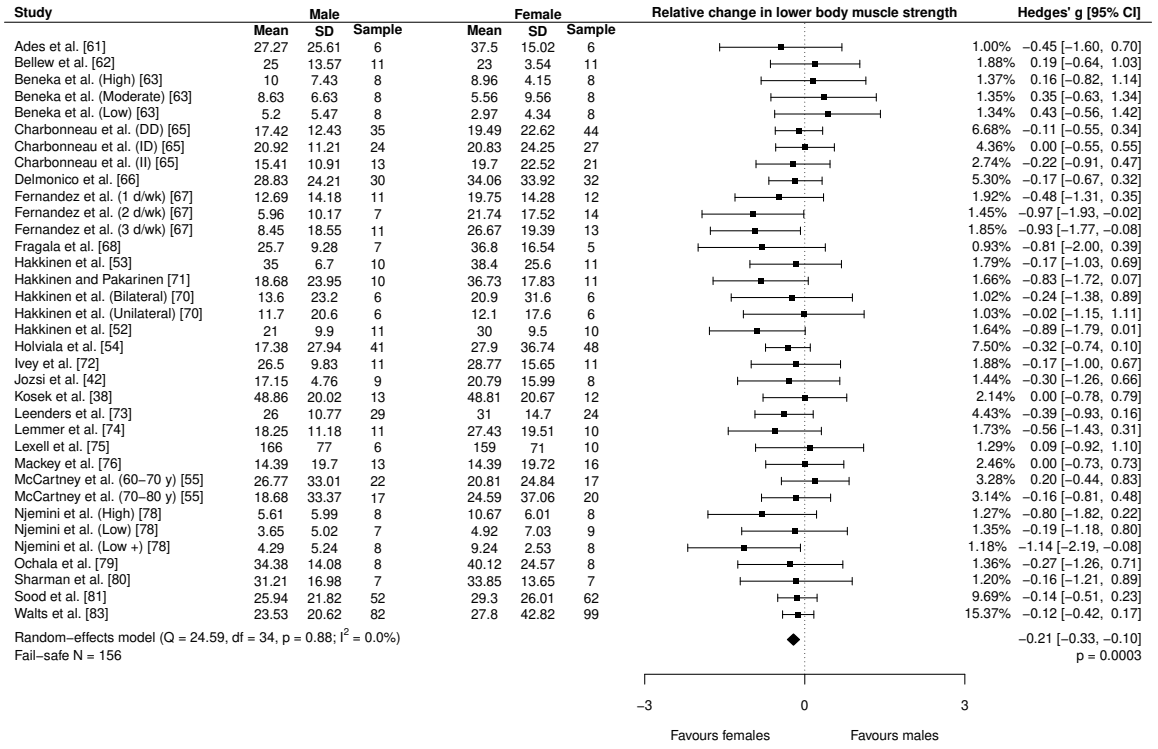
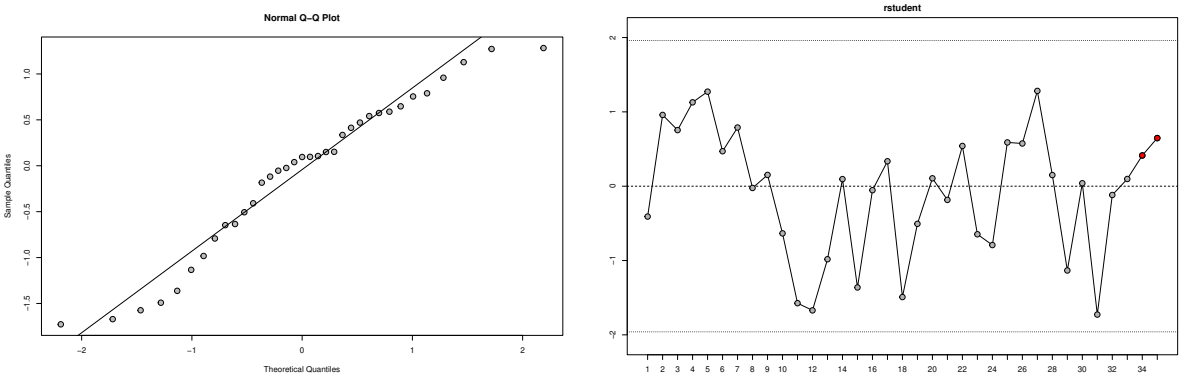


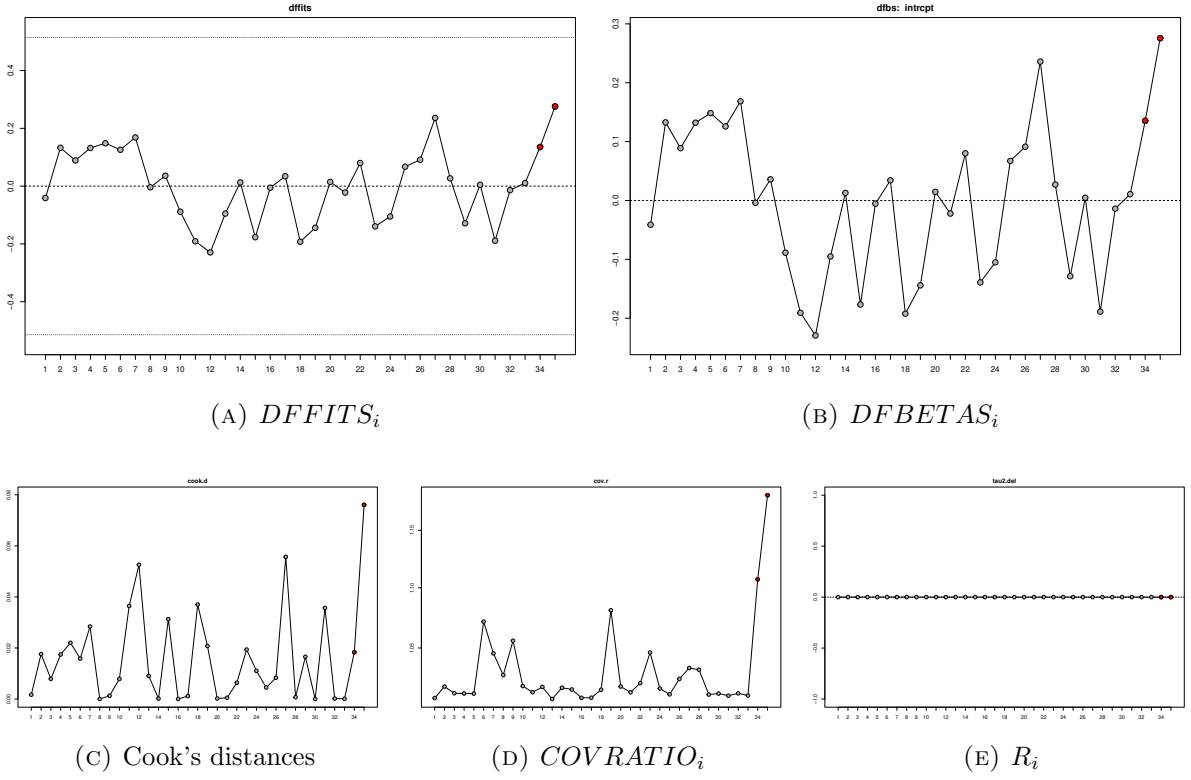
FIGURE 2.9: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom



(A) Residuals Q-Q plot (B) residuals vs study numbers

FIGURE 2.10: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention.

Two effect measures stand out as influential values, namely "Sood et al. [81]" and "Walts et al. [83]". Although these do not influence the values predicted by the model (Figure 2.11a) or even the coefficients (Figure 2.11b), the $COVRATIO_i$ is particularly greater than one if these studies are removed. It is not complex to explain the reason for this, since in the RE model the weights of the two studies are 9.69% and 15.37%, respectively. In fact, the two studies have very low within-study variance, and considering that the τ^2 estimate is zero, the weights coincide with those of the CE model. Thus, the two studies are "outliers" in terms of the precision of the effect measure, and the reason is related to the sample size, see Figure 2.9).



(A) $DFFITS_i$ (B) $DFBETAS_i$

(C) Cook's distances (D) $COVRATIO_i$ (E) R_i

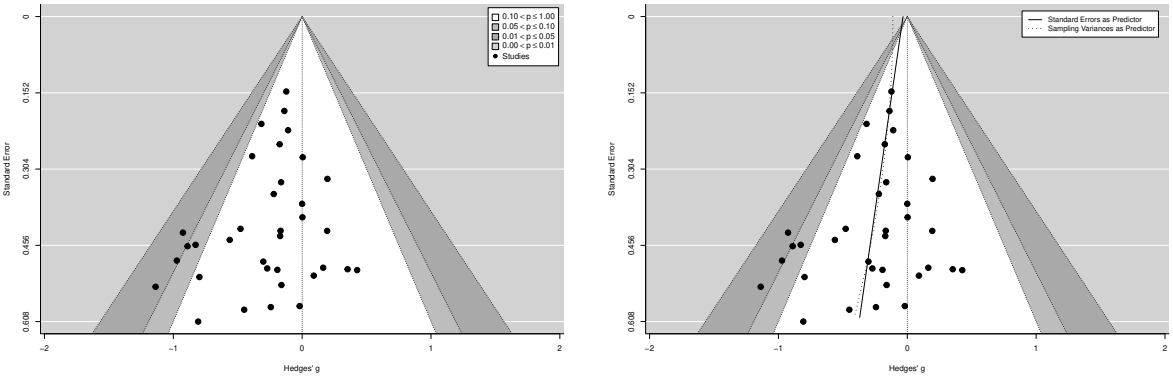
FIGURE 2.11: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention. The statistics are plotted against a study progressive number.

While there is no particular reason to remove the two studies, the β estimate obtained their removal is $\hat{\beta} = -0.24, \quad se(\hat{\beta})) = 0.07, \quad p < 0.001$, not changing the inference.

2.7.2 Evaluation of risk of publication bias

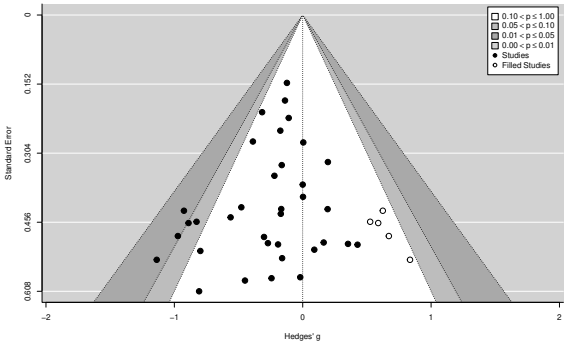
Looking at the contour-enhanced funnel plot in Figure 2.12a, there seems to be indication of publication bias, as one would expect a larger number of studies with positive and not significant effect sizes. However, Egger's test for asymmetry of the funnel plot

is not significant neither using the standard error ($Z = -1.29, p = 0.20$) nor the variance ($Z = -1.34, p = 0.18$) as regressors. Figure 2.12b shows the Egger regression line calculated using standard error and sample variance as predictors.



(A) Contour enhanced funnel plot

(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.12: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention.

Figure 2.12c shows the contour-enhanced funnel plot where the number of missing trials was added to achieve symmetry using the trim and fill method. The results support the idea that studies with positive not significant effect size are missing, indicating the lack of 5 studies. The sensitivity analysis performed with the trim and fill method yields $\hat{\beta} = -0.1507$ with a 95% confidence interval of $(-0.26, -0.04)$, which sligthly change the inferential conclusion in term of magnitude but not direction. Finally, Rosenthal's fail-safe N calculation is equal to 156.

2.7.3 Comparison with more advanced methods

The point estimate of β and its standard error are identical for DerSimonian & Laird (1986), maximum likelihood, restricted maximum likelihood, and Paul-Mandel methods.

This is not surprising given that all methods use a plug-in of $\hat{\tau^2}$ in estimating β, and $\hat{\tau^2} = 0$ for all of them. Accordingly, the confidence intervals for β are all very similar with almost the same amplitude (Table 2.10) and the p-value of all evaluated methods is close to zero at the second decimal place.

TABLE 2.10: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β of the RE model fitted to estimate the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	LL	UL	Width
DL_Wald	-0.33	-0.10	0.23
ML_Wald	-0.33	-0.10	0.23
REML_Wald	-0.33	-0.10	0.23
ML_prof_Wilks	-0.33	-0.10	0.23
REML_prof_Wilks	-0.33	-0.10	0.23
Hartung-Knapp-Sidik-Jonkman	-0.39	-0.08	0.30
prof_Wilks_Skovgaard	-0.33	-0.10	0.23

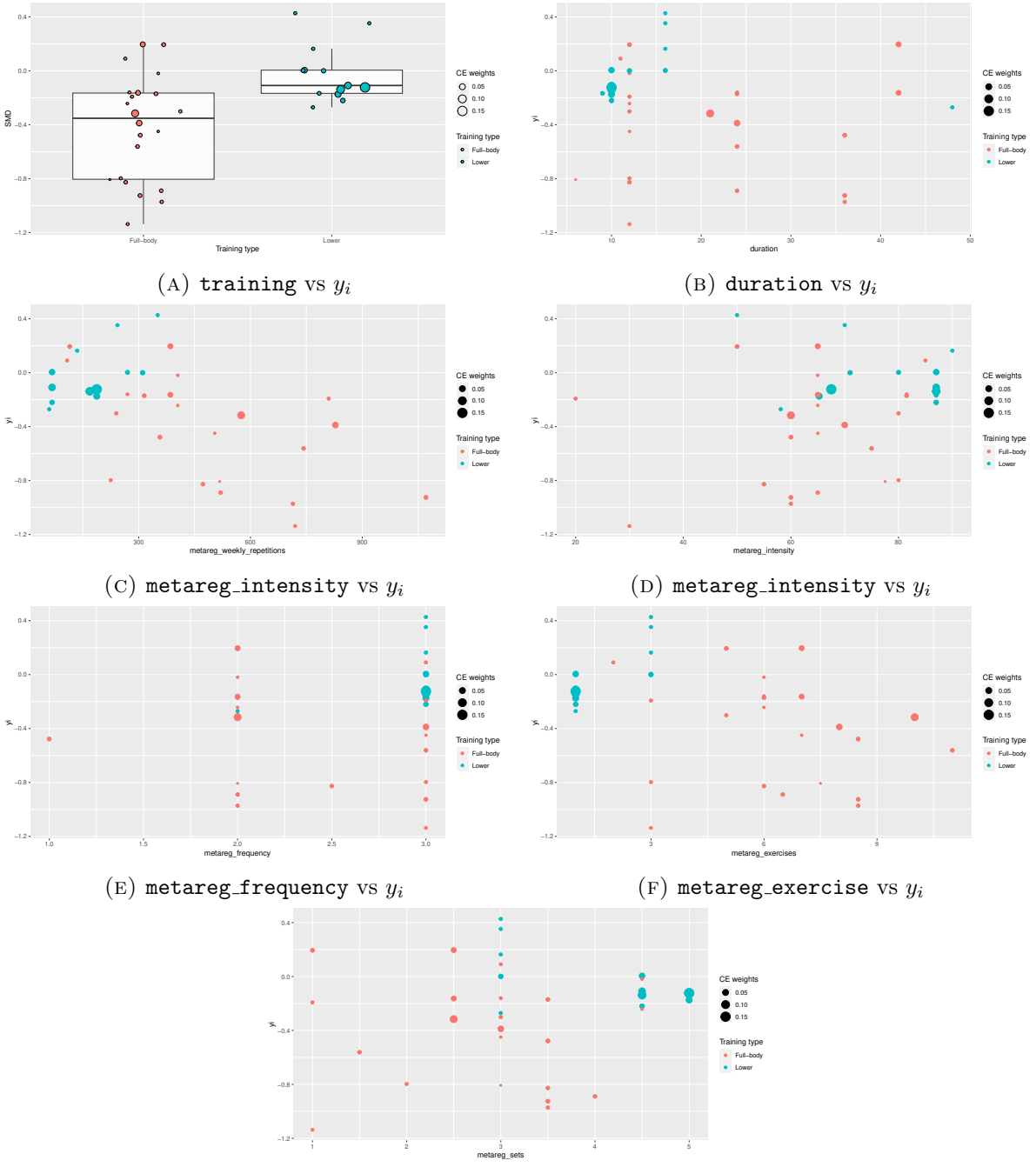
TABLE 2.11: Point estimate and standard error of τ^2 of the RE model fitted to estimate the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	$< 10^{-4}$	0.03
ML	$< 10^{-4}$	0.02
REML	$< 10^{-4}$	0.02
PM	$< 10^{-4}$	0.03

2.7.4 Meta-regression

The exploratory bivariate scatter plots for each covariates are presented in Figure 2.13. The variable `metareg_exercise` was excluded from the analysis as it measures the concept of volume similarly to `metareg_weekly_repetitions` but with less precision. Furthermore, it was observed that the variable `training` consistently showed higher values in full-body programs compared to lower-body programs, likely due to the inclusion of a greater total number of exercises. In fact, in a full-body program, the number of muscles trained is higher. However, `training` was included in the initial metaregression

model since there was a suspected interaction with `metareg_intensity`.The model selection was performed using backward regression based on the p-value criterion, without correcting for multiple comparisons.



(A) `training` vs y_i

(B) `duration` vs y_i

(C) `metareg_intensity` vs y_i

(D) `metareg_intensity` vs y_i

(E) `metareg_frequency` vs y_i

(F) `metareg_exercise` vs y_i

(G) `metareg_sets` vs y_i

FIGURE 2.13: Bivariate exploratory analysis of the relationship of the most interesting covariates with respect to the measure of effect of the single study in the analysis of relative changes in upper body strength. The size of the points in the box is proportional to the weights of the CE model.

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 \text{training} + \beta_2 \text{metareg\_weekly\_repetitions} + \\ &+ \beta_3 \text{training} \times \text{metareg\_weekly\_repetitions} + \beta_4 \text{duration} + \\ &+ \beta_5 \text{metareg\_intensity} + \beta_6 \text{metareg\_frequency} + \delta_i + \epsilon_i. \end{aligned}$$

Except for `training`, which is a binary variable, all the other variables are continuous. Then the backward selection led to the identification of a considerably simpler model with the following linear predictor:

$$Y_i = \beta_0 + \beta_1 \text{metareg\_weekly\_repetitions} + \delta_i + \epsilon_i.$$

The results of the selection are available in the code. The estimated value of τ^2 is $\hat{\tau}^2 = 0$, and the p-value of the Cochran's Q test is particularly high ($Q = 15.79, \quad df = 33, \quad p = 0.99$), indicating that the hypothesis homogeneity of the true effects cannot be rejected. The estimated model yields a coefficient of $\hat{\beta}_0 = 0.0194$ with a Wald 95% confidence interval of $(-0.17, 0.21)$, and a slope of $\hat{\beta}_1 = -0.07 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.12 \times 10^{-2}, -0.02 \times 10^{-2})$. Hence, the slope is significantly different from zero. The meta-regression line is represented in Figure 2.14. As the weekly volume increases, the effect size becomes increasingly negative, favoring females. In other words, the difference between males and females in relative changes in muscle strength is a function of volume and it appears to be more pronounced and favorable for women at higher training volumes. This moderator is also capable of explaining the heterogeneity present in the intercept only meta-analysis model.

Furthermore, the model does not exhibit significant deviations from the assumption of normality. The Shapiro-Wilk test conducted on the standardized deleted residuals is not significant ($W = 0.95, p = 0.14$). None of the studies appears to be influential. However, since the estimate of $\hat{\tau^2} = 0$, the model reduces to the CE model, and therefore the weights are highly correlated with the study's sample size. The studies "Fernandez et al. (3 d/wk)," "Leenders et al.," and "Walts et al." are considered suspicious as their removal causes an inflation of the $COVRATIO_i$ above the value of 1.25. As for the last two studies, it is clear that the motivation is related to the weight $1/\sigma_i \times 100$. Specifically, "Leenders et al." has a weight of 4.43%, and "Walts et al." has a weight of 15.37%, ranking as the third and first highest weights, respectively. However, the removal of these three studies does not change the estimate of τ^2 at the second decimal place or the value of the slope. In fact, we have $\hat{\beta}_1 = -0.08 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.14 \times 10^{-2}, -0.02 \times 10^{-2})$, which does not alter the inferential

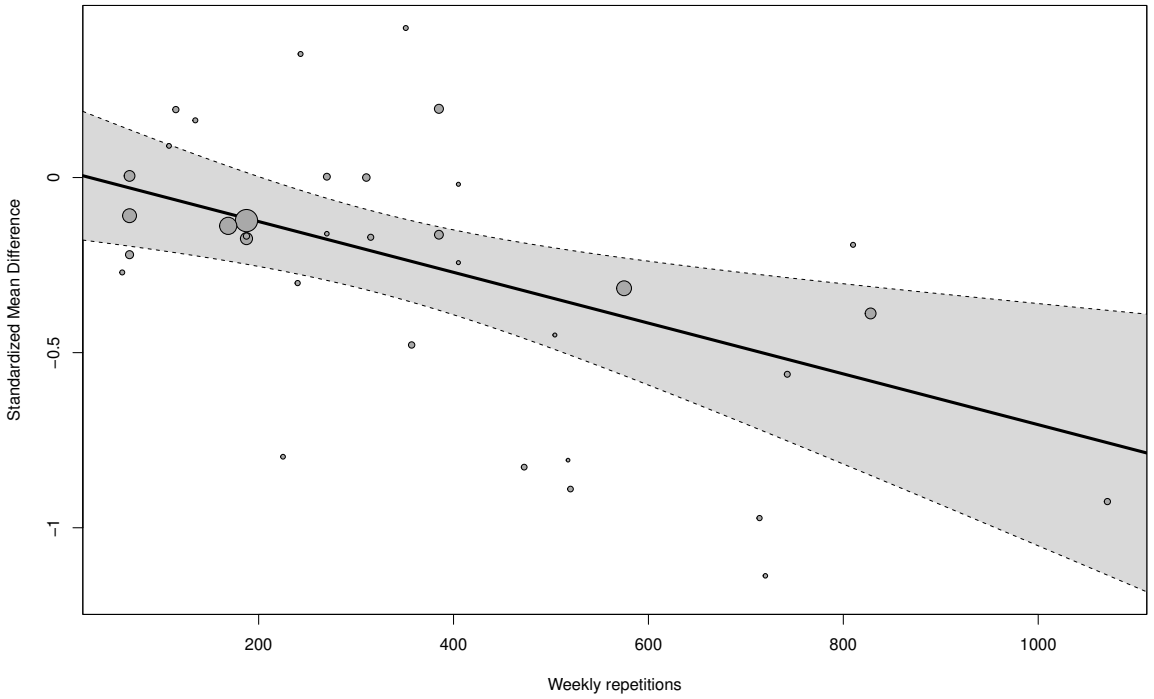


FIGURE 2.14: Estimated model variation of the effect measure y_i as a function of the number of weekly repetitions in the dataset of the relative changes in upper body muscle strength. The size of the circle is proportional to $1/(s_i^2 + \hat{\tau}^2)$

.

conclusions.

Tables 2.12, 2.13, 2.14, 2.15 report respectively the estimates of the β coefficients, their confidence intervals, the significance test for each coefficient and the estimate of τ^2 using the methods employed in all previous analyses. It should be noted that the inferential conclusions remain unchanged.

2.8 Analysis of the absolute changes in lower body muscle strength

The RE model estimated to test the sex effect in the absolute change in lower body muscle strength is based on the same number of studies and participants as the model on the relative differences. The point estimate of the parameter of interest is $\hat{\beta} = 0.33$, with an associate 95% confidence interval of $(0.19, 0.47)$. Accordingly, the p-value for the test of the null hypothesis of no effect is $p < .01 \times 10^{-2}$. Therefore, the null hypothesis of no difference between males and females can be rejected.

TABLE 2.12: Point estimate and standard error of β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Coefficient	Estimator	Estimate	Standard Error
β_0	DL	0.02	0.10
β_0	ML	0.02	0.10
β_0	REML	0.02	0.10
β_0	PM	0.02	0.10
β_1	DL	-0.07×10^{-2}	0.02×10^{-2}
β_1	ML	-0.07×10^{-2}	0.02×10^{-2}
β_1	REML	-0.07×10^{-2}	0.02×10^{-2}
β_1	PM	-0.07×10^{-2}	0.02×10^{-2}

TABLE 2.13: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β_0 and β_1 in the meta-regression model fitted on the relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Coefficient	Estimator	LL	UL	Width
β_0	DL_Wald	-0.17	0.21	0.38
β_0	ML_Wald	-0.17	0.21	0.38
β_0	REML_Wald	-0.17	0.21	0.38
β_0	ML_prof_Wilks	-0.17	0.21	0.38
β_0	REML_prof_Wilks	-0.17	0.21	0.38
β_0	Hartung-Knapp-Sidik-Jonkman	-0.21	0.28	0.49
β_0	prof_Wilks_Skovgaard	-0.17	0.21	0.38
β_1	DL_Wald	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}
β_1	ML_Wald	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}
β_1	REML_Wald	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}
β_1	ML_prof_Wilks	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}
β_1	REML_prof_Wilks	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}
β_1	Hartung-Knapp-Sidik-Jonkman	-0.14×10^{-2}	-0.02×10^{-2}	0.11×10^{-2}
β_1	prof_Wilks_Skovgaard	-0.12×10^{-2}	-0.02×10^{-2}	0.10×10^{-2}

TABLE 2.14: Significance test on the parameter β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	β_0	β_1
DL	0.84	0.30×10^{-2}
Hartung-Knapp-Sidik-Jonkman	0.77	0.76×10^{-2}
ML_Wald	0.84	0.30×10^{-2}
REML_Wald	0.84	0.30×10^{-2}
ML_prof_Wilks	0.84	0.30×10^{-2}
Skovgaard	0.66	0.19×10^{-2}
Bartlett	0.85	0.42×10^{-2}
Permutations	0.35	$< 10^{-4}$

TABLE 2.15: Point estimate and standard error of τ^2 in the meta-regression mixed effect model fitted to estimate on relative changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	$< 10^{-4}$	0.03
ML	$< 10^{-4}$	0.02
REML	$< 10^{-4}$	0.02
PM	$< 10^{-4}$	0.03

The Cochran's Q test is not significant ($Q = 47.45, \quad df = 34, \quad p = 0.06$) and thus the hypothesis of a common true effect is not rejected. The percentage of variance explained by heterogeneity is $I^2 = 18.5\%$, indicating that the main source of variance is due to within-study variability but that heterogeneity is considerably higher than in the previous model.

Moreover, the point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0.03$, and a maximum likelihood standard error $se(\hat{\tau^2}) = 0.04$. The results presented so far were shown in the forest plot in Figure 2.15.

2.8.1 Evaluation of model assumption and influential values

According to the Figure 2.16a, the standardized deleted residuals of the model are normally distributed. The Shapiro-Wilk test to support the graphical analysis (which is

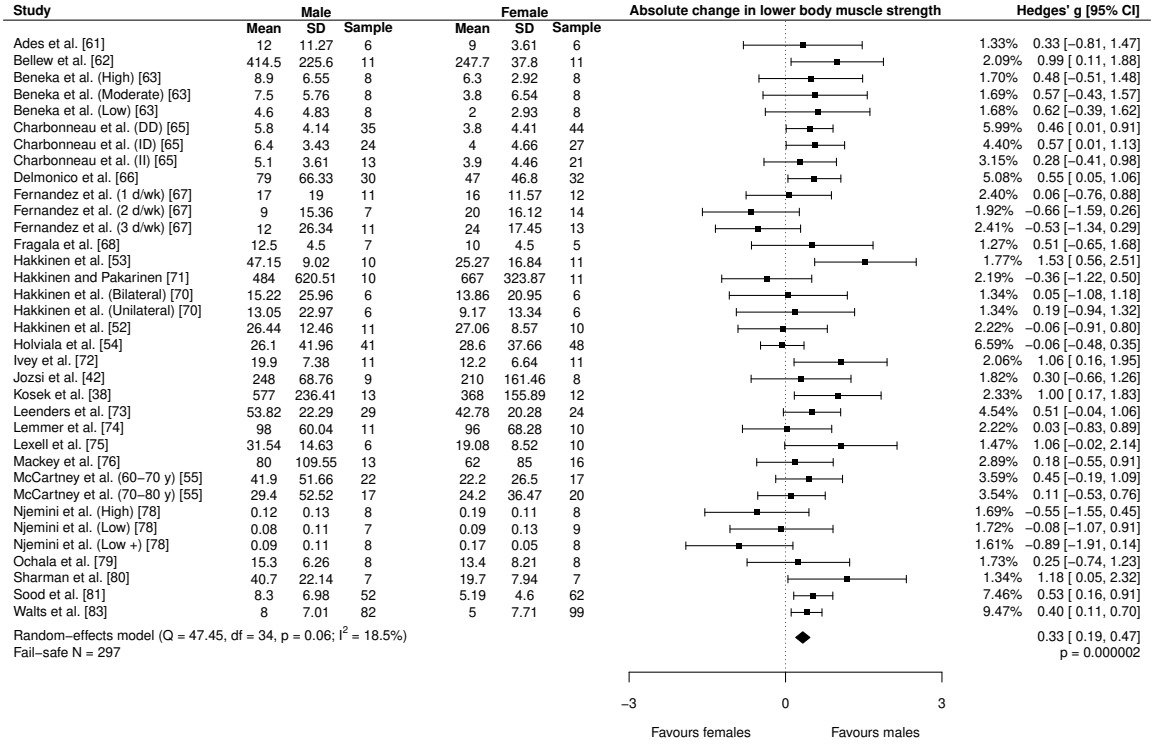
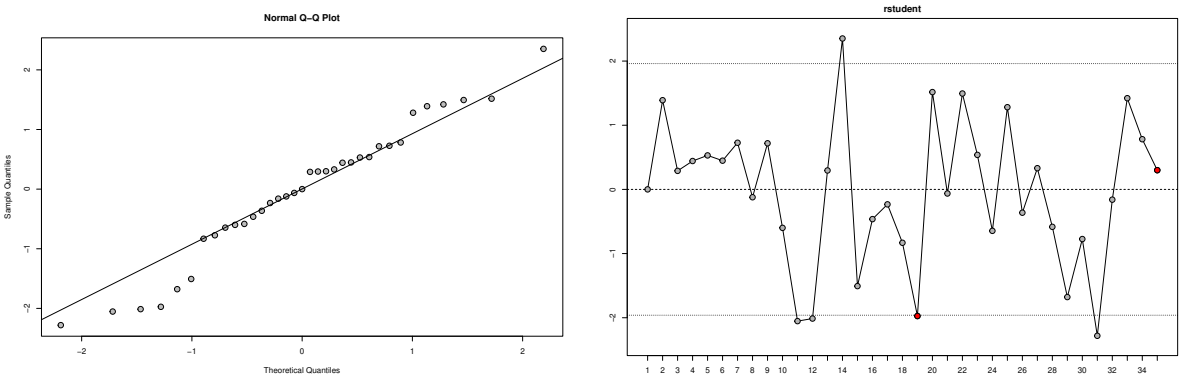


FIGURE 2.15: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom

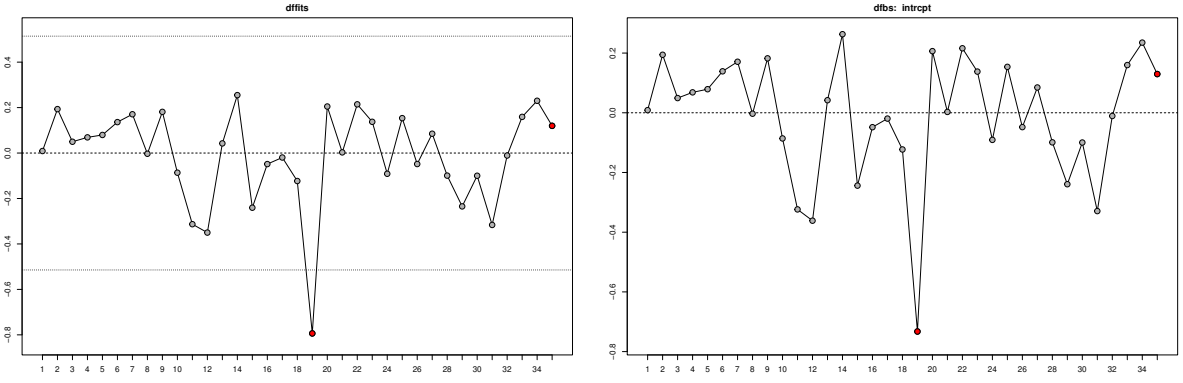
rather limited due to the small number of studies), being not significant ($W = 0.97, p = 0.41$). The graph in Figure 2.16b shows that four standardized deleted residuals are above 1.96 or below −1.96. This could cast doubt on the hypothesis of normaility as one would expect at most two points beyond the indicated limits, however it is not considered particularly serious given the result of the formal test.

The study of "Holviala et al. [54]" appears to be an influential value, particularly for the heterogeneity estimation. In fact, if it were removed, the point estimate of τ^2 would be equal to zero (Figures 2.17e and 2.17d). It also affects the estimation of beta, in fact if removed the point estimate is $\hat{\beta} = 0.37\,(se(\hat{\beta}) = 0.06)$, see Figure 2.17b.

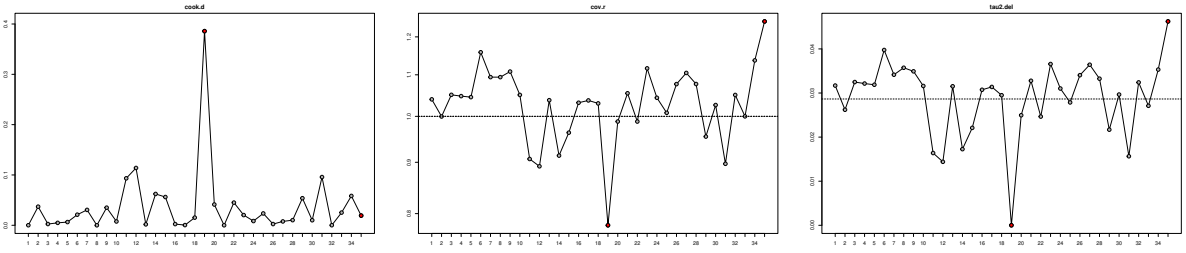


(A) Residuals Q-Q plot (B) residuals vs study numbers

FIGURE 2.16: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention.



(A) $DFFITS_i$ (B) $DFBETAS_i$

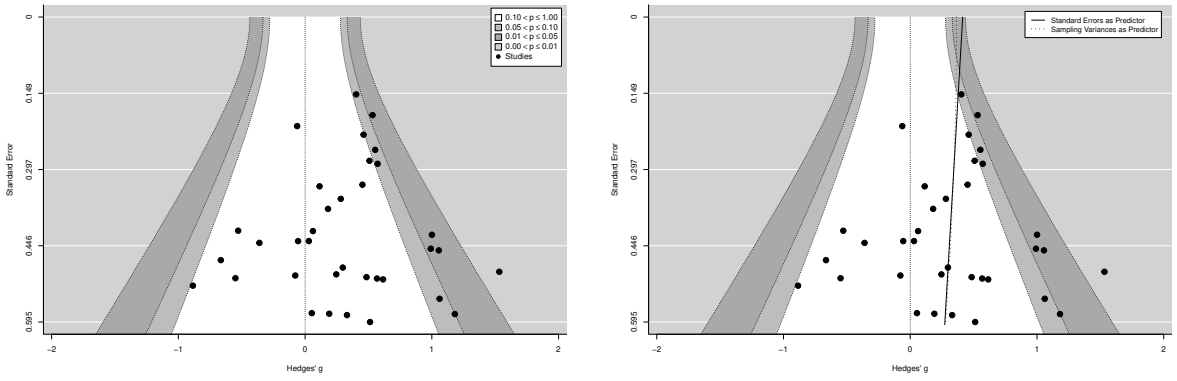


(C) Cook's distances (D) $COVRATIO_i$ (E) R_i

FIGURE 2.17: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention. The statistics are plotted against a study progressive number.

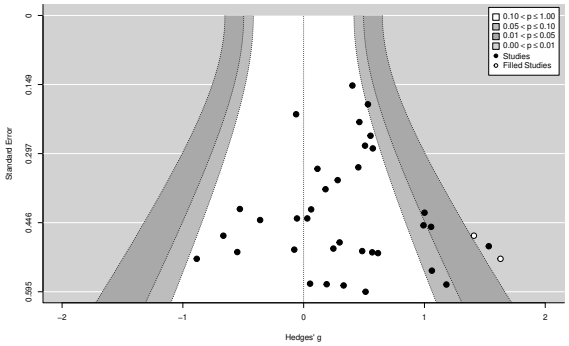
2.8.2 Evaluation of risk of publication bias

Looking at the contour-enhanced funnel plot in Figure 2.18a, there seems to be no lackness of points, indicating no suspect for systematica bias. The Egger's test for asymmetry is not significant both using variance ($Z = -0.44, \quad p = 0.66$) or the standard error ($Z = -0.34, p = 0.74$) as a predictor (Figure 2.18b). However, the trim and fill methods suggest that the estimated number of missing study on the left is 2 (Figure 2.18c) and with a significant and large effect size. Finally,the Rosenthal's fail-safe N calculation suggested that 297 studies with nil effects are needed to have a non-significant estimate of β (Figure 2.15).



(A) Contour enhanced funnel plot

(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.18: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention.

2.8.3 Comparison with more advanced methods

Considering that the results are absolutely similar to those obtained in Section 2.7.3, for simplicity the discussion is omitted. However, for the interested reader, they are

reported in the code on lines 1059-1099.

2.8.4 Meta-regression

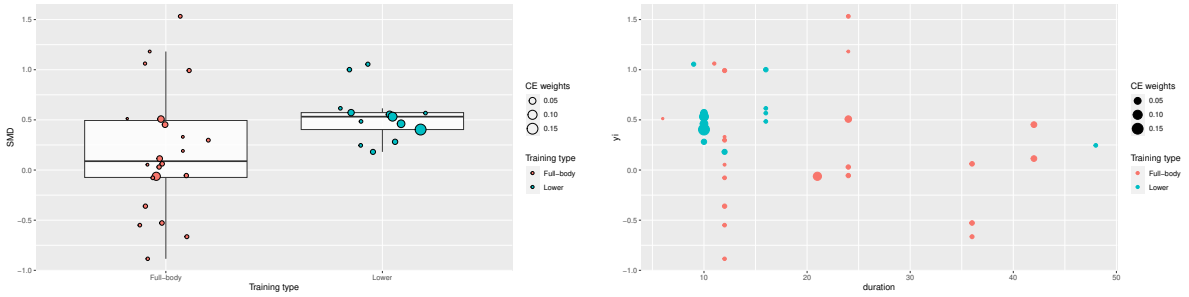
The exploratory bivariate scatter plots are presented in Figure 2.19. From their observation and considering what has been discussed in Section 2.7.4 the following initial meta-regression mixed effect model (of the type described in the equation 1.40) was proposed:

$$\begin{aligned} Y_i =& \beta_0 + \beta_1 \text{training} + \beta_2 \text{metareg\_weekly\_repetitions} + \\ &+ \beta_3 \text{training} \times \text{metareg\_weekly\_repetitions} + \beta_4 \text{duration} + \\ &+ \beta_5 \text{metareg\_intensity} + \beta_6 \text{metareg\_frequency} + \delta_i + \epsilon_i, \end{aligned}$$

where all the variables are continous, except for `training` which is a dummy.
Then the backward selection led to the identification of a considerably simpler model with the following linear predictor:

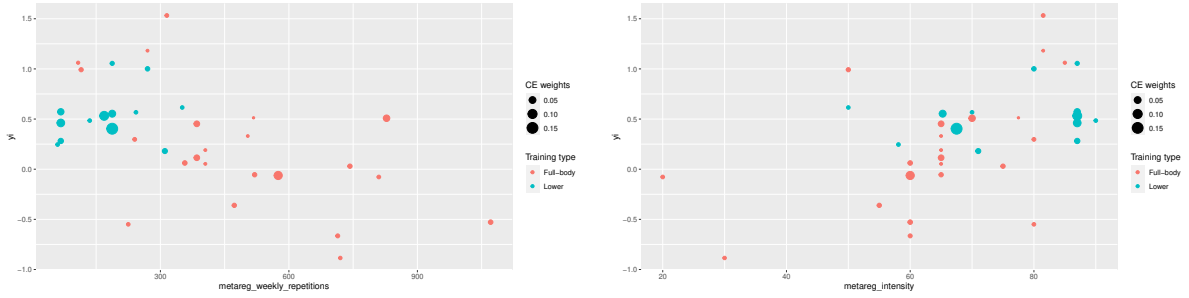
$$Y_i = \beta_0 + \beta_1 \text{metareg\_weekly\_repetitions} + \delta_i + \epsilon_i.$$

The results of the selection are available in the code. The estimated value of τ^2 is $\hat{\tau}^2 = 0$, and the p-value of the Cochran's Q test is particularly high ($Q = 33.24, \quad df = 33, \quad p = 0.46$), indicating that the hypothesis homogeneity of the true effects cannot be rejected. The estimated model yields a coefficient of $\hat{\beta}_0 = 0.64$ with a Wald 95% confidence interval of $(0.45, 0.83)$, and a slope of $\hat{\beta}_1 = -0.09 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.14 \times 10^{-2}, -0.04 \times 10^{-2})$. Hence both, the intercept β_0 and the slope β_1 are significantly different from zero. The meta-regression line is represented in Figure 2.20, where it is evident that as the weekly volume increases, the effect size becomes increasingly negative, favoring females. The moderator is also capable of explaining the heterogeneity present in the intercept only meta-analysis model.
Furthermore, the model does not exhibit significant deviations from the assumption of normality. The Shapiro-Wilk test conducted on the standardized deleted residuals is not significant ($W = 0.97, p = 0.41$). The study "Leenders et al." seems to be an influential value, in fact its standardized deleted residuals is above 2, the dffits value is above its threeshold of $3 * \sqrt{(p/(k-p)} = 0.74$ and the Cook's distance is greater than $\chi_{0.5,p+1} = 2.37$. Moreover, the $COVRATIO_i$ for "Leenders et al.", Fernandez et al. (3 d/wk)" and "Walts et al." is greated than 1.25, as in the previous analysis. It is worth



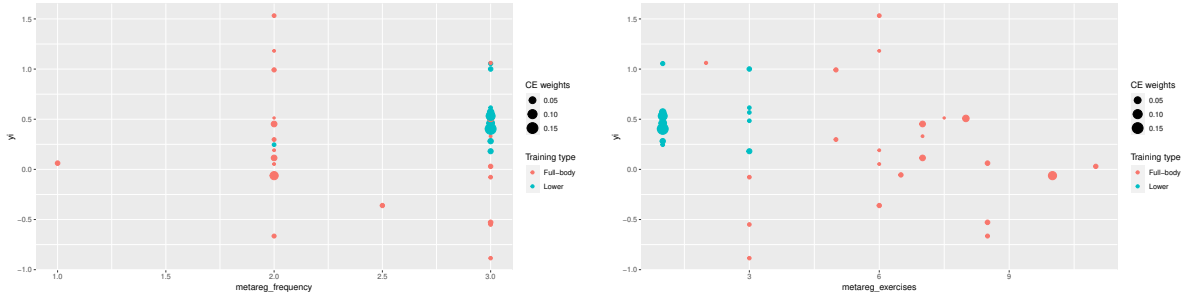
(A) `training` vs y_i

(B) `duration` vs y_i



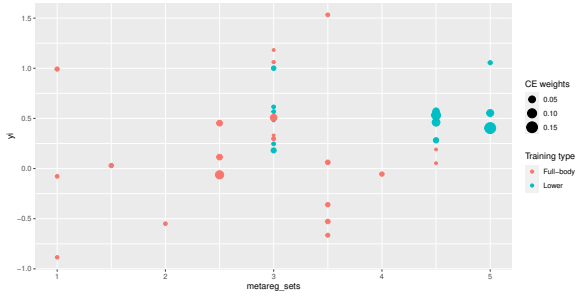
(C) `metareg_intensity` vs y_i

(D) `metareg_intensity` vs y_i



(E) `metareg_frequency` vs y_i

(F) `metareg_exercise` vs y_i



(G) `metareg_sets` vs y_i

FIGURE 2.19: Bivariate exploratory analysis of the relationship of the most interesting covariates with respect to the measure of effect of the single study in the analysis of relative changes in lower body strength. The size of the points in the box is proportional to the weights of the CE model.

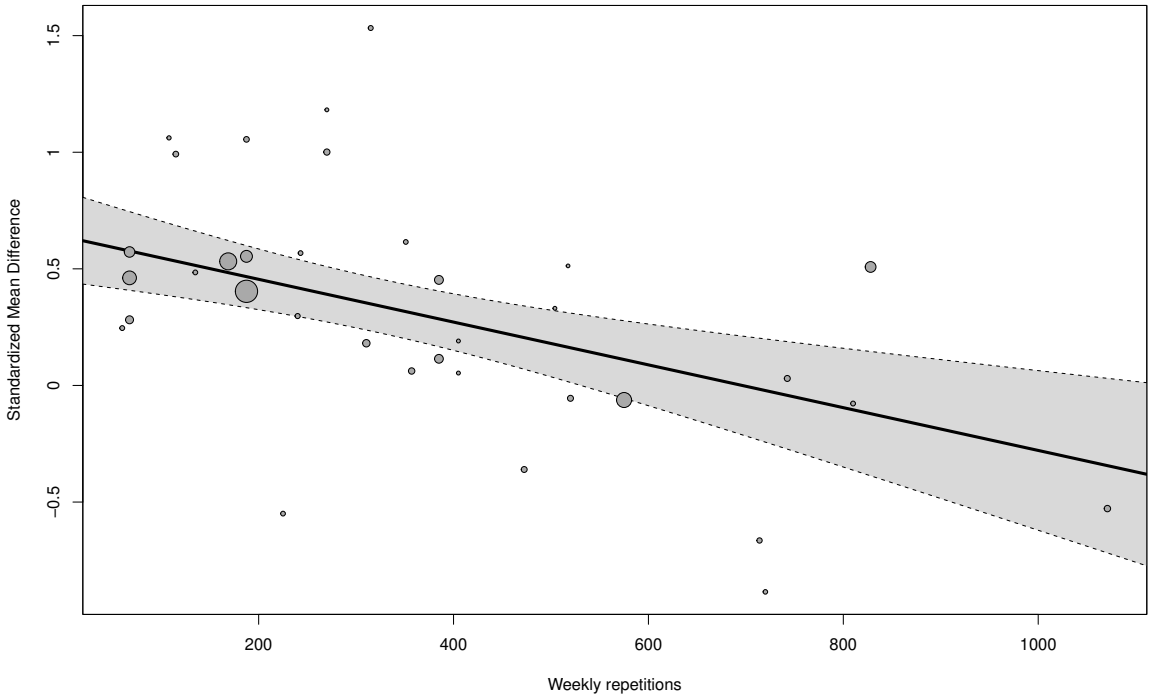


FIGURE 2.20: Estimated model variation of the effect measure y_i as a function of the number of weekly repetitions in the dataset of the absolute changes in lower body muscle strength. The size of the circle is proportional to $1/(s_i^2 + \hat{\tau}^2)$

.

noting that "Leenders et al." impact the estimated of the the slope but not the intercept, while the other two studies did not. However, the removal of "Leenders et al." did not change practically the estimated value of β to the fourth decimal place. In fact, we have $\hat{\beta}_1 = -0.08 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.14 \times 10^{-2}, -0.03 \times 10^{-2})$, which does not alter the inferential conclusions.

The Tables 2.16, 2.17, 2.18 and 2.19 report respectively the estimates of the β coefficients, their confidence intervals, the significance test for each coefficient and the estimate of τ^2. It should be noted that the inferential conclusions remain unchanged.

2.9 Analysis of the relative changes in muscle size

The RE model estimated on the relative effect measures of differences between males and females in the relative change in muscle size is based on the results of 30 studies and 1064 individuals, of whom 560 are females. The point estimate of the parameter of interest is $\hat{\beta} = 0.1$, with an associate 95% confidence interval of$(-0.04, 0.23)$. Accordingly, the p-value for the test of the null hypothesis of no effect is $p = 0.16$. Therefore,

TABLE 2.16: Point estimate and standard error of β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Coefficient	Estimator	Estimate	Standard Error
β_0	DL	0.64	0.10
β_0	ML	0.64	0.10
β_0	REML	0.64	0.10
β_0	PM	0.64	0.10
β_1	DL	-0.92×10^{-3}	0.24×10^{-3}
β_1	ML	-0.92×10^{-3}	0.24×10^{-3}
β_1	REML	-0.92×10^{-3}	0.24×10^{-3}
β_1	PM	-0.92×10^{-3}	0.24×10^{-3}

TABLE 2.17: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β_0 and β_1 in the meta-regression model fitted on the absolute changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Coefficient	Estimator	LL	UL	Width
β_0	DL_Wald	0.44	0.83	0.39
β_0	ML_Wald	0.45	0.83	0.38
β_0	REML_Wald	0.45	0.83	0.38
β_0	ML_prof_Wilks	0.45	0.83	0.38
β_0	REML_prof_Wilks	0.45	0.83	0.38
β_0	Hartung-Knapp-Sidik-Jonkman	0.39	0.99	0.60
β_0	prof_Wilks_Skovgaard	0.45	0.83	0.38
β_1	DL_Wald	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-3}
β_1	ML_Wald	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-3}
β_1	REML_Wald	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-3}
β_1	ML_prof_Wilks	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-3}
β_1	REML_prof_Wilks	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-3}
β_1	Hartung-Knapp-Sidik-Jonkman	-0.35×10^{-2}	-0.17×10^{-2}	0.14×10^{-3}
β_0	prof_Wilks_Skovgaard	-0.44×10^{-3}	-0.14×10^{-2}	0.96×10^{-2}

TABLE 2.18: Significance test on the parameter β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in absolute changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	β_0	β_1
DL	0.10×10^{-9}	0.16×10^{-3}
Hartung-Knapp-Sidik-Jonkman	0.63×10^{-5}	0.31×10^{-2}
ML_Wald	0.10×10^{-9}	0.16×10^{-3}
REML_Wald	0.10×10^{-9}	0.16×10^{-3}
ML_prof_Wilks	0.84×10^{-7}	0.19×10^{-3}
Skovgaard	0.10×10^{-5}	0.68×10^{-3}
Bartlett	1.00×10^{-6}	0.32×10^{-3}
Permutations	$< 10^{-4}$	$< 10^{-4}$

TABLE 2.19: Point estimate and standard error of τ^2 in the meta-regression mixed effect model fitted to estimate on absolute changes in lower body muscle strength after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.92×10^{-3}	0.03
ML	0.14×10^{-5}	0.02
REML	0.07×10^{-5}	0.02
PM	0.11×10^{-2}	0.03

the null hypothesis that there is no difference between males and females in upper body strength gain cannot be rejected when the change is expressed relative to baseline. The Cochran's Q test is not significant ($Q = 34.21, \quad df = 29, p = 0.23$) and thus the hypothesis of a common true effect is not rejected. Nevertheless, the the point estimate of the statistic indicating the percentage of variance explained by heterogeneity is $I^2 = 10.4\%$, indicating that the main source of variance is due to within-study variability. Moreover, the point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0.01$, with an associate 95% confidence interval of $(0, 0.17)$, using Viechtbauer (2007). The results presented so far were shown in the forest plot in Figure 2.21.

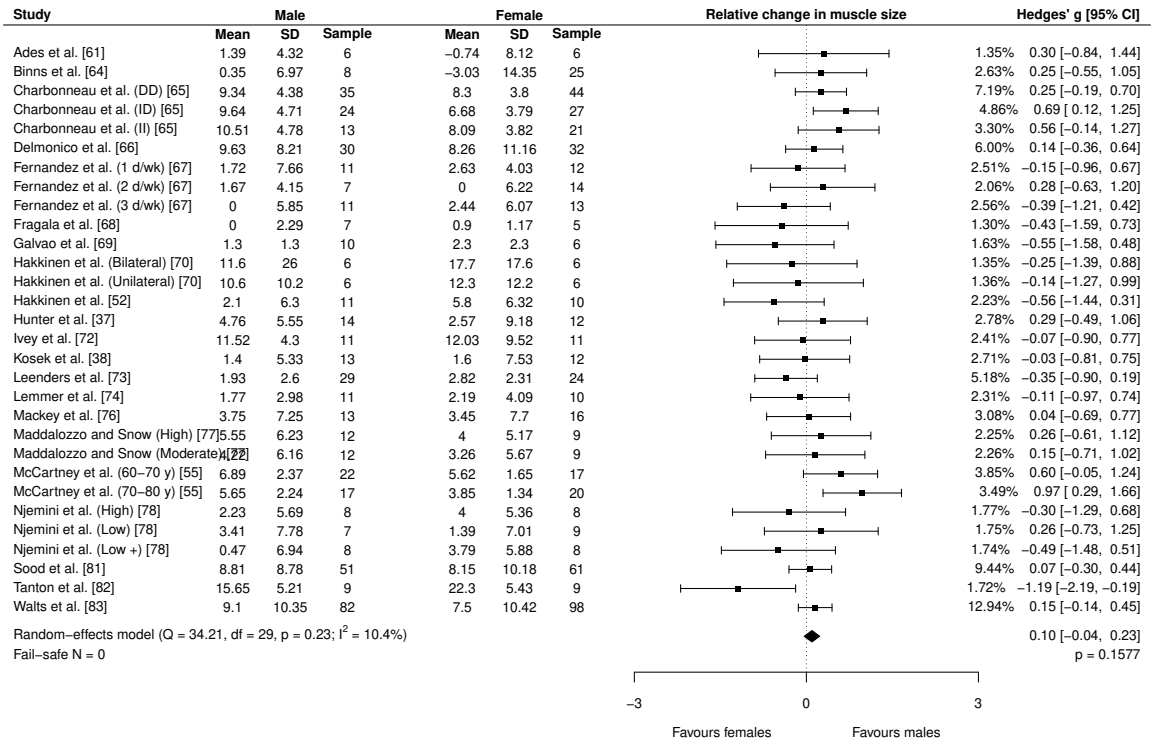
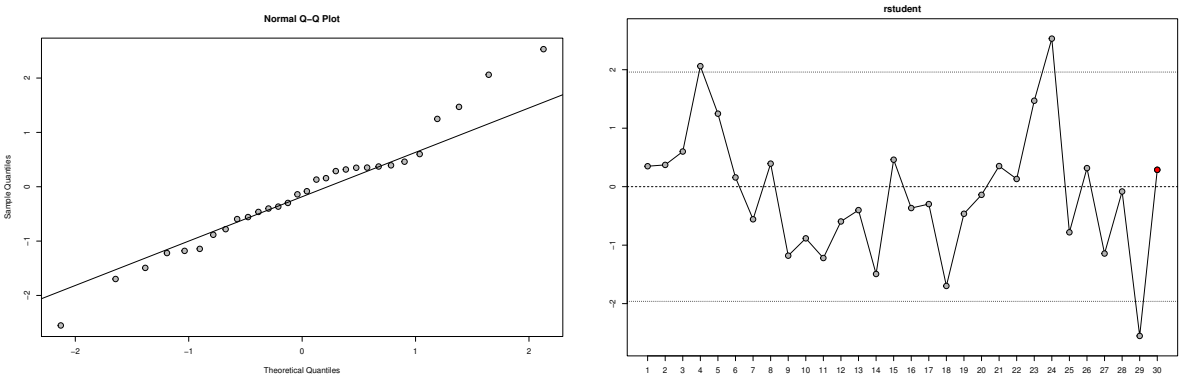


FIGURE 2.21: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in relative changes in muscle size after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom

2.9.1 Evaluation of model assumption and influential values

According to the Figure 2.22a, the standardized residuals are not arranged according to a normal distribution, particularly with respect to the right tail of the distribution. However, the Shapiro-Wilk test does not support these conclusions, as it is definitely not significant ($W = 0.98, p = 0.69$). Moreover, the plot in Figure 2.22b shows that only 3 points have a standardized deleted residual beyond the limits of 1.96 or -1.96, which turns out to be 10% of the points, although this does not raise any particular suspicion since the sample of 30 studies is not that large.

From the graphs in Figure 2.23, no single study is configured as influential. However, unlike previous analyses, the graph of the $COVRATIO_i$ in Figure 2.23d shows how



(A) Residuals Q-Q plot (B) residuals vs study numbers

FIGURE 2.22: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in relative changes in muscle size after a resistance training intervention.

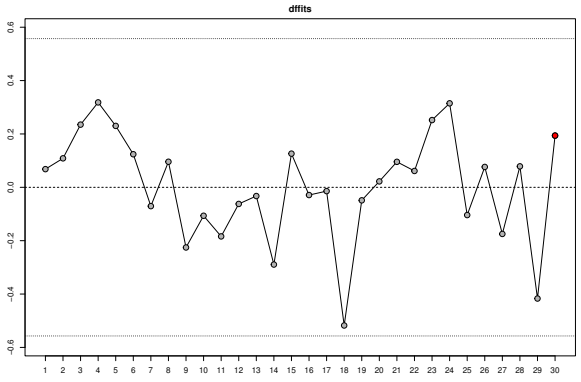
removing studies one by one has two opposite effects on covariance, depending on which study is considered. Particularly, the value of τ^2 in the model calculated after one-to-one removal of the studies "Charbonneau et al (ID) [65]," "Leenders et al [73]," "McCartney et al (70-80 y) [55]," and "Tanton et al [82]" is $\hat{\tau^2} = 0$, suggesting that some commonality in these studies could lead to heterogeneity in the meta-analysis. Note, however, that the amount of overall heterogeneity is small and not significant.

Note also that the weight of "Walts et al [83]" is particularly high (12%), suggesting that this study is particularly influential in estimating the effect. Such a high weighting is partly due to the fact that the estimate of τ^2 is small and not significant, thus regressing the model on the CE model. In addition, the study has a significantly larger sample size than the other included studies. It is well explain in the Chapter 1 that the weights of the CE model are the reciprocal of the standard error and that the standard error is a function of sample size.

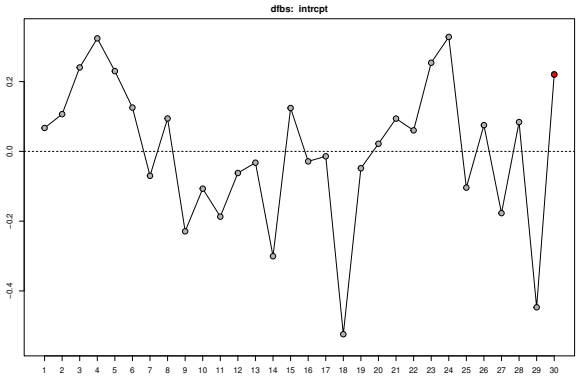
2.9.2 Evaluation of risk of publication bias

Looking at the contour-enhanced funnel plot in Figure 2.24a, there seems to be a lack of study with positive large effect size and large standard error. The Egger's test for asymmetry of the funnel plot partially confirms the conclusion, in fact it is significant with variance as predictor ($Z = -2.07, p = 0.04$) but not with standard error ($Z = -1.79, p = 0.07$). Figure 2.24b shows the Egger regression line calculated using standard error and sample variance as predictors.

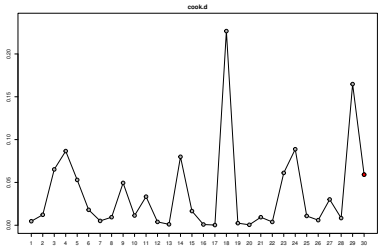
Figure 2.24c shows the contour-enhanced funnel plot where the number of missing trials was added to achieve symmetry using the trim and fill method. The results support the



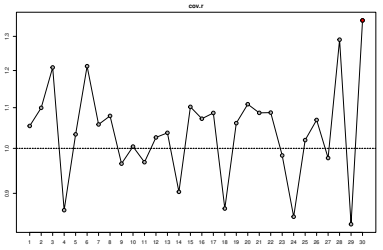
(A) $DFFITS_i$



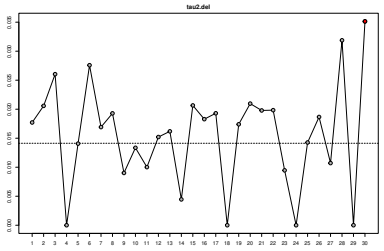
(B) $DFBETAS_i$



(C) Cook's distances



(D) $COVRATIO_i$



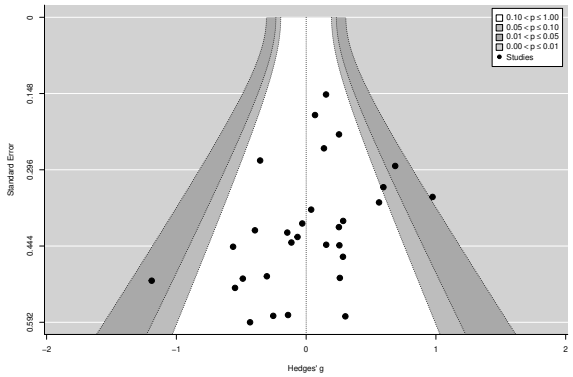
(E) R_i

FIGURE 2.23: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in relative changes in muscle size after a resistance training intervention. The statistics are plotted against a study progressive number.

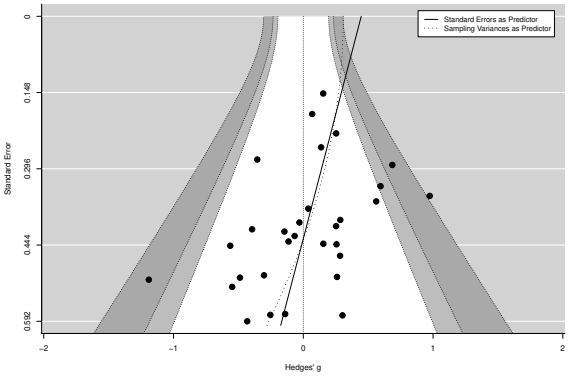
graphical analysis, suggesting the lack of 6 studies on the right side of the plot. It is also noteworthy that the sensitivity analysis for asymmetry performed with the trim and fill method yields $\hat{\beta} = 0.18$ with a 95% confidence interval of $(0.03, 0.33)$, which changes the conclusions of the meta-analysis, indicating that the relative gain in muscles size favours male against female. Finally, Rosenthal's fail-safe N calculation is equal to zero because the estimated effect in the model based only on the data is not statistically significant (Figure 2.21).

2.9.3 Comparison with more advanced methods

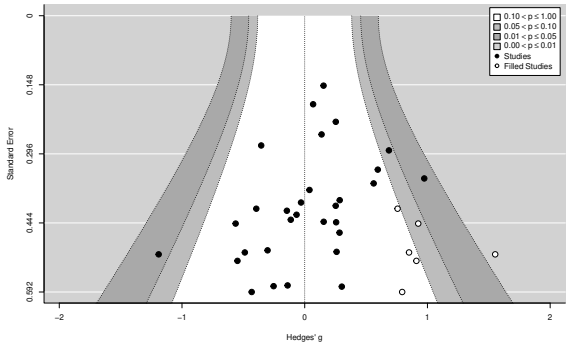
As can be seen in Table 2.20, the point estimate of the intercept of the model is particularly robust to the choice of method used. In fact, the method proposed by DerSimonian & Laird (1986), the maximum likelihood estimation, the restricted maximum likelihood estimation (REML), and Paule-Mandel's method yield similar results. This is not surprising since the estimator $\hat{\beta}$ is the same with exception of the weights used, which vary between methods based on τ^2. The percentage of heterogeneity in the model is very small, as indicated by previous analyses, and therefore has little effect on the β



(A) Contour enhanced funnel plot



(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.24: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in relative changes in muscle size after a resistance training intervention.

estimate. Similarly, the standard errors of $\hat{\beta}$ in the different models differ based on τ^2 and therefore also appear very similar.

TABLE 2.20: Point estimate and standard error of β of the RE model fitted to estimate the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.09	0.07
ML	0.11	0.06
REML	0.10	0.07
PM	0.09	0.07

The confidence intervals for β also seem to offer a common interpretation, with similar endpoints and comparable width, except for the Hartung-Knapp-Sidik-Jonkman method (refer to Table 2.21).

TABLE 2.21: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β of the RE model fitted to estimate the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	LL	UL	Width
DL_Wald	-0.05	0.23	0.28
ML_Wald	-0.02	0.23	0.24
REML_Wald	-0.04	0.23	0.27
ML_prof_Wilks	-0.02	0.23	0.24
REML_prof_Wilks	-0.04	0.23	0.27
Hartung-Knapp-Sidik-Jonkman	-0.11	0.25	0.36
prof_Wilks_Skovgaard	-0.06	0.23	0.29

It should also be noted that all the p-values are larger than 5%.

TABLE 2.22: Significance test on the parameter β of the RE model fitted to estimate the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	p
DL	0.16
Hartung-Knapp-Sidik-Jonkman	0.45
ML_Wald	0.09
REML_Wald	0.16
ML_prof_Wilks	0.18
Skovgaard	0.91
Bartlett	0.22
Permutations	0.07

The point estimate of τ^2 is also very similar among the methods used, with comparable standard errors (see Table 2.23). This is likely due to the negligible amount of heterogeneity.

2.9.4 Meta-regression

The exploratory bivariate scatter plots are presented in Figure 2.25. Note from Figure 2.25a that there is only one study where participants are training only the "upper body".

TABLE 2.23: Point estimate and standard error of τ^2 of the RE model fitted to estimate the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.02	0.04
ML	$< 10^{-4}$	0.02
REML	0.01	0.03
PM	0.03	0.04

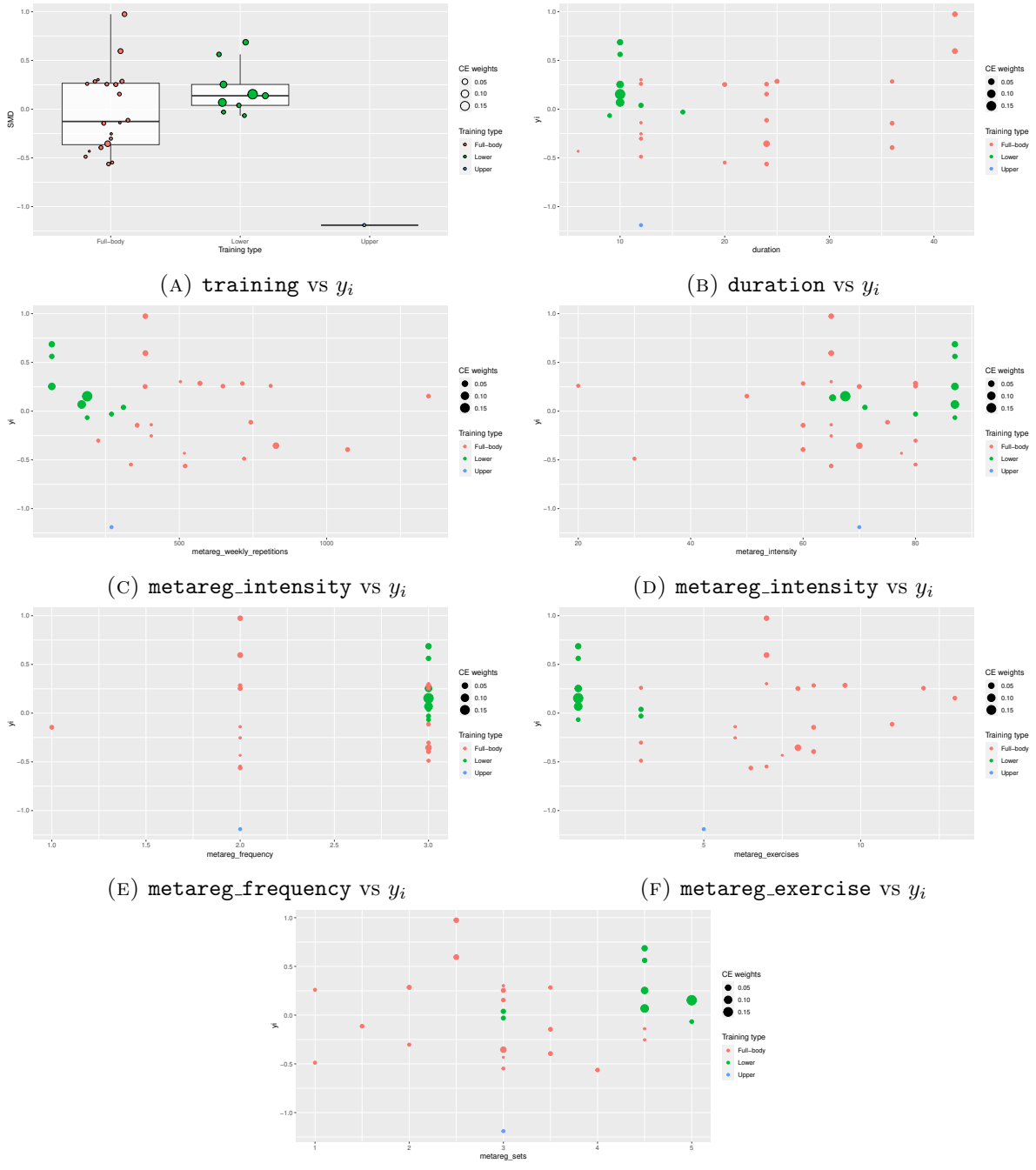
Moreover, in the previous analysis, none of study was focused on upper body training only. From the scatter plot observation and considering what has been discussed in the Section 2.7.4 the following initial meta-regression mixed effect model (of the type described in the equation 1.40) was proposed:

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 \text{training} + \beta_2 \text{metareg\_weekly\_repetitions} + \\ &+ \beta_3 \text{training} \times \text{metareg\_weekly\_repetitions} + \beta_4 \text{duration} + \\ &+ \beta_5 \text{metareg\_intensity} + \beta_6 \text{metareg\_frequency} + \delta_i + \epsilon_i, \end{aligned}$$

where all the variables are continuous, except for `training` which is a dummy. Although the same initial model has been proposed as in previous meta-regressions, from the observation of Figures 2.25d and 2.25e, it is quite certain that the variables `metareg_intensity` and `metareg_frequency` will be excluded from the final model. Moreover, from the observation of the graph in Figure 2.25c, it actually seems that the slope of the line is different for those who train only the lower body compared to those who train the whole body and that the slope may not be linear.
Then the backward selection led to the identification of a considerably simpler model with the following linear predictor:

$$Y_i = \beta_0 + \beta_1 \text{metareg\_weekly\_repetitions} + \beta_2 \text{duration} + \delta_i + \epsilon_i.$$

The results of the selection are available in the code. The estimated value of τ^2 is $\hat{\tau}^2 = 0$, and the p-value of the Cochran's Q test is particularly high ($Q = 25.68, \quad df = 27, \quad p = 0.54$), indicating that the hypothesis homogeneity of the true effects cannot be rejected. The estimated model yields a coefficient $\hat{\beta}_0 = 0.09$ with a Wald 95% confidence interval of $(-0.15, 0.33)$, $\beta_1 = -0.07 \times 10^{-2}$ with a a Wald $1 - \alpha = 0.95$ confidence interval of $(-0.13 \times 10^{-2}, -0.02 \times 10^{-2})$ and a $\beta_2 = -0.02$ with a a



(A) `training` vs y_i

(B) `duration` vs y_i

(C) `metareg_intensity` vs y_i

(D) `metareg_intensity` vs y_i

(E) `metareg_frequency` vs y_i

(F) `metareg_exercise` vs y_i

(G) `metareg_sets` vs y_i

FIGURE 2.25: Bivariate exploratory analysis of the relationship of the most interesting covariates with respect to the measure of effect of the single study in the analysis of relative changes in lower body strength. The size of the points in the box is proportional to the weights of the CE model.

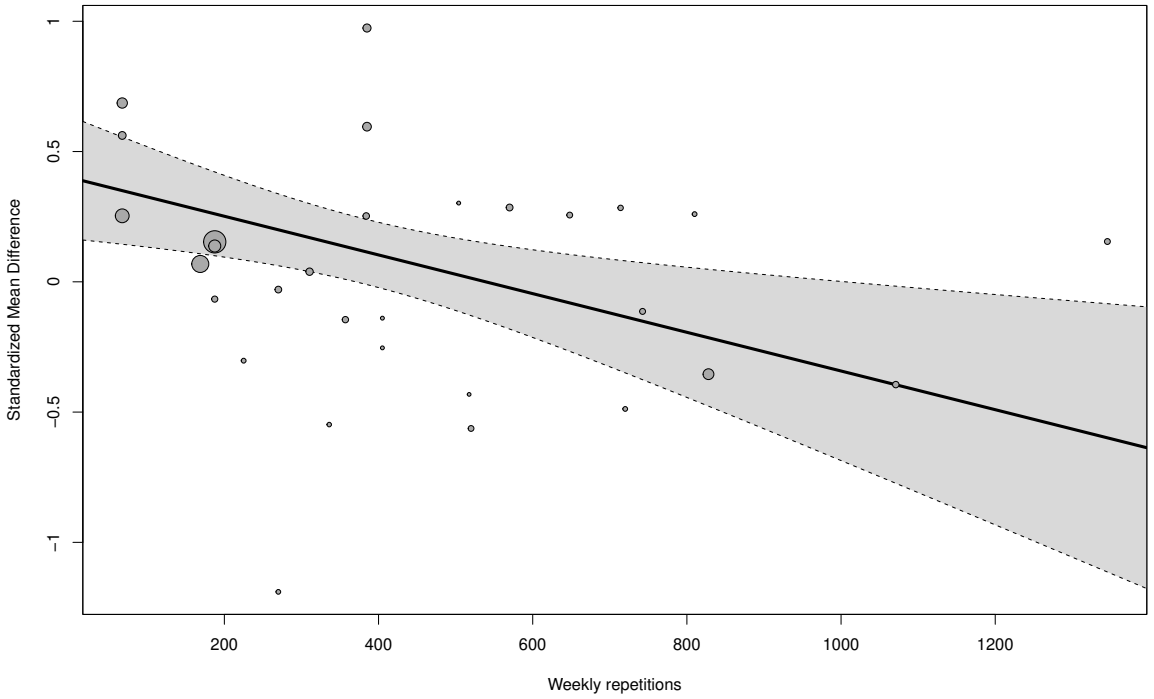


FIGURE 2.26: Estimated model variation of the effect measure y_i as a function of the number of weekly repetitions in the dataset of the relative changes in muscle size. The size of the circle is proportional to $1/(s_i^2 + \hat{\tau}^2)$

.

Wald $1 - \alpha = 0.95$ confidence interval of $(0.22 \times 10^{-2}, 0.03 \times 10^{-2})$. Hence, although the intercept of the model is not significantly different from zero, the coefficient for `metareg_weekly_repetitions` and `duration` are. The interpretation of the model is quite straighforward: the effect estimates decrease with when the number of weekly repetitions increase, meaning that it favours female. On the other side, the increase in the study duration seems to favour males. The marginal model showing the relationship between `metareg_weekly_repetitions` and the effect estimate y_i is reported in Figure 2.26, while Figure 2.27 shows the relationship with `duration`. The moderator is also capable of explaining the heterogeneity present in the intercept only meta-analysis model.

Furthermore, the model does not exhibit significant deviations from the assumption of normality. The Shapiro-Wilk test conducted on the standardized deleted residuals is not significant ($W = 0.98, p = 0.78$). The diagnosis of influential studies identifies three possible studies with outliers in at least one of dffits, $COVRATIO_i$ and dfbetas. The three the studies "Maddalozzo and Snow (Moderate)," "McCartney et al. (60-70 y)" and "McCartney et al. (70-80 y)" have a $COVRATIO_i$ greater than 1.5, however, the

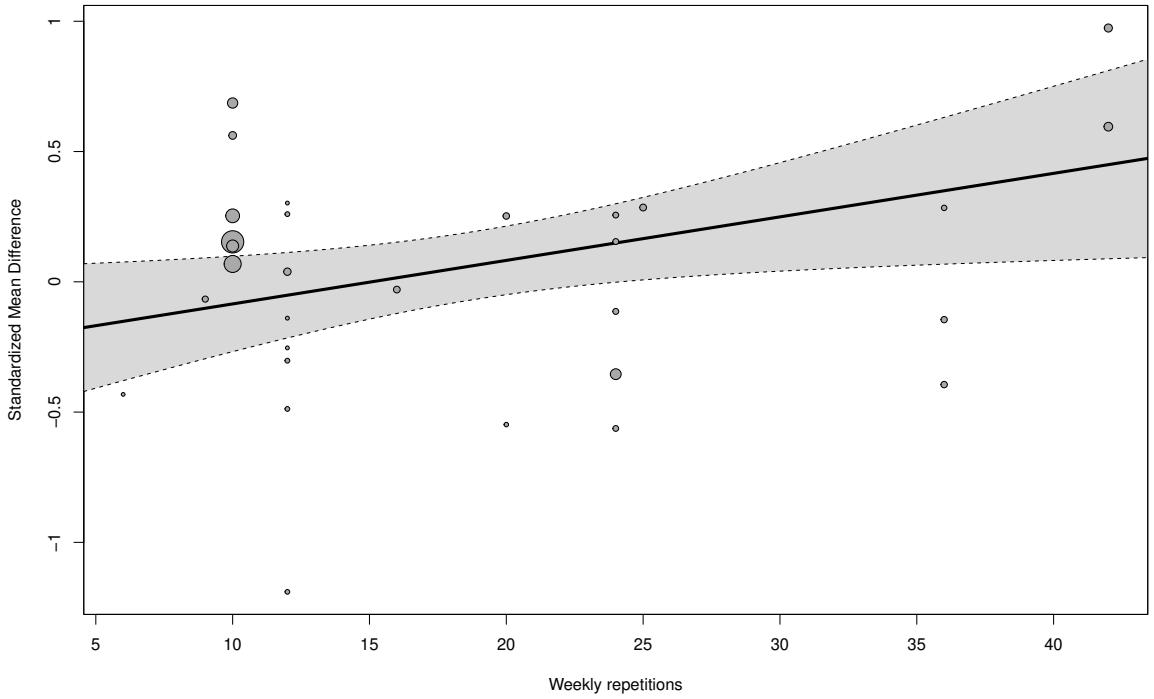


FIGURE 2.27: Estimated model variation of the effect measure y_i as a function of the number of study duration in the dataset of the relative changes in muscle size. The size of the circle is proportional to $1/(s_i^2 + \hat{\tau}^2)$

.

variance of τ^2 is such that to the second decimal place results in $\hat{\tau}^2 = 0$. "Maddalozzo and Snow (Moderate)" and "McCartney et al. (70-80 y)" present a value of diifs greater than $3 \times \sqrt{(p/K - p)} = 1$, indicating that their removal is influential on the model predictions. Cook's distances are also found to be high compared to the other studies, although well below their limit of $\chi_{0.5,p+1}$. "Maddalozzo and Snow (Moderate)" is also inflential in estimating β_1 and "McCartney et al. (70-80 y)" in estimating β_2, with a dfbs greater than 1. Their removal change the inferential conclusion, in fact $\hat{\beta}_1 = -0.07 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.15 \times 10^{-2}, < \times 10^{-4})$ and the respective p-value is $p = 0.07$, while $\hat{\beta}_2 = 0.44 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.02, 0.03)$ and the respective p-value is $p = 0.7327$. These results indicate that `duration` can be removed from the model. Doing it in the reduced dataset leads to $\hat{\beta}_0 = 0.25$ with a Wald 95% confidence interval of $(0.05, 0.46)$ and $\hat{\beta}_1 = -0.06 \times 10^{-2}$ with a Wald 95% confidence interval of $(-0.11 \times 10^{-2}, -0.01 \times 10^{-2})$. The intercept results significantly different from zero and positive, indicating that protocol with low volume still favours male. The slope is reduced but the direction of the interpretation is not changed.

TABLE 2.24: Point estimate and standard error of β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Coefficient	Estimator	Estimate	Standard Error
β_0	DL	0.09	0.12
β_0	ML	0.09	0.12
β_0	REML	0.09	0.12
β_0	PM	0.09	0.12
β_1	DL	-0.07×10^{-2}	-0.03×10^{-2}
β_1	ML	-0.07×10^{-2}	-0.03×10^{-2}
β_1	REML	-0.07×10^{-2}	-0.03×10^{-2}
β_1	PM	-0.07×10^{-2}	-0.03×10^{-2}
β_2	DL	0.02	0.01
β_2	ML	0.02	0.01
β_2	REML	0.02	0.01
β_2	PM	0.02	0.01

The Tables 2.24, 2.25, 2.26 and 2.27 report respectively the estimates of the β coefficients, their confidence intervals, the significance test for each coefficient and the estimate of τ^2.

2.10 Analysis of the absolute changes in muscle size

The RE model estimated on the absolute change between males and females in muscle size is based on the results of 28 studies and 1040 individuals, of whom 548 are female. The point estimate of the parameter of interest is $\hat{\beta} = 0.45$, with an associate 95% confidence interval of$(0.23, 0.67)$. Accordingly, the null hypothesis of no effect can be rejected ($p < .01 \times 10^{-2}$).

The Cochran's Q test is significant ($Q = 63.57, \quad df = 27, p =< .01 \times 10^{-2}$) and thus the hypothesis of a common true effect is rejected. The percentage of variance explained by heterogeneity is $I^2 = 62.1\%$, indicating that the main source of variance is due to between-study variability.

Moreover, the point estimate of the nuisance parameter obtained with the method of moments is $\tau^2 = 0.20$, and 95% confidence interval is $(< 10^{-4}, 0.49)$, using the method of Viechtbauer (2007). The results presented so far were shown in the forest plot in Figure 2.28.

TABLE 2.25: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β_0 and β_1 in the meta-regression model fitted on the relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Coefficient	Estimator	LL	UL	Width
β_0	DL_Wald	-0.15	0.33	0.48
β_0	ML_Wald	-0.15	0.33	0.48
β_0	REML_Wald	-0.15	0.33	0.48
β_0	ML_prof_Wilks	-0.15	0.33	0.48
β_0	REML_prof_Wilks	-0.15	0.33	0.48
β_0	Hartung-Knapp-Sidik-Jonkman	-0.33	0.40	0.73
β_0	prof_Wilks_Skovgaard	-0.15	0.33	0.48
β_1	DL_Wald	-0.01×10^{-1}	-0.02×10^{-2}	0.01×10^{-1}
β_1	ML_Wald	-0.01×10^{-1}	-0.02×10^{-2}	0.01×10^{-1}
β_1	REML_Wald	-0.01×10^{-1}	-0.02×10^{-2}	0.01×10^{-1}
β_1	ML_prof_Wilks	-0.01×10^{-1}	-0.02×10^{-2}	0.01×10^{-1}
β_1	REML_prof_Wilks	-0.01×10^{-1}	-0.02×10^{-2}	0.01×10^{-1}
β_1	Hartung-Knapp-Sidik-Jonkman	-0.01×10^{-1}	0.02×10^{-3}	0.01×10^{-1}
β_1	prof_Wilks_Skovgaard	-0.01×10^{-1}	-0.02×10^{-3}	0.01×10^{-1}
β_2	DL_Wald	0.02×10^{-1}	0.031	0.03
β_2	ML_Wald	0.02×10^{-1}	0.031	0.03
β_2	REML_Wald	0.02×10^{-1}	0.031	0.03
β_2	ML_prof_Wilks	0.02×10^{-1}	0.03	0.03
β_2	REML_prof_Wilks	0.02×10^{-1}	0.03	0.03
β_2	Hartung-Knapp-Sidik-Jonkman	-0.03×10^{-1}	0.03	0.04
β_2	prof_Wilks_Skovgaard	0.02×10^{-1}	0.03	0.03

TABLE 2.26: Significance test on the parameter β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	β_0	β_1	β_2
DL	0.48	0.02	0.01
Hartung-Knapp-Sidik-Jonkman	0.84	0.09	0.06
ML_Wald	0.48	0.02	0.01
REML_Wald	0.48	0.02	0.01
ML_prof_Wilks	0.48	0.02	0.01
Skovgaard	0.02	0.01	0.01
Bartlett	0.53	0.03	0.01
Permutations	0.20	0.01	0.01

TABLE 2.27: Point estimate and standard error of τ^2 in the meta-regression mixed effect model fitted to estimate on relative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	$< 10^{-4}$	0.03
ML	$< 10^{-4}$	0.02
REML	$< 10^{-4}$	0.02
PM	$< 10^{-4}$	0.03

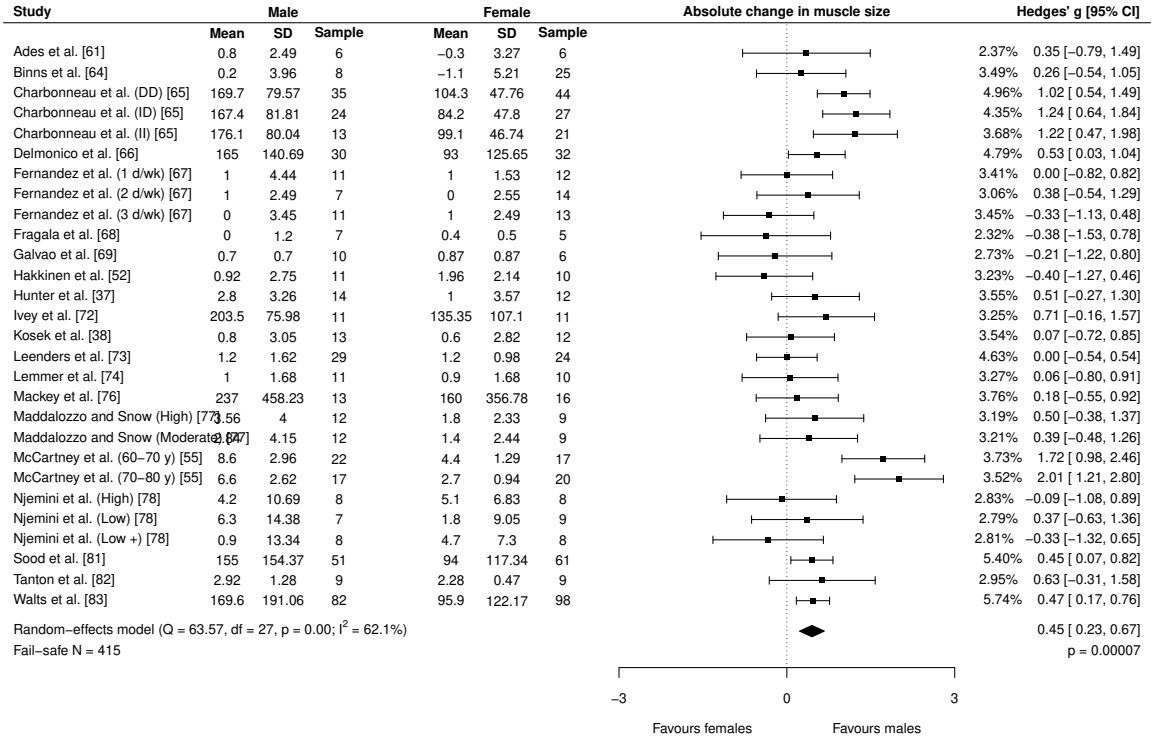
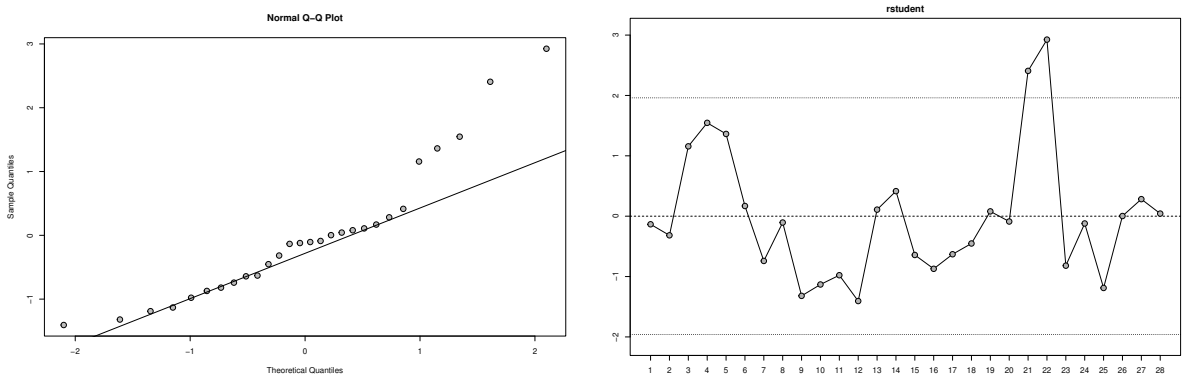


FIGURE 2.28: Forest plot depicting the results of a meta-analysis investigating the difference between males and females in absolute changes in muscle size after a resistance training intervention. The forest plot displays the effect size estimates for each individual study along with their corresponding 95% confidence intervals. The size of each square corresponds to the weight assigned to each study according to the RE formulation. The overall effect size estimate based on the random effect model proposed by DerSimonian & Laird (1986) is represented by the diamond at the bottom, with its width indicating the 95% CI. Cochran's Q statistic, a measure of heterogeneity and the Rhosenthal's fail-safe number of study needed to obtain a non-significan effect are reported in the bottom left corner. Abbreviations: SD, Standard Deviation; p, p-value; df, degrees of freedom

2.10.1 Evaluation of model assumption and influential values

According to the Figure 2.29a, the standardized deleted residuals of the model are not normally distributed, with a main detachment in the top-right corner of the plot. The Shapiro-Wilk test to support the graphical analysis, being highly significant ($W = 0.90, p = 0.01$). The graph in Figure 2.29b shows that two standardized deleted residuals is above the limit of 1.96. Under the normality hypothesis, the 5% of the residuals are expected to exceed those limits, and thus what is observed could be compatible with what is expected. In fact the 5% of 28 is 1.4. However, the evidence against normality obtained from the Shapiro-Wilk test and the normal qq-plot are enough to doubt about on this assumption.



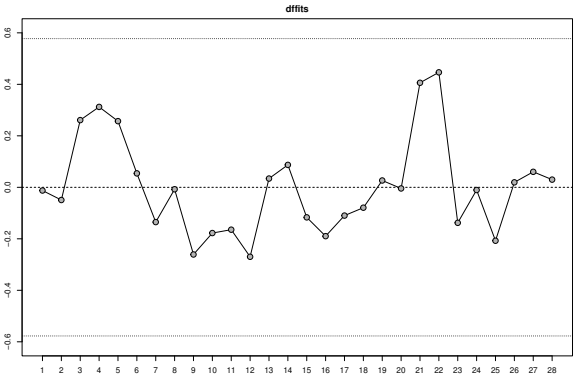
(A) Residuals Q-Q plot (B) residuals vs study numbers

FIGURE 2.29: Evaluation of the normality assumption on the standardized deleted residuals for the model evaluating the difference between males and females in absolute changes in muscle size after a resistance training intervention.

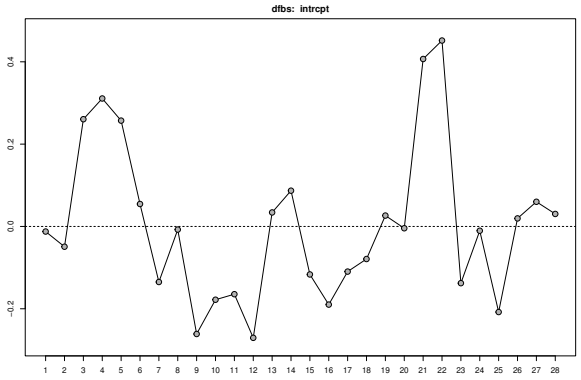
As for the influential values, we can see that none of the studies has an impact on the β estimate (Figure 2.30). However, looking at the $COVRATIO_i$ figure (Figure 2.30d), it is clear that removing the studies "McCartney et al. (60-70 y) [55]" and "McCartney et al. (70-80 y) [55]" significantly reduces the estimate of τ^2 and, more importantly, removing "McCartney et al. (70-80 y) [55]" implies $\hat{\tau}^2 < \times 10^{-4}$.

2.10.2 Evaluation of risk of publication bias

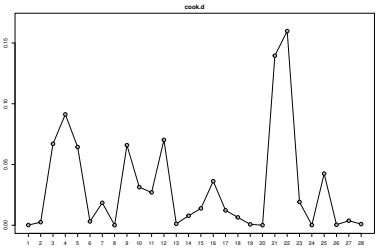
Looking at the contour-enhanced funnel plot in Figure 2.31a, there seems to be an indication of asimmetry, as one would expect a larger number of studies with positive and significant effect size and relatively big standard error. The Egger's test for asymmetry is not significant both using variance ($Z = -1.76, p = 0.08$) or the standard error ($Z = -1.51, p = 0.13$) as a predictor(Figure 2.31b), not confirming our observations. However, the trim and fill methods suggest that the estimated number of missing study



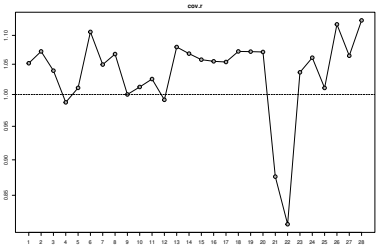
(A) $DFFITS_i$



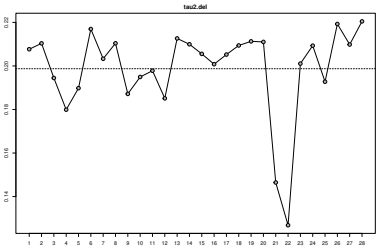
(B) $DFBETAS_i$



(C) Cook's distances



(D) $COVRATIO_i$



(E) R_i

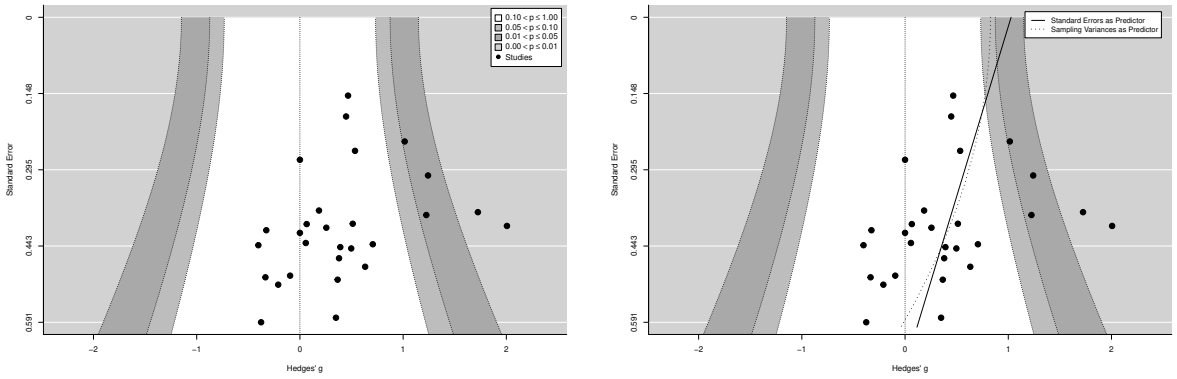
FIGURE 2.30: Evaluation of the influential studies on the study predict average value, on the β coefficints and on τ^2 on the model investigating the difference between males and females in absolute changes in muscle size after a resistance training intervention. The statistics are plotted against a study progressive number.

on the bottom right corner is 7, see Figure 2.31c). In accordance with the Trim and Fill sensitivity analysis, the results of the meta-analysis would be slightly different, not in direction but in magnitude ($\hat{\beta} = 0.64$, with a 95% confidence interval is $(0.41, 0.86)$). Finally,the Rosenthal's fail-safe N calculation suggested that 415 studies with nil effects are needed to have a non-significant estimate of β (Figure 2.28).

2.10.3 Comparison with more advanced methods

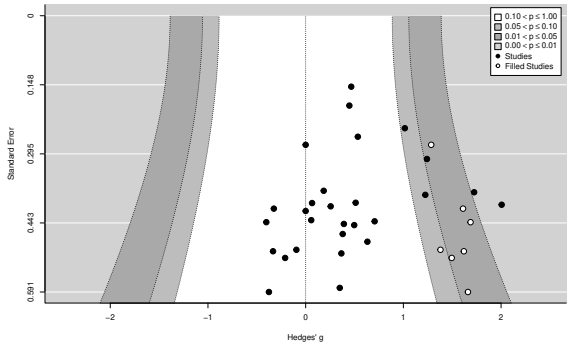
The discussion of the robustness of the point estimate of β and its standard error with respect to the estimation method is similar to that in Section 2.5.3. The values are given in Table 2.28.

It should also be noted that all the p values used to assess the significance of the intercept test are significant at the third decimal place (Table 2.30), hence we can conclude that the effect robust and that the detachment from the normality are not a concern to establish the significance of the test.



(A) Contour enhanced funnel plot

(B) Egger's regression 'functions'



(C) Trim and Fill

FIGURE 2.31: Publication bias assessment evaluated via contour enhanced funnel plot in the model investigating the difference between males and females in absolute changes in muscle size after a resistance training intervention.

TABLE 2.28: Point estimate and standard error of β of the RE model fitted to estimate the difference between males and females in absolute changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.45	0.11
ML	0.45	0.11
REML	0.45	0.11
PM	0.45	0.11

Also, the discussion of the robustness of the estimator of τ^2 is analogous to Section 2.5.3, but for all the methods the amount of heterogeneity is substantial in this model, see Table 2.31.

TABLE 2.29: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β of the RE model fitted to estimate the difference between males and females in absolute changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	LL	UL	Width
DL_Wald	0.24	0.66	0.42
ML_Wald	0.23	0.67	0.43
REML_Wald	0.23	0.67	0.44
ML_prof_Wilks	0.23	0.67	0.43
REML_prof_Wilks	0.23	0.67	0.44
Hartung-Knapp-Sidik-Jonkman	0.21	0.68	0.47
prof_Wilks_Skovgaard	0.21	0.67	0.45

TABLE 2.30: Significance test on the parameter β of the RE model fitted to estimate the difference between males and females in absolute changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	p
DL	0.01×10^{-2}
Hartung-Knapp-Sidik-Jonkman	0.02×10^{-2}
ML_Wald	0.05×10^{-3}
REML_Wald	0.01×10^{-2}
ML_prof_Wilks	0.05×10^{-2}
Skovgaard	0.08×10^{-2}
Bartlett	0.07×10^{-2}
Permutations	0.05×10^{-2}

2.10.4 Meta-regression

The exploratory bivariate scatter plots are presented in Figure 2.32. As for the analysis of the relative changes in muscle size, there is only one study training only the "upper body". The following initial meta-regression mixed effect model (of the type described in the equation 1.40) was proposed:

TABLE 2.31: Point estimate and standard error of τ^2 of the RE model fitted to estimate the difference between males and females in absolute changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

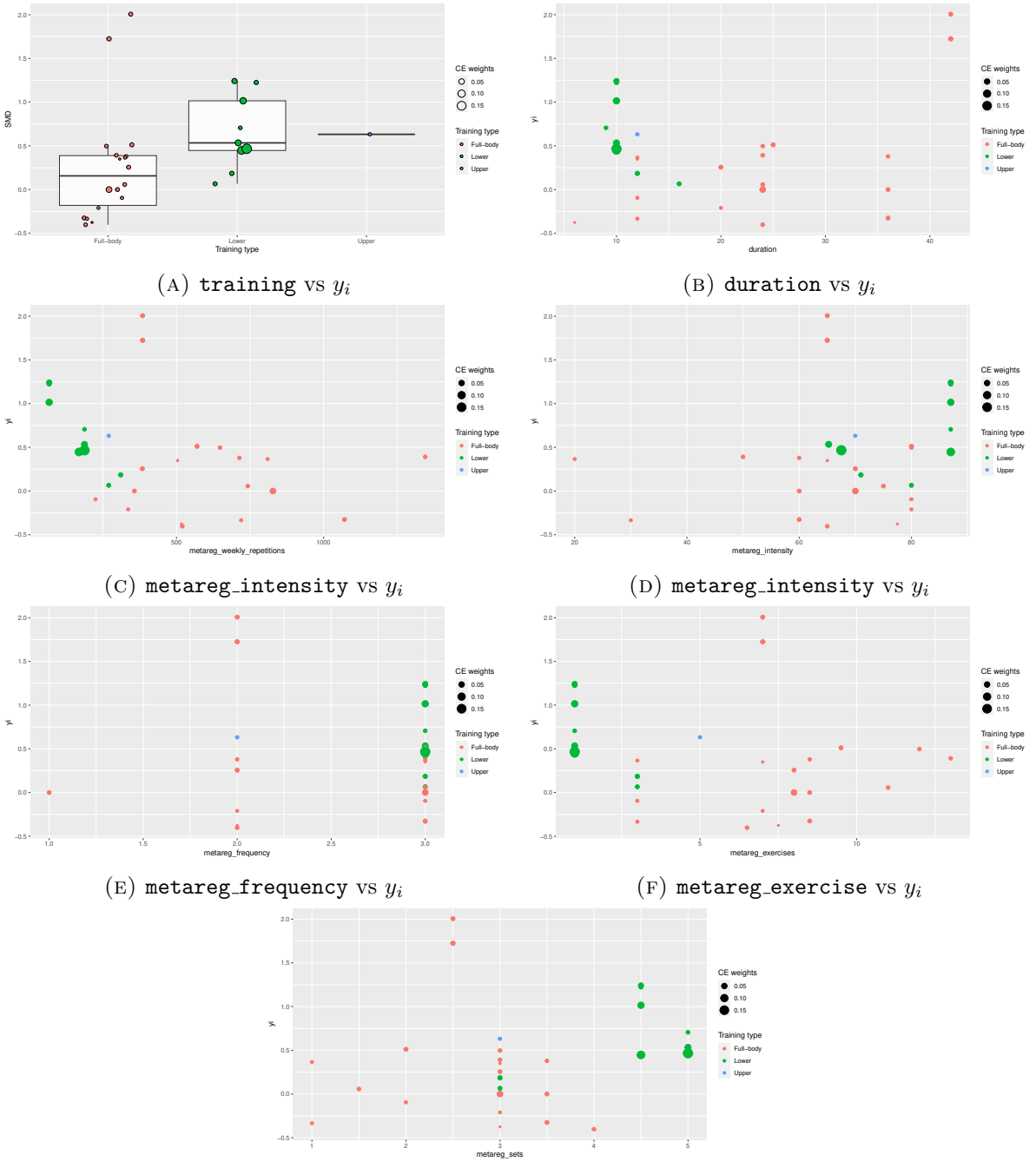
Estimator	Estimate	Standard Error
DL	0.16	0.09
ML	0.19	0.09
REML	0.20	0.09
PM	0.20	0.10

$$\begin{aligned} Y_i =& \beta_0 + \beta_1 \text{training} + \beta_2 \text{metareg\_weekly\_repetitions} + \\ &+ \beta_3 \text{training} \times \text{metareg\_weekly\_repetitions} + \beta_4 \text{duration} + \beta_5 \text{UpperBody} + \\ &+ \beta_6 \text{metareg\_intensity} + \beta_7 \text{metareg\_frequency} + + \delta_i + \epsilon_i, \end{aligned}$$

where all the variables are continuous, except for `training` and `UpperBody` which are dummy variables.
The backward selection led to the identification of the following linear predictor:

$$\begin{aligned} Y_i =& \beta_0 + \beta_1 \text{metareg\_weekly\_repetitions} + \\ &+ \beta_2 \text{training} + \beta_3 \text{duration} + \\ &+ \beta_4 \text{training} \times \text{metareg\_weekly\_repetitions} \\ &+ \beta_5 \text{UpperBody} + \delta_i + \epsilon_i. \end{aligned}$$

The results of the selection are available in the code. Please note that `UpperBody` is an dummy variable indentifying the only study that training the upper boby, i.e. "Tanton et al.". The estimated value of τ^2 is $\hat{\tau}^2 < \times 10^{-4}$, and the p-value of the Cochran's Q test is particularly high ($Q = 26.95, \quad df = 22, \quad p = 0.21$), indicating that the hypothesis homogeneity of the true effects cannot be rejected. The estimated model is presented in the Table 2.32.
The interaction coefficient β_4 explains the the different response to training volume between studies involving the lower body and those involving the upper body. Note also that the range of variation in training volume is significantly smaller in studies involving only the lower body. In addition, an effect for type of training is present, even taking into account the difference in volume. Finally, the test on the single study



(A) `training` vs y_i

(B) `duration` vs y_i

(C) `metareg_intensity` vs y_i

(D) `metareg_intensity` vs y_i

(E) `metareg_frequency` vs y_i

(F) `metareg_exercise` vs y_i

(G) `metareg_sets` vs y_i

FIGURE 2.32: Bivariate exploratory analysis of the absationship of the most interesting covariates with respect to the measure of effect of the single study in the analysis of absolute changes in muscle size. The size of the points in the box is proportional to the weights of the CE model.

Variable	β	Standard error	Z-value	p
intrcpt	-0.31	0.38	-0.83	0.41
Lower	1.34	0.39	3.42	0.06×10^{-2}
Upper	0.63	0.53	1.19	0.23
repetitions	-0.06×10^{-2}	0.04×10^{-2}	-1.68	0.09
duration	0.04	0.01	3.91	-0.1×10^{-2}
Lower:repetitions	-0.04×10^{-1}	-0.01×10^{-1}	-3.25	-0.01×10^{-1}

TABLE 2.32: Results of the regression model after the selection of variables for absolute changes in muscle size. Abbreviations: p, p-value

Variable	β	Standard error	Z-value	p
intrcpt	-0.41	0.48	-0.85	0.40
Lower	1.49	0.48	3.09	0.02×10^{-2}
Upper	0.71	0.55	1.29	0.20
repetitions	-0.03×10^{-2}	0.06×10^{-2}	-0.54	0.59
duration	0.04	0.01	2.89	0.03×10^{-2}
Lower:repetitions	-0.05×10^{-2}	$0.14 \times 10^{-}2$	-3.24	0.01×10^{-2}

TABLE 2.33: Results of the regression model after the selection of variables for absolute changes in muscle size after the removal of the influential studies.

involving the upper body is not significant, indicating that this is not an influential value. Again, the introduction of covariates is sufficient to explain the heterogeneity present in the model. Furthermore, the model does not exhibit significant deviations from the assumption of normality. The Shapiro-Wilk test conducted on the standardized deleted residuals is not significant ($W = 0.98, \quad p = 0.80$) and only the residuals of the study "Fernandez et al. (3 d/wk)", "Mackey et al." and "Maddalozzo and Snow (Moderate)" are above the cut-off value of 1.96. However, the same studies possess high value of diffits, very close to the limit of 1.57 and the Cook's distances are very high with respect to the other studies. Note that "Fernandez et al. (3 d/wk)" and "Mackey et al." are also found to be significantly influential for estimating the regression coefficient for the variable `metareg_weekly_repetitions`, presenting a value of dfbs greater than 1, while "Maddalozzo and Snow (Moderate)" is found to be influential for estimating the coefficient of `duration`. It is also worth noting that the estimate of τ^2 is particularly sensitive to the removal of one of the studies. Note, however, that at most you have $\hat{\tau}^2 = 0.20$ should you remove the "Sood et al." study. The removal of the on the influential studies led to the coefficient estimation in Table 2.33. Removing the influential values shows that the largest difference lies in `metareg_weekly_repetitions` which is no longer significant. The p-value increases from $p = 0.09$ to $p = 0.59$. However, it is preferred not to remove it from the model, since the interaction still remains significant.

Tables 2.34, 2.35, 2.36 and 2.37 report respectively the estimates of the β coefficients, their confidence intervals, the significance test for each coefficient and the estimate of

TABLE 2.34: Point estimate and standard error of β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in absative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Coefficient	Estimator	Estimate	Standard Error
β_0	DL	-0.31	0.40
β_0	ML	-0.31	0.38
β_0	REML	-0.31	0.38
β_0	PM	-0.31	0.41
β_1	DL	-0.06×10^{-2}	0.04×10^{-2}
β_1	ML	-0.06×10^{-2}	0.04×10^{-2}
β_1	REML	-0.06×10^{-2}	0.04×10^{-2}
β_1	PM	-0.06×10^{-2}	0.04×10^{-2}
β_2	DL	1.36	0.43
β_2	ML	1.34	0.39
β_2	REML	1.34	0.39
β_2	PM	1.36	0.44
β_3	DL	0.04	0.01
β_3	ML	0.04	0.01
β_3	REML	0.04	0.01
β_3	PM	0.04	0.01
β_4	DL	-0.45×10^{-2}	0.16×10^{-2}
β_4	ML	-0.44×10^{-2}	0.14×10^{-2}
β_4	REML	-0.44×10^{-2}	0.14×10^{-2}
β_4	PM	-0.45×10^{-2}	0.16×10^{-2}
β_5	DL	0.63	0.56
β_5	ML	0.64	0.53
β_5	REML	0.64	0.53
β_5	PM	0.63	0.57

τ^2. It should be noted that the inferential conclusions remain unchanged.

TABLE 2.35: Confidence intervals lower limit (LL), upper limit (UL) and width for the effect β_0 and β_1 in the meta-regression model fitted on the absative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood (ML_Wald) and the restricted maximum likelihood (REML_Wald) using the Wald's methods, the maximum likelihood (ML_prof_Wilks) and the restricted maximum likelihood (REML_prof_Wilks) using Wilks' method, the Hartung-Knapp-Sidik-Jonkman estimator, and the profile likelihood Skovgaard's method (prof_Wilks_Skovgaard).

Estimator	Estimator	LL	UL	Width
β_0	DL_Wald	-1.09	0.47	1.56
β_0	ML_Wald	-1.05	0.43	1.48
β_0	REML_Wald	-1.05	0.43	1.48
β_0	ML_prof_Wilks	-1.05	0.43	1.48
β_0	REML_prof_Wilks	-1.05	0.43	1.48
β_0	Hartung-Knapp-Sidik-Jonkman	-1.25	0.64	1.89
β_0	prof_Wilks_Skovgaard	-1.05	0.43	1.48
β_1	DL_Wald	-0.14×10^{-2}	0.02×10^{-2}	0.15×10^{-2}
β_1	ML_Wald	-0.14×10^{-2}	0.01×10^{-2}	0.15×10^{-2}
β_1	REML_Wald	-0.14×10^{-2}	0.01×10^{-2}	0.15×10^{-2}
β_1	ML_prof_Wilks	-0.14×10^{-2}	0.01×10^{-2}	0.15×10^{-2}
β_1	REML_prof_Wilks	-0.14×10^{-2}	0.01×10^{-2}	0.15×10^{-2}
β_1	Hartung-Knapp-Sidik-Jonkman	-0.15×10^{-2}	0.04×10^{-2}	0.20×10^{-2}
β_1	prof_Wilks_Skovgaard	-0.14×10^{-2}	0.02×10^{-2}	$0.15 \times 10^{-}2$
β_2	DL_Wald	0.52	2.20	1.68
β_2	ML_Wald	0.57	2.11	1.54
β_2	REML_Wald	0.57	2.11	1.54
β_2	ML_prof_Wilks	0.57	2.11	1.54
β_2	REML_prof_Wilks	0.57	2.11	1.54
β_2	Hartung-Knapp-Sidik-Jonkman	0.30	2.47	2.17
β_2	prof_Wilks_Skovgaard	0.57	2.11	1.54
β_3	DL_Wald	0.02	0.06	0.04
β_3	ML_Wald	0.02	0.06	0.04
β_3	REML_Wald	0.02	0.06	0.04
β_3	ML_prof_Wilks	0.02	0.06	0.04
β_3	REML_prof_Wilks	0.02	0.06	0.04
β_3	Hartung-Knapp-Sidik-Jonkman	0.01	0.06	0.05
β_3	prof_Wilks_Skovgaard	0.02	0.06	0.04
β_4	DL_Wald	-0.75×10^{-2}	-0.14×10^{-2}	0.61×10^{-2}
β_4	ML_Wald	-0.71×10^{-2}	-0.18×10^{-2}	0.54×10^{-2}
β_4	REML_Wald	-0.71×10^{-2}	-0.18×10^{-2}	0.54×10^{-2}
β_4	ML_prof_Wilks	-0.71×10^{-2}	-0.18×10^{-2}	0.54×10^{-2}
β_4	REML_prof_Wilks	-0.71×10^{-2}	-0.18×10^{-2}	0.54×10^{-2}
β_4	Hartung-Knapp-Sidik-Jonkman	-0.87×10^{-2}	-0.04×10^{-2}	0.83×10^{-2}
β_4	prof_Wilks_Skovgaard	-0.71×10^{-2}	-0.18×10^{-2}	0.54×10^{-2}
β_5	DL_Wald	-0.47	1.74	2.21
β_5	ML_Wald	-0.41	1.68	2.09
β_5	REML_Wald	-0.41	1.68	2.09
β_5	ML_prof_Wilks	-0.41	1.68	2.09
β_5	REML_prof_Wilks	-0.41	1.68	2.09
β_5	Hartung-Knapp-Sidik-Jonkman	-0.70	1.96	2.66
β_5	prof_Wilks_Skovgaard	-0.41	1.68	2.09

TABLE 2.36: Significance test on the parameter β_0 and β_1 of the meta-regression mixed effect model estimated on the difference between males and females in absative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the Hartung-Knapp-Sidik-Jonkman method, the maximum likelihood (ML_Wald) and the restricted maximum likelihood methods (REML_Wald) using the Wald's statistic, the profile maximum likelihood method (ML_prof_Wilks) and the Skovgaard's method using the Wilk's test, the Wilk's method with Bartlett's correction and the permutation method.

Method	β_0	β_1	β_2	β_3	β_4	β_5
DL	0.41	0.09	0.06×10^{-2}	0.01×10^{-2}	0.12×10^{-2}	0.23
Hartung-Knapp-Sidik-Jonkman	0.53	0.27	0.01	0.45×10^{-2}	0.03	0.35
ML_Wald	0.41	0.09	0.06×10^{-2}	0.01×10^{-2}	$0.12 \times 10^{-}2$	0.23
REML_Wald	0.41	0.09	0.06×10^{-2}	0.01×10^{-2}	$0.12 \times 10^{-}2$	0.23
ML_prof_Wilks	0.41	0.09	0.06×10^{-2}	0.01×10^{-2}	0.16×10^{-2}	0.23
Skovgaard	0.40	0.07	0.06×10^{-2}	0.06×10^{-2}	0.05×10^{-1}	0.23
Bartlett	0.41	0.10	0.08×10^{-2}	0.02×10^{-2}	0.02×10^{-2}	0.24
Permutations	0.19	0.04	$< 10^{-4}$	0.05×10^{-2}	0.15×10^{-2}	0.13

TABLE 2.37: Point estimate and standard error of τ^2 in the meta-regression mixed effect model fitted to estimate on absative changes in muscle size after a resistance training intervention, obtained by DerSimonian & Laird (1986), the maximum likelihood method (ML), the restricted maximum likelihood method (REML) and the Paule-Mandel estimator (PM).

Estimator	Estimate	Standard Error
DL	0.03	0.05
ML	$< 10^{-4}$	0.02
REML	$< 10^{-4}$	0.03
PM	0.04	0.05

Chapter 3

Integrated likelihood inference in meta-analysis

3.1 Small studies in meta-analysis

This chapter focuses on small sample inference in meta-analysis as a consequence of small samples within each study. Indeed, one of the basic assumptions of the standard meta-analysis model is that the within-study variance is assumed to be known, specifically $\sigma_i^2 = s_i^2, \quad \forall i = 1, \ldots, K$. This assumption holds when the study sample is sufficiently large, because s_i^2 is an unbiased and consistent estimator of σ_i. In case of small within-study sample size if the uncertainty in the estimation of σ_i^2 is not considered, then results can be misleading. Several solutions have been proposed in the literature of two-stage approaches to meta-analysis: the use of a population-averaged study specific variance (Böhning et al., 2002), the use a non parametric estimator of the between-study variance that consider the within-study variance (Malzahn et al., 2000), the use of shrinkage approaches for variance estimation (Di Gessa, 2008) and an improvement of the likelihood results using the second-order asymptotic on likelihood ratio statistic (Sharma & Mathew, 2011). Some authors reasoned on the advantages of using the standardized mean difference in place of the un-standardized version to incorporate within-study variances directly in the effect measure (Johnson & Huedo-Medina, 2013). This chapter will present an approach based on integrated the likelihood (Bellio & Guolo, 2016), which eliminates variance components by integrating the likelihood function. It should be noted that other solutions are possible, such as the use of limit meta-analysis, which is a model that allows the effect of small studies to be estimated under appropriate assumptions (Rücker et al., 2011).

3.2 Integrated likelihood

Consider a statistical model dependent on a k-dimensional parameter θ, partitionable into $\theta = (\psi, \lambda)^T$, and assume that the parameter of interest is the p-dimensional ψ, while λ is a $(k-p)$-dimensional nuisance parameter. Let $L(\theta) = L(\psi, \lambda)$ be the likelihood function of the statistical model. The inference on the parameter of interest can always be based on the average of the likelihood functions $L(\psi, \lambda)$ with respect to a function of weights $\pi(\lambda|\theta)$ (Chapter 8, Section 8.4, Severini, 2000). Note that although $\pi(\lambda|\theta)$ is often referred to as a 'conditional prior density', it need not to be a probability density function in order to apply this method (Severini, 2007). Let Λ the space of all admissible λ. The integrated likelihood is defined as

$$L_{Int}(\psi; \pi) = \log \int_{\Lambda} L(\psi)\pi(\lambda)d\lambda. \tag{3.1}$$

The integrated log likelihood function can be expressed as $\ell_{Int}(\psi; \pi) = \log L_{Int}(\psi; \pi)$ and may be utilized as well for inference purposes. It should be noted that, while the method is quite general and can be applied to any parametric model, the resulting integrals may be complex to compute. The integrated likelihood approach offers a distinct advantage over other pseudo-likelihoods, such as the profile likelihood, in that it obtains the likelihood of the parameter of interest $L(\theta)$ by averaging rather than maximization, which renders it numerically more stable. Moreover, when a conditional or marginal likelihood function is available, it often corresponds to a integrated likelihood function with respect to a specific weighting function (Severini, 2000).

3.2.1 Desirable properties

In the present context, integrated likelihood will be employed in a non-Bayesian inference framework. To ensure its suitability for non-Bayesian inference, it is important to obtain certain desirable properties, which can be achieved by selecting an appropriate weights function $\pi(\boldsymbol{\lambda}|\boldsymbol{\theta})$. The rationale for this selection, as outlined below, is based on Severini (2000).

The first consideration in selecting the weight function $\pi(\boldsymbol{\lambda}|\boldsymbol{\psi})$ is that it should be chosen so that if γ is a nuisance parameter 'unrelated' to ψ, then γ and ψ should be independent under $\pi(\boldsymbol{\lambda}|\boldsymbol{\psi})$. In this context, the term 'unrelated' has a specific meaning: namely, that the maximum likelihood estimator of γ fixed ψ, denoted as $\hat{\gamma}_{\psi}$, should be approximately constant as a function of ψ. More precisely, we can define:

- γ 'weekly unrelated' to ψ if $\hat{\gamma}_{\psi} = \hat{\gamma} + O(n^{-1})$ for moderate deviations of ψ;

- γ 'strongly unrelated' to ψ if $\hat{\gamma}_\psi = \hat{\gamma} + O\left(n^{-1/2}\right) O(|\psi - \hat{\psi}|)$.

Note that orthogonal nuisance parameters are weekly but not strongly unrelated to the parameter of interest.

For the second desirable requirement, consider the factorizing likelihood in the form $L(\theta) = L_1(\psi)L_2(\gamma)$. Accordingly, $L_1(\psi)$ can be used as the likelihood for the inference on ψ. Hence, function $\pi(\lambda \mid \psi)$ should be chosen so that the unrelated parameters are independent. With this solution, the integrated likelihood for the inference on ψ would correspond to $L_1(\psi)$.

A third consideration for the integrated likelihood is that it should exhibit the frequentist properties of a genuine likelihood function, at least approximately. These include satisfying the two Bartlett identities, which are known as score and information unbiasedness. However, if $\pi(\gamma \mid \psi)$ does not depend on ψ:

- $E\left\{\ell'_{Int}(\psi); \psi, \lambda\right\} = O\left(n^{-1}\right)$, if γ is weakly unrelated to ψ,
- $E\left\{\ell''_{Int}(\psi) + \ell'_{Int}(\psi)\ell'_{Int}(\psi)^{\mathrm{T}}; \theta\right\}$ is also $O\left(n^{-1}\right)$, if γ is strongly unrelated to ψ,

suggesting that $\pi(\lambda \mid \psi)$ should be chosen so that, if a nuisance parameter γ is strongly unrelated to ψ, then ψ and γ are independent under $\pi(\lambda \mid \psi)$.

A fourth consideration is that the integrated likelihood should be not sensitive to the choice of the prior density. In other words, the inferential results should be the same disrespectfully to the prior choice. Consider the Laplace approximation of the integrated likelihood reported by Severini (2007):

$$L_{int}(\psi) = \int_\Lambda L(\psi, \lambda)\pi(\lambda \mid \psi)d\lambda = cL_p(\psi)\left|-\ell_{\lambda\lambda}\left(\psi, \hat{\lambda}_\psi\right)\right|^{-1/2} \pi\left(\hat{\lambda}_\psi \mid \psi\right)\left\{1 + O\left(n^{-1}\right)\right\},$$

where c denotes a constant not depending on ψ and $\ell_{\lambda\lambda}(\psi, \lambda) = \partial^2\ell(\psi, \lambda)/\partial\lambda\partial\lambda^{\mathrm{T}}$. It suggests that

- if γ is weakly unrelated to ψ and $\pi(\gamma \mid \psi)$ does not depend on ψ, then, for ψ in the moderate deviation range, $L_{Int}(\psi)$ does not depend on the form of $\pi(\gamma)$, if terms of order n^{-1} are ignored,
- If γ is strongly unrelated to ψ, then, in addition, $L_{Int}(\psi)$ does not depend on the form of $\pi(\gamma)$ for ψ in the large deviation range, if terms of order $n^{-1/2}$ are ignored.

The last desirable property is that the integrated likelihood remains invariant under the re-parametrization of the parameter of interest. It is noteworthy that this property does not impose any constraints on the selection of the weight function.
The desirability of these properties suggests the use of a re-parametrization such that the new nuisance parameter ϕ is strongly unrelated with ψ and the choice of a weight function for ϕ, namely $\pi(\phi|\psi)$ that does not depend on $\pi(\psi)$. So the integrated likelihood takes the following form

$$L_{Int}(\psi) = \int \tilde{L}(\psi, \phi)\pi(\phi)d\phi. \tag{3.2}$$

where $\tilde{L}(\psi, \phi)$ is the re-parametrized likelihood function.

3.2.2 Strongly unrelated nuisance parameter

Choosing which re-parameterization to adopt for the nuisance parameters is not trivial, as it is well known that the integrated likelihood function varies with the adopted parameterization. As an illustration, example 8.36 in Severini (2000) is reported. Let $Y_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma), \quad \forall i = 1, \ldots, N$, where μ is the parameter of interest and σ is the nuisance parameter, and a uniform density is used as the weighting function, $\pi(\sigma|\mu) = 1$. The integrated likelihood function results

$$\begin{aligned} L_U(\mu, \sigma) &= \int_0^\infty \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_i^n (y_i - \mu)^2\right\} d\sigma \\ &= \left\{1 + \frac{(\mu - \bar{y})^2}{\hat{\sigma}^2}\right\}^{-(n-1)/2}, \end{aligned}$$

where $\bar{y} = \frac{1}{n}\sum_i^n y_i$, $\hat{\sigma}^2 = \frac{1}{n}\sum_i^n (y_i - \bar{y})^2$ and L_U is used to highlight the uniform weight. Please note that the model is parametrized in term of $(\mu, \sigma)^T$ and not $(\mu, \sigma^2)^T$. If the model parametrized in terms of $(\mu, \sigma^2)^T$ was considered, the likelihood function would have been

$$L_U(\mu, \sigma^2) = \left\{1 + \frac{(\mu - \bar{y})^2}{\hat{\sigma}^2}\right\}^{-(n-2)/2},$$

whereas if we considered the model in terms of $(\mu, \log(\sigma))$

$$L_U(\mu, \log\sigma) = \left\{1 + \frac{(\mu - \bar{y})^2}{\hat{\sigma}^2}\right\}^{-n/2}.$$

To fulfill the requirements mentioned in the previous paragraph, a data-dependent re-parameterization of the nuisance parameters, known as the zero-score-expectation parameterization, was suggested by Severini (2007). This parameterization involves obtaining a parameter $\phi \equiv \phi(\psi, \lambda; \hat{\psi})$, by solving the equation:

$$E\left\{\ell_\lambda(\psi, \lambda); \hat{\psi}, \phi\right\} \equiv E\left\{\ell_\lambda(\psi, \lambda); \psi_0, \lambda_0\right\}\Big|_{(\psi_0, \lambda_0)=(\hat{\psi}, \phi)} = 0. \tag{3.3}$$

It is necessary to fix the value of (ψ, λ, ψ) and solving for ϕ yields $\phi(\psi, \lambda; \psi)^T$. It has been shown in the Appendix 1 of Severini (2007) that $\hat{\phi} = \hat{\lambda}$ and that ϕ is strongly unrelated to ψ. Please note that a data-dependent parameter in the likelihood function is not a source of problems as the data are considered fixed. The full rationale behind the derivation of the parameter ϕ is explained in the original publication (Severini, 2007). It is useful to dwell on the following properties of the integrated likelihood obtained by the zero-score-expectation re-parameterization. As exposed in Severini (2007),

- supposed a model in which the likelihood factorizes in the form $L_1(\psi)L_2(\gamma)$. The parameter ϕ is dependent on ψ and λ only through $\gamma(\psi, \lambda)$, which means that $L_{Int}(\psi) \propto L_1(\psi)$ for any $\pi(\phi)$ such that 3.1 is finite.
- $L_{Int}(\psi)$ is score and information unbiased to the order $O(n^-1)$.
- For a fixed type of prior, such as a uniform prior, $L_{Int}(\psi)$ is parameterization invariant to order $O(n^{-1/2})$ for fixed ψ and to order $O(n^{-1})$ for $\psi = \hat{\psi} + O(n^{-1/2})$.

3.2.3 Likelihood ratio statistics based on integrated likelihood

The log-likelihood ratio statistic is a primary inference tool in likelihood theory since it is re-parameterization invariant and provides confidence intervals that respect the parametric space boundaries. In the context of integrated likelihood, as stated in Severini (2010), a similar statistic is defined as

$$R_{Int} = sign(\psi_{Int} - \psi)\sqrt{2\{L_{Int}(\psi_{Int}) - L_{Int}(\psi)\}}, \tag{3.4}$$

where ψ_{Int} represents the maximizer of the integrated log likelihood for ψ. The test is apparently preferable to the profile likelihood test when the nuisance parameter has large dimensions due to computational problems related to the maximization. The quantity R_{Int} is obtained using a type of integration commonly used in Bayesian inference, and so any Bayesian software can help in its computation (Severini, 2010).

Indicating with R the profile log likelihood ratio statistics, Severini (2010) showed that $R_{Int} = R + O(n^{1/2})$. Hence R_{Int} is asymptotically normally distributed. Moreover, it has been demonstrated that R_{Int} can be modified to be normally distributed at the second order and the modification is well described in Severini (2010). An approximate confidence interval of level α can be obtained as:

$$L_{Int}(\psi) > L_{Int}(\psi_{Int}) - Z_{1-\alpha/2},$$

while the hypothesis $H_0 : \psi = 0$ against $H_1 : \psi \neq 0$ can be tested

$$|R_{Int}| \geq Z_{1-\alpha/2}$$

3.3 Integrated likelihood in meta-analysis

This section focuses on the application of integrated likelihood in meta-analysis where the measure of effect is on a continuous scale. A discussion on the binary outcomes can be found in the supporting information of Bellio & Guolo (2016). Please note that the following discussion is entirely based on the same paper.

Data from a meta-analysis are typically presented in pairs, denoted as (y_i, s_i^2) for all $i = 1, \ldots, K$. Assuming a RE model, the first element of the $i-th$ pair is an estimate of β_i and s_i^2 of the uncertainty associated with the study i. Given that the measures of effects are assumed to be independent, it follows that $y_i \sim \mathcal{N}(\beta, \sigma_i^2 + \tau^2)$.

The estimate s_i^2 is used to approximate σ_i^2 and is calculated with f_i degrees of freedom, which can vary based on the effect measure used. For example, when comparing the averages of two groups of size n_{i1} and n_{i2}, the degrees of freedom will be $n_{i1} + n_{i2} - 2$, whereas when measuring the difference between pre- and post-treatment in a single group, the degrees of freedom will be $n - 1$ (where n is the sample size of the single group). By applying a variable change formula, it is possible to derive the distribution of s_i^2 from the known distribution of the quantity

$$\frac{s_i^2 f_i}{\sigma_i^2} \sim \chi_{f_i}.$$

Given any absolutely continuous random variable X with density $X \sim f_X(x)$ and given a strictly increasing monotone transform of X of the type $Y = g(X)$, the probability density function $f_Y(y)$ is

$$f_Y(y) = f_X\{g^{-1}(y)\} \cdot \frac{d}{dy} g^{-1}(y).$$

For fixed f_i and σ_i and since $(f_i, \sigma_i) \in [0, +\infty) \times (0, +\infty)$, function $h(x) = \frac{xf_i}{\sigma_i}$ is strictly increasing monotone. Let $Z = \frac{s_i^2 f_i}{\sigma_i} \sim \chi^2_{f_i}$, $s_i^2 = g(Z) = \frac{Z\sigma_i}{f_i}$ and $\frac{d}{dy}g^{-1}(y) = f_i/\sigma_i$. Then

$$f_{S_i^2}(s_i^2) = \frac{(\frac{s_i^2 f_i}{\sigma_i})^{(f_i/2)-1} e^{\left(-\frac{s_i^2 f_i}{\sigma_i}\right)/2}}{2^{f_i/2}\Gamma(f_i/2)} \cdot \frac{f_i}{\sigma_i} \tag{3.5}$$

where Γ is the gamma function.

Hence, assuming the RE model and given that y_i and s_i^2 are independent, the joint probability density function for the i-th study is

$$f_{(Y_i,S_i^2)}(y_i, s_i^2) = \frac{1}{\sqrt{2\pi(\sigma_i^2+\tau^2)}} e^{-\frac{(x-\beta_i)^2}{2(\sigma_i^2+\tau^2)}} \cdot \frac{(\frac{s_i^2 f_i}{\sigma_i})^{(f_i/2)-1} e^{-\left(\frac{s_i^2 f_i}{\sigma_i}\right)/2}}{2^{f_i/2}\Gamma(f_i/2)} \cdot \frac{f_i}{\sigma_i}.$$

Since the studies are considered independent, the joint probability function of the whole sample is

$$f_{(Y_1,S_1^2),\ldots,(y_K,S_K^2)}\{(y_1, s_1^2), \ldots, (y_K, s_K^2)\} = \prod_i^K f_{Y_i,S_i^2}(y_i, s_i^2) =$$

$$= \prod_i^K \frac{1}{\sqrt{2\pi(\sigma_i^2+\tau^2)}} e^{-\frac{(x-\beta_i)^2}{2(\sigma_i^2+\tau^2)}} \cdot \frac{(\frac{s_i^2 f_i}{\sigma_i})^{(f_i/2)-1} e^{-(\frac{s_i^2 f_i}{\sigma_i})/2}}{2^{f_i/2}\Gamma(f_i/2)} \cdot \frac{f_i}{\sigma_i},$$

which, is formally equal to the likelihood function

$$L(\beta, \tau, \sigma_1, \ldots, \sigma_K | (y_1, s_1^2), \ldots, (y_K, s_K^2)),$$

henceforth denoted as $L(\beta, \tau, \sigma_1, \ldots, \sigma_K)$ for convenience. The log-likelihood function is obtained as

$$\ell(\beta, \tau, \sigma_1, \ldots, \sigma_K) = \log L(\beta, \tau, \sigma_1, \ldots, \sigma_K)$$

and it is proportional to

$$l(\beta, \tau, \sigma_1, \ldots, \sigma_K) = \sum_{i=1}^K \left\{ -\frac{1}{2}\log(\sigma_i^2+\tau^2) - \frac{1}{2}\frac{(x_i-\beta_i)^2}{(\sigma_i^2+\tau^2)} - \frac{f_i}{2}\log\sigma_i^2 - \frac{f_i s_i^2}{2}\log 2\sigma_i^2 \right\}. \tag{3.6}$$

Note that the likelihood (3.6) differs from (1.20) in that it considers the joint probability function of (y_i, s_i^2), whereas in (1.20) the σ_i^2 were assumed to be known and equal to s_i^2.

The number of parameter is different. In fact, in (1.20) the estimand is $(\beta, \tau^2)^T$ while in (3.6) it is $(\beta, \tau, \sigma_1, \ldots, \sigma_K)^T$. Furthermore, it is clear that the number of parameters increases with the number of studies. For convenience, let $\theta = (\beta, \tau, \sigma_1, \ldots, \sigma_K)^T$.
The object of interest for inferential purpose is β and the components of variance $(\tau^2, \sigma_1^2, \ldots, \sigma_K^2)^T$ are considered nuisance parameters. Therefore, the parameter θ is partitioned into $\theta = (\beta, \boldsymbol{\lambda})^T$ with $\boldsymbol{\lambda} = (\tau^2, \sigma_1^2, \ldots, \sigma_K^2)^T$. It follows that the integrated maximum log-likelihood in the form of (3.1) is

$$\ell_{Int}(\beta) = \log \int_0^\infty \int_0^\infty \cdots \int_0^\infty L(\beta, \tau, \sigma_1^2, \ldots, \sigma_K^2)\pi(\tau, \sigma_1^2, \ldots, \sigma_K^2|\beta) d\sigma_1^2 \ldots d\sigma_K^2 \tau \quad (3.7)$$

The intergral was calculated over $\boldsymbol{\lambda}$ in $\Lambda = [0, \infty) \times (0, \infty)^K$. The inference can be based on the integrated log-likelihood ratio test as in (3.4).
As explained above, it is necessary to re-parametrized the model using the zero-expectation parametrization solving (3.3).For the model considered so far, the score vector for $\boldsymbol{\lambda}$ is

$$\begin{aligned}
\ell'_\tau(\beta, \boldsymbol{\lambda}) &= \tau \sum_{i=1}^n \left\{ -\frac{1}{(\sigma_i^2 + \tau^2)} + \frac{(y_i - \beta)^2}{(\sigma_i^2 + \tau^2)^2} \right\}, \\
\ell'_{\sigma_i}(\beta, \boldsymbol{\lambda}) &= \sigma_i \left\{ -\frac{1}{(\sigma_i^2 + \tau^2)} + \frac{(y_i - \beta)^2}{(\sigma_i^2 + \tau^2)^2} - \frac{f_i}{\sigma_i^2} + \frac{f_i s_i^2}{\sigma_i^4} \right\}, \quad i = 1, \ldots, K.
\end{aligned}$$

The evaluation of the expected value of the above expressions at the parameter value $(\beta_0, \boldsymbol{\lambda_0})^T$, setting $\beta_0 = \hat{\beta}$ and $\boldsymbol{\lambda_0} = \boldsymbol{\phi}$, where $\boldsymbol{\phi} = (\zeta, \delta_1, \ldots, \delta_n)^T$, is

$$E\{\ell_\tau(\beta, \lambda); \beta_0, \lambda_0\}\big|_{(\beta_0, \lambda_0) = (\hat{\beta}, \boldsymbol{\phi})} = \tau \sum_{i=1}^K \left\{ \frac{(\delta_i^2 - \sigma_i^2) + (\zeta^2 - \tau^2) + (\hat{\beta} - \beta)^2}{(\sigma_i^2 + \tau^2)^2} \right\} \quad (3.8)$$

and

$$E\{\ell_{\sigma_i}(\beta, \lambda); \beta_0, \lambda_0\}\big|_{(\beta_0, \lambda_0) = (\hat{\beta}, \boldsymbol{\phi})} = \sigma_i \left\{ \frac{(\delta_i^2 - \sigma_i^2) + (\zeta^2 - \tau^2) + (\hat{\beta} - \beta)^2}{(\sigma_i^2 + \tau^2)^2} + \frac{f_i\,(\delta_i^2 - \sigma_i^2)}{\sigma_i^4} \right\}. \quad (3.9)$$

Equating (3.8) and (3.9) to zero and solving for $(\zeta, \delta_1, \ldots, \delta_n)^T$ leads to

$$\begin{aligned}
\tau^2 &= \zeta^2 + (\hat{\beta} - \beta)^2, \\
\delta_i &= \sigma_i, \qquad i = 1, \ldots, n.
\end{aligned}$$

The chosen weight function for ζ and $\sigma_1, \ldots, \sigma_K$ is $\pi(\zeta) \propto 1$ and $\pi(\sigma_i) \propto 1/\sigma_i^b$, for fixed b and specifically $b = 1$, given that different bs do not change the results Bellio & Guolo (2016).

Under the assumption that the study participating in the meta-analysis are independent and using a weight function with separated components for $\boldsymbol{\phi}$ imply that

$$\ell_{Int}(\beta) = \log \int_0^\infty \left\{ \prod_{i=1}^{K} g_i(\beta, \zeta) \right\} \pi(\zeta) d\zeta, \tag{3.10}$$

where $g_i(\beta, \zeta) = \int_0^\infty L(\beta, \zeta, \sigma_i)\, \pi(\sigma_i)\, d\sigma_i$ and $L(\beta, \zeta, \sigma_i)$ is the likelihood term for study i. Each of the K integrals $g_i(\beta, \zeta)$ and the main integral in (3.10) are one-dimensional integrals that can be approximated via standard numerical methods.

3.4 Integrated likelihood in meta-regression

Where the interest is in explaining some of the heterogeneity of the treatment effect by one or more covariates, integrated likelihood can be used to make inference on a meta-regression model. Let $\boldsymbol{x_i}$ denote the vector of d covariates available at the aggregated meta-analysis level for each study, with $x_{1i} = 1$ and $\boldsymbol{\beta}$ a d-dimensional vector of parameter to be estimated, the meta-regression model can be written as

$$Y_i \sim N(\boldsymbol{x_i}^T \boldsymbol{\beta}, \sigma_i^2 + \tau^2). \tag{3.11}$$

The log likelihood function is similar to 3.7 with the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \tau, \sigma_1, \ldots, \sigma_K)^T$ including the covariates $\boldsymbol{x_i}$ in the mean of Y_i. Note that the parameter of interest here is the vector $\boldsymbol{\beta}$ representing the effect of the covariates on the effect measure of the meta-analysis. Therefore, in order to obtain confidence intervals for the single parameter or tests for the single coefficient, it is recommended to use an integrated profile likelihood. The zero-score-expectation re-parametrization can be obtained with the same approach previously described, however it is not possible to obtain an expression of τ^2 in terms of ζ^2. In fact ζ^2 is related to τ^2 as follows,

$$\tau^2 = \zeta^2 + \frac{\sum_{i=1}^{K} (\sigma_i^2 + \tau^2)^{-2} (\boldsymbol{x_i}^T \hat{\boldsymbol{\beta}} - \boldsymbol{x_i}^T \boldsymbol{\beta})^2}{\sum_{i=1}^{K} (\sigma_i^2 + \tau^2)^{-2}}.$$

when, in case of $\sigma_i^2 = \sigma^2, \quad \forall i = 1, \ldots, K$,

$$\tau^2 = \zeta^2 + \sum_{i=1}^{n} (\boldsymbol{x_i}^T \hat{\boldsymbol{\beta}} - \boldsymbol{x_i}^T \boldsymbol{\beta})^2.$$

Although the case in which the within-study variances are equal is a borderline and unrealistic case, this approximation of zero-score expectation parameterization is generally used to have a closed form for τ^2.

3.5 Details of the implementation of the algorithm

The `R` to implement the integrated likelihood in meta-analysis is in the Appendix. In this section, we will focus on some computational details useful for efficiency gains. First, the code used the log-likelihood function (3.10), since solving one-dimensional integrals is easier than the multidimensional counterpart. All integrals were calculated using the `R` function `integrate`, which implements an adaptive Gauss-Kronrod quadrature based on the C QUAPACK library Piessens et al. (2012). The computation of the integrated likelihood was not possible in all the cases, especially when some of the studies under investigation had a particularly high sample size. Function $g_i(\beta, \zeta)$ in (3.10)contains the likelihood function, which grows with increasing sample size and it is approximately proportional to $\exp(n_i)$, where n_i is the sample size of the i-th study. The operation outside the calculation of $g_i(\beta, \zeta)$ is the product over all sample elements of the the meta-analysis. This product was often too large and therefore the R software returned the value 'Inf'. According to the IEC 60559 (also known as IEEE 754) standards, the largest number that can be processed by the `R` software with a standard precision of 53 bits is actually 1.79769×10^{308}. To simplify the calculation, the 3.10 was rewritten as follows

$$\begin{aligned}
\ell_{Int}(\beta) &= \log \int_0^\infty \left\{ \prod_{i=1}^{K} g_i(\beta, \zeta) \right\} \pi(\zeta) d\zeta \\
&= \log \int_0^\infty \left\{ \prod_{i=1}^{K} \int_0^\infty L(\beta, \zeta, \sigma_i)\, \pi(\sigma_i)\, d\sigma_i \right\} \pi(\zeta) d\zeta \\
&= \log \int_0^\infty \left\{ \prod_{i=1}^{K} \int_0^\infty L^*(\beta, \zeta, \sigma_i) \exp(n_i) \pi(\sigma_i)\, d\sigma_i \right\} \pi(\zeta) d\zeta \\
&= \sum_{i=1}^{K} n_i + \log \int_0^\infty \left\{ \prod_{i=1}^{K} \int_0^\infty L^*(\beta, \zeta, \sigma_i)\, \pi(\sigma_i)\, d\sigma_i \right\} \pi(\zeta) d\zeta,
\end{aligned}$$

where $L^*(\beta, \zeta, \sigma_i) = L(\beta, \zeta, \sigma_i) / \exp(n_i)$. In this way, the invariance of the likelihood function with respect to a multiplicative constant was exploited, and the individual

contributions of each trial were of similar magnitude. Moreover, the 'exponential sum logarithm' trick was used to calculate the product where possible, i.e.,

$$\prod_{i}^{K} g_i(\beta, \zeta) = \exp\left[\sum_{i}^{K} \log\{g_i(\beta, \zeta)\}\right].$$

3.6 REML vs integrated likelihood

To illustrate the performance of the integrated maximum likelihood method, a comparison with the data set chosen as an example in Chapter 2 was proposed. The results are compared with the REML method, as this was used by the authors in the original publication Jones et al. (2021). The aim is to assess the validity of the conclusions when uncertainty regarding the precision of the estimate of each study true effect is taken into account. The comparison is proposed only for pairwise meta-analyses and not for meta-regression models because it is not possible to obtain a closed form for τ^2 as a function of ζ^2 (see Section 3.4).

The results of the comparisons are presented in Table 3.1 and Table 3.2. Overall, the integrated likelihood leads to more conservative p-values and wider confidence intervals than the REML. The reason for this is the additional uncertainty introduced considering the uncertainty on the estimation of σ_i^2. It should be noted that most of the studies included in the meta-analyses have a sample size smaller than 20 participants, so the bias of s_i^2 may not be negligible. However, it is worth noting that the conclusion regarding H_0 remained unchanged in each analysis and so, the main conclusions of the authors in the paper are valid. It is worth noting that in the analysis of the 'Relative changes in lower body muscle strength' and 'Relative changes in muscle size' the p-values obtained with the integrated likelihood method are smaller than the one obtained with the REML and so, the confidence intervals are smaller.

A plot depicting the integrated likelihood function for the analysis 'Relative changes in upper body muscle strength' can be seen in Figure 3.1. It is evident that the integrated likelihood takes the form of a convex function. Therefore, it is plausible that we identify an absolute maximum. The same conclusion can be drawn for the plot of each analysis and for these reason, we avoid to report all of them. They are available running the code. Furthermore, from Figure 3.1 it can be noticed that the integrated likelihood has a symmetric shape and so the confidence intervals.

	Integrated Likelihood			REML		
Analysis	$\bar{\beta}$	r_{Int}	p	$\hat{\beta}$	Z	p
Rel. upper	-0.28	-1.40	0.17	-0.29	0.08	0.08
Abs. upper	0.47	2.09	0.04	0.48	0.02	0.02
Rel. lower	-0.21	-3.71	0.02×10^{-2}	-0.21	-3.62	0.03×10^{-2}
Abs.lower	0.34	3.87	0.01×10^{-2}	0.33	4.75	$< 0.01 \times 10^{-2}$
Rel. size	0.10	1.28	0.10	0.10	1.41	0.16
Abs. size	0.45	3.28	0.01×10^{-1}	0.45	3.97	$< 0.01 \times 10^{-2}$

TABLE 3.1: Comparisons of the p-value obtained using the integrated likelihood and REML methods. Abbreviations: Rel. upper: relative changes in upper body muscle strength; Abs. upper: absolute changes in upper body muscle strength; Rel. lower: relative changes in lower body muscle strength; Abs. lower: absolute changes in lower body muscle strength; Rel. muscle: relative changes in muscle size; Abs. muscle: absolute changes in muscle size.

	Integrated Likelihood		REML	
Analysis	$\bar{\beta}$	$(1-\alpha)$ CI	$\hat{\beta}$	$(1-\alpha)$ CI
Rel. upper	-0.28	$(-0.69, 0.16)$	-0.29	$(-0.62, 0.04)$
Abs. upper	0.47	$(0.04, 1.02)$	0.48	$(0.09, 0.88)$
Rel. lower	-0.21	$(-0.30, -0.11)$	-0.21	$(-0.33, -0.10)$
Abs.lower	0.34	$(0.19, 0.48)$	0.33	$(0.20, 0.47)$
Rel. size	0.10	$(-0.06, 0.23)$	0.10	$(-0.04, 0.23)$
Abs. size	0.45	$(0.20, 0.68)$	0.45	$(0.23, 0.67)$

TABLE 3.2: Comparisons of the confidence intervals obtained using the integrated likelihood and REML methods. Abbreviations: Rel. upper: relative changes in upper body muscle strength; Abs. upper: absolute changes in upper body muscle strength; Rel. lower: relative changes in lower body muscle strength; Abs. lower: absolute changes in lower body muscle strength; Rel. muscle: relative changes in muscle size; Abs. muscle: absolute changes in muscle size.

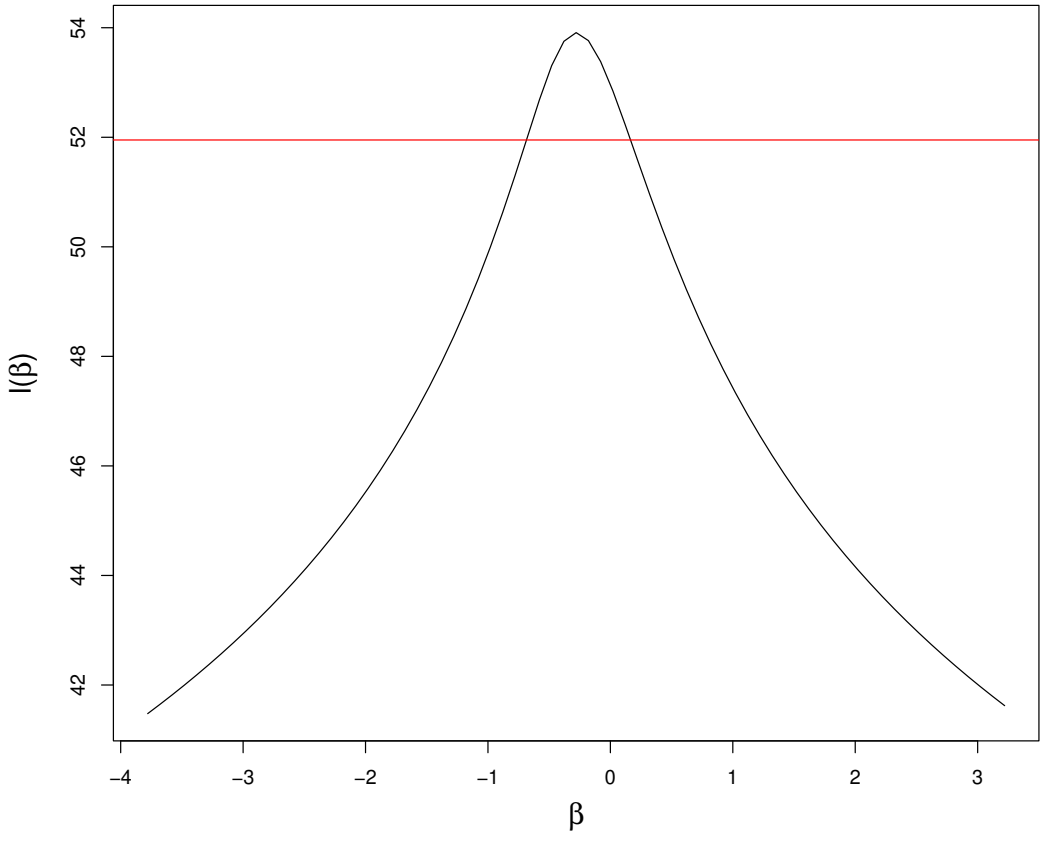


FIGURE 3.1: Integrated likelihood function for meta-analysis regarding relative changes in the upper body muscle strength. The red line indicates the quantile $1-\alpha/2 = 0.975$ of a $\mathcal{N}(0,1)$ distribution. Its intersection with the curve indicates the extremes of a confidence interval of magnitude $1-\alpha = 0.95$.

Chapter 4

Conclusion

4.1 Discussion of the findings

In this work, we analyzed the dataset created by Jones and collegues (Jones et al., 2021) to examine sex differences in hypertrophic and strength adaptations in response to a resistance training program in older adults. We replicated the analysis proposed in the original paper and then applied more advanced methods aimed at relaxing some of the assumptions underlying the models most commonly used in pairwise meta-analysis. In particular, we focused on methods for small sample sizes and presented some alternatives in Chapter 1 when the number of studies included in the meta-analysis is small, while in Chapter 3 we proposed integrated likelihood as a strategy for analyzing datasets where the number of studies is reasonable but the number of participants in each study is small. As shown in Chapter 2, the dataset under study exhibits both the problems. Indeed, the analysis of 'relative and absolute changes in upper body strength' involved 7 studies and 160 subjects, whereas the analysis of 'relative and absolute changes in lower body strength' and 'relative and absolute changes in muscle size' involved 34 and 30 studies and 1196 and 1064 subjects, respectively.

4.1.1 Muscle strength

Females exhibited greater relative increases in lower body strength compared with males, with a standardized mean difference of $\hat{\beta} = -0.21 \quad (95\% CI : -0.33, -0.10)$. However, no sex differences were observed for upper body strength ($\hat{\beta} = -0.29 \quad (95\% CI : -0.62, 0.04)$). Absolute changes in muscle strength were greater in men, with a standardized mean difference of $\hat{\beta} = 0.3312 \quad (95\% CI : 0.19, 0.47)$ and $\hat{\beta} = 0.48 \quad (95\% CI : 0.09, 0.88)$, respectively. These results were obtained using the DerSimonian & Laird

(1986) method. However, none of the other methods tested in Chapter 2 changed the conclusion, although the test based on the Skovgaard statistic and the Hartung-Knapp-Sidik-Jonkman method leads to considerably conservative p-values, especially for upper body muscle strength. The estimates based on integrated likelihood (Table 3.2) do not change the inferential conclusion. However, it can be seen that the results of the analysis on the 'relative and absolute changes in upper body strength' are subject to considerable higher uncertainty. In general, baseline strength is greater in adult males than females, likely due to greater muscle size in males (Bishop et al., 1987) rather than a sex difference in the ability of the nervous system to voluntarily activate the muscles (Molenaar et al., 2013). It is therefore likely that the greater absolute increase in males is a function of their larger body size. However, normalizing for initial strength, it appears that the two sexes make the same progress in the upper body, whereas women progress more rapidly in the lower body. It should be noted, however, that the number of studies included for the upper body is very small and the conclusions should be treated with appropriate caution, especially given the width of the integrated likelihood confidence interval. The problem is also recognize by the authors of the original paper (Jones et al., 2021). However, the non significant difference in upper body strength testing may be related to the accuracy of strength assessment in the original studies. Indeed, the meta-analysis includes only studies using dynamic resistance training, and the most prevalent assessment is the one (or multiple) repetition maximum (i.e., the maximum amount of weight the person can lift for a given number of repetitions). Typically, the minimum load increment that can be used in these tests is 1.25 kg, unless the laboratory is equipped with appropriate 250g and 500g or possibly smaller plates. Since males have higher baseline strength than females, the test can be more accurate and the increment in load represent a smaller percentage of the maximal force of the individual. Note also that the results are at odds with those of an earlier, similar meta-analysis conducted by Roberts and colleagues in young people (Roberts et al., 2020). In this study, women actually appeared to be favored in the upper body, while there were no differences in the lower body and authors attributed the differences to the fact that women were less trained in the upper body than men. In our analysis, it is possible that the difference is due to the more rapid decline in lower limb muscle mass in older men than in women (Cheng et al., 2014), which could potentially impair the ability to adapt to a resistance training protocol. For the analysis of the upper body, it was considered inappropriate to perform a meta-regression because of the limited number of included studies. In the analysis of 'relative changes in lower body muscle strength', the estimated value of τ^2 was found to be $\hat{\tau}^2 < 10^{-4}$, although the 95% confidence

interval was $(< 10^{-4}, 0.11)$ according to the method of Viechtbauer (2007). Analysis of 'absolute changes in lower body muscle strength' yielded an estimated value for τ^2 of $\hat{\tau}^2 = 0.03$ with a 95% confidence interval of $(< 10^{-4}, 0.49)$. The uncertainty in determining the presence of heterogeneity and the intent to replicate the analyses of the original work necessitated meta-regression. Effect size was found to vary with weekly volume performed. Specifically, when analyzing relative changes in lower body muscle strength, the effect size decreases by -0.07×10^{-2} for each additional repetition per week. This result remains consistent after excluding influential studies and remains significant regardless of correction for the small sample applied. Interestingly, it is noteworthy that the intercept of the model is not significantly different from zero, regardless of the correction used. It seems apparent that there are no sex differences in protocols with a low weekly volume, while adjustments to protocols with a high number of repetitions seem to favor female individuals. The same effect is seen in the analysis of relative changes in lower body muscle strength, where the effect size for each repetition decreases by -0.09×10^{-2} $(95\%CI : -0.14 \times 10^{-2}, -0.04 \times 10^{-2})$; however, in this case, the intercept is significantly different from zero, 0.64 $(95\%CI : 0.45, 0.83)$. The results remain fairly consistent regardless of the correction applied. The conclusion of this section differs in part from that proposed in the original Jones et al. (2021) paper, as it appears that the sex differences in force adaptation compared to baseline are related to the type of protocol proposed and are only evident in high volume training protocols. It is also worth noting that the slope of the regression line has only a small magnitude, since the number of weekly repetitions has a wide variation, ranging from 60 to 1071, and it should also be noted that all protocols involving the lower body have less than 350 repetitions. Finally, it is apparent from looking at the figures in Chapter 2 that all of the studies with a high number of repetitions have low weight. A possible explanation for this phenomenon could be related to the fact that males start with a higher level of strength not only because of their body size, but also because they are more likely to perform tasks that require the use of strength in their daily lives. Thus, because they are more accustomed to expressing high levels of force, they must expose themselves to heavier loads in order to improve, and these can only be utilized if the number of repetitions is lower. Women, on the other hand, starting from a lower level of training, improve even with lower loads and therefore with protocols that provide higher repetitions. Of course, this is pure speculation, as there is no evidence to confirm or disprove this theory.

4.1.2 Muscle size

The analysis of the 'relative changes in muscle size' yielded an estimate of $\beta = 0.10$ with a 95% confidence interval of $(-0.04, 0.23)$, whereas the analysis of the 'absolute changes in muscle size' yielded an estimate of $\beta = 0.45 \quad (95\%CI : 0.23, 0.67)$. These results indicate that there is no significant difference in relative terms, but in absolute terms males are favored. The findings are supported by the fact that the extent of hypertrophic adaptations appears to be primarily related to the presence of androgen receptors rather than circulating levels of systemic hormones such as testosterone (Morton et al., 2018). However, the presence of these receptors is not altered by exercise, and there are no studies examining whether their expression differs in men and women. Furthermore, hypertrophic adaptations are associated with protein synthesis and signaling systems that do not differ between the sexes (Dreyer et al., 2010). These considerations could offer an explanation of why relative hypertrophic adaptations are the same between the sexes. However, because males have a larger body size, the differences observed in absolute adaptations are easily explained. It should be noted, however, that the cited studies were conducted in a population of young adults and that there is no evidence to confirm these results in elderly.

Both proposed models have a substantial amount of heterogeneity, with τ^2 in the relative differences model estimated to be $\hat{\tau}^2 = 0.01$ $(95\%CI :< 10^{-4}, 0.1730)$ and $\hat{\tau}^2 = 19.87$ $(95\%CI :< 10^{-4}, 0.49)$. This corresponds to 10.43% and 62.12% of the variance attributed to heterogeneity, respectively. For this reason, we perfomed a meta-regression. Regarding the 'relative changes in muscle size', the regression model includes duration and weekly volume. Notably, the intercept is not significantly different from zero and the effect favors women for each additional weekly repetition $(\hat{\beta_1} = -0.07 \times 10^{-2}, p = 0.05 \times 10^{-1})$, while it seems to favor men for each additional week of study duration $(\hat{\beta_1} = 0.02, p = 0.02)$. The heterogeneity of the meta-regression model is estimated as $\hat{\tau}^2 = 0(95\%CI :< \times 10^{-4}, 0.17)$. Thus, although the point estimate decreases substantially, the uncertainty behind the variance estimate remains the same. The results remain significant under all proposed correction methods, except for the coefficient β_2 when using the "Hartung-Knapp-Sidik-Jonkman" method. This is not surprising, considering that the number of study in this model is substantial, while the real issue is the number of participants in each study. The effect of weekly volume can be explained by the same rationale used for the effect on muscle strength. If we assume that women are less inclined (i.e., "trained") to perform strength exercises, we can assume that they already benefit from shorter training periods. From this perspective, the

effect tends to balance out in studies with longer duration, which explains the positive coefficient β_2.

Regarding absolute changes in muscle size, the regression model is more complex. In fact, it includes the training type, study duration, weekly volume, and the interaction between training type and weekly volume. Specifically, the expected value of the effect size for 'absolute changes in muscle size' is $-0.31 - 0.06 \times 10^{-2} \cdot \text{metareg\_weekly\_repetitions} + 0.04 \cdot \text{duration}$, when the study used a 'full body body' exercise protocol, while it is $1.03 - 0.05 \times 10^{-1} \cdot \text{metareg\_weekly\_repetitions} + 0.04 \cdot \text{duration}$ for protocols involving only the 'lower body'. Thus, the effect of number of repetitions appears to be much more pronounced for training protocols that focus on the 'lower body'. However, it should be noted that none of these protocols exceed 350 repetitions per week, so the result should be interpreted with caution. In addition, the intercepts are particularly different, and this could be a result of the hierarchy used to choose the effect measures. It should also be noted that the coefficient for the fixed effect of `metareg_weekly_repetitions` is not significantly different from zero for any of the methods used. However, it is retained in the model because it was chosen to keep the interaction with the variable `training`. The interpretation of this model is not different from that of the previous model for 'relative changes in muscle size', once we acknowledge that males have bigger body size.

4.2 Limitations

This work is not without limitations. Several methods have been proposed to correct for a small number of studies in a meta-analysis and the integrated likelihood for small sample sizes within individual studies. To the best of our knowledge, there is currently no method that addresses both problems simultaneously. Unfortunately, the analyses on 'upper body muscle strength' were conducted with a sample of 7 studies and a total of 160 participants, which means an average of 22.86 participants per study. However, several studies have a sample size of less than 15 participants. In this context, it would be optimal to propose a correction that takes into account the 'double low sample size' problem.

In addition, as mentioned in Chapter 2, some included studies have multiple effect measures. The authors used hierarchy to reduce dimensionality. Originally, there were two sources of multiple effect measures in the same study: the fact that more than one outcome was measured to represent the same construct, and the presence of multiple intervention groups. And so, it would have been more informative to have access to all effect sizes in order to use a hierarchical multilevel model. In the present analysis, it

would have been possible to account for multiple groups within a single study by using a three-level model rather than assuming independent effect measures.
Finally, muscle strength and muscle size are expected to be highly correlated, especially in older adults. For this reason, it would have been interesting to use a multivariate meta-analysis model. However, this was not possible because not all studies reported on all three effect measures and, consequently, the sample size included in each analysis varied.

Appendix

Appendix 1

```
#' This function performs multiple meta-analyses using different methods and
#' returns various results, including estimator estimates, confidence intervals,
#' p-values, and more.
#'
#' @param y_i      A numeric vector of effect sizes for each study.
#' @param si2      A numeric vector of sampling variances for each study.
#' @param X        An optional matrix or dataframe of moderators.
#'                 Defaults to NULL. Interaction must be included as a new
#'                 column. Intercept is included by default.
#' @param alpha    The desired level of confidence/significance for confidence
#'                 intervals and test. Defaults to 0.05.
#' @param B        The number of permutations for the permutation test.
#'                 Defaults to 2000.
#'
#' @return A list containing various results:
#'          - a dataframe for the estimate of \beta and its standard error;
#'          - a dataframe for the estimate of \tau^2 and its standard error;
#'          - a dataframe for the confidence intervals of \beta;
#'          - a dataframe for the confidence intervals of \tau^2;
#'          - a dataframe for thesignificance tests on \beta;
#'
#' @examples
#' # Example 1: Perform meta-analyses without moderators
#' y_i <- c(0.5, 0.8, 1.2)
#' si2 <- c(0.1, 0.2, 0.3)
#' results <- meta_many(y_i, si2
#'
#' # Example 2: Perform meta-analyses with moderators
#' # Generate example data
#' set.seed(123)
#' y_i <- rnorm(10, 0.5, 0.2)
#' si2 <- runif(10, 0.1, 0.3)
#' X <- data.frame(Moderator1 = rnorm(10), Moderator2 = rnorm(10))
#'
#' Perform meta-analyses
#' results <- meta_many(y_i, si2, X)


meta_many <- function(y_i, si2, X = NULL,
                      alpha = 0.05, B=2000){
```

```
##############################################################################
#Estimation
##############################################################################

# Argument list
argument_list <- list(yi = y_i, vi = si2)

# Moderators
if (!is.null(X)) {
  argument_list$mod <- X
}

# Calls metafor rma
methods <- c("DL", "PM", "ML", "REML", "SJ")
fitted_obj <-list()

argument_list_tmp <- argument_list
for (i in 1:length(methods)){
  argument_list_tmp$method <- methods[i]
  fitted_obj[[i]] <- do.call(rma, argument_list_tmp)
}

names(fitted_obj) <- methods

#Add method = GENQ
argument_list_tmp$method <- "GENQ"
argument_list_tmp$weights <- argument_list_tmp$vi
fitted_obj$GENQ <- do.call(rma, argument_list_tmp)

#Dataframe for metalik and metatest

# ncol X
if(!is.null(X)){
  if(!is.null(ncol(X))){
    mod_num <- ncol(X)
  } else {
    mod_num <- 1
  }
}

if (!is.null(X)) {
    tmp_df <- data.frame(do.call(cbind, argument_list))
    colnames(tmp_df) <- c("yi", "vi", paste("v", 1:mod_num, sep="_"))

    f <- paste("yi ~ ",
               paste(grep("v_", x = colnames(tmp_df), value = TRUE),
                     collapse = " + "))
} else {
  tmp_df <- data.frame(do.call(cbind, argument_list))
  f <- "yi ~ 1"
}

#Add metalik
fitted_obj$metaLik_fit  <- metaLik(as.formula(f),
```

```
                                    sigma2 = vi,
                                    data = tmp_df)

#Add metatest
fitted_obj$metatest_fit <- metatest(as.formula(f),
                                     variance = vi,
                                     npermut = B,
                                     data = tmp_df)

###############################################################################
#Results
###############################################################################

#Results
res_list <-list()

###############################################################################
#Estimators Beta and Tau^2
###############################################################################

res_list$Betas <- as.data.frame(cbind(fitted_obj$DL$beta,
                                      fitted_obj$DL$se,
                                      fitted_obj$ML$beta,
                                      fitted_obj$ML$se,
                                      fitted_obj$REML$beta,
                                      fitted_obj$REML$se,
                                      fitted_obj$PM$beta,
                                      fitted_obj$PM$se))

colnames(res_list$Betas) <- c("DL_Estimate", "DL_se",
                              "ML_Estimate", "ML_se",
                              "REML_Estimate", "REML_se",
                              "PM_Estimate", "PM_se")

res_list$Tau2s <- as.data.frame(cbind(fitted_obj$DL$tau2,
                                      fitted_obj$DL$se.tau2,
                                      fitted_obj$ML$tau2,
                                      fitted_obj$ML$se.tau2,
                                      fitted_obj$REML$tau2,
                                      fitted_obj$REML$se.tau2,
                                      fitted_obj$PM$tau2,
                                      fitted_obj$PM$se.tau2))

colnames(res_list$Tau2s) <- c("DL_Estimate", "DL_se",
                              "ML_Estimate", "ML_se",
                              "REML_Estimate", "REML_se",
                              "PM_Estimate", "PM_se")

###############################################################################
#CI - Beta
###############################################################################
#Skovgaard CI
if (!is.null(X)) {
  s_confint <- list()
```

```
  for(p in 1:(mod_num + 1)){
    prof_tmp <- metaLik:::.profLik(fitted_obj$metaLik_fit, param = p)
    s_confint[[p]] <- predict(prof_tmp$smooth.rs.inv, x=c(-1.96, 1.96))$y

  }
  s_confint <- do.call(rbind, s_confint)
  s_confint <- as.data.frame(s_confint)

} else {
  prof <- metaLik:::.profLik(fitted_obj$metaLik_fit)
  s_confint <- predict(prof$smooth.rs.inv, x=c(-1.96, 1.96))$y
  s_confint <- as.data.frame(matrix(s_confint, ncol = 2))
}

colnames(s_confint) <- c("Upper", "Lower")

#Lower bound before
res_list$Beta_CIs <- as.data.frame(cbind(
  fitted_obj$DL$beta - qnorm(1-alpha/2) * fitted_obj$DL$se, #Wald DL
  fitted_obj$DL$beta + qnorm(1-alpha/2) * fitted_obj$DL$se,
  fitted_obj$ML$beta - qnorm(1-alpha/2) * fitted_obj$ML$se, #Wald ML
  fitted_obj$ML$beta + qnorm(1-alpha/2) * fitted_obj$ML$se,
  fitted_obj$REML$beta - qnorm(1-alpha/2) * fitted_obj$REML$se, #Wald REML
  fitted_obj$REML$beta + qnorm(1-alpha/2) * fitted_obj$REML$se,
  confint(fitted_obj$ML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 2], #ML prof lik
  confint(fitted_obj$ML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 3],
  confint(fitted_obj$REML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 2], #REML prof lik
  confint(fitted_obj$REML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 3],
  confint(fitted_obj$SJ, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 2],  #Hartung-Knapp-Sidik-Jonkman
  confint(fitted_obj$SJ, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$fixed[, 3],
  s_confint$Lower, s_confint$Upper))

colnames(res_list$Beta_CIs) <- c("DL_Wald_LL", "DL_Wald_UL",
                                 "ML_Wald_LL", "ML_Wald_UL",
                                 "REML_Wald_LL", "REML_Wald_UL",
                                 "ML_prof_Wilks_LL", "ML_prof_Wilks_UL",
                                 "REML_prof_Wilks_LL", "REML_prof_Wilks_UL",
                                 "Hartung-Knapp-Sidik-Jonkman_LL",
                                 "Hartung-Knapp-Sidik-Jonkman_UL",
                                 "prof_Wilks_Skovgaard_LL",
                                 "prof_Wilks_Skovgaard_UL")

#############################################################################
#CI - Tau^2
#############################################################################
#Lower bound before
res_list$Tau2_CIs <- as.data.frame(cbind(
  fitted_obj$ML$tau2 - qnorm(1-alpha/2) * fitted_obj$ML$se.tau2,   #Wald ML
  fitted_obj$ML$tau2 + qnorm(1-alpha/2) * fitted_obj$ML$se.tau2,
```

```
  confint(fitted_obj$ML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$random[1, 2], #prof lik ML
  confint(fitted_obj$ML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$random[1, 3],
  confint(fitted_obj$REML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$random[1, 2], #prof lik REML
  confint(fitted_obj$REML, random = TRUE, fixed = TRUE, type = "PL",
          level = alpha)$random[1, 3],
  confint(fitted_obj$GENQ, random = TRUE, fixed = TRUE, type = "GENQ",
          level = alpha)$random[1, 2], ##Generalized Q method (Viechtbauer)
  confint(fitted_obj$GENQ, random = TRUE, fixed = TRUE, type = "GENQ",
          level = alpha)$random[1, 3]))

colnames(res_list$Tau2_CIs) <- c("ML_Wald_LL", "ML_Wald_UL",
                                 "ML_prof_Wilks_LL", "ML_prof_Wilks_UL",
                                 "REML_prof_Wilks_LL", "REML_prof_Wilks_UL",
                                 "GENQ_LL", "GENQ_UL")

##############################################################################
#Pvalue -Fixed effects
##############################################################################

res_list$Beta_pvals <- as.data.frame(cbind(
  #DL
  fitted_obj$REML$pval,
  #Hartung-Knapp-Sidik-Jonkman
  fitted_obj$SJ$pval,
  #Wald (ML)
  fitted_obj$ML$pval,
  #Wald (REML)
  fitted_obj$REML$pval,
  #Signed profile loglik (ML)
  if(!is.null(X)){
    sapply(1:(mod_num+1), function(p)
      test.metaLik(fitted_obj$metaLik_fit, param = p, print = FALSE)$pvalue.r)
  } else {
      test.metaLik(fitted_obj$metaLik_fit, param = 1, print = FALSE)$pvalue.r
  },
  #Skovgaard
  if(!is.null(X)){
    sapply(1:(mod_num+1), function(p)
      test.metaLik(fitted_obj$metaLik_fit,
                   param = p, print = FALSE)$pvalue.rskov)
  } else {
      test.metaLik(fitted_obj$metaLik_fit,
                   param = 1, print = FALSE)$pvalue.rskov
  },
  #Bartlett's correction
  fitted_obj$metatest_fit$pBartlett,
  #Permutation test
  fitted_obj$metatest_fit$ppermtest))

colnames(res_list$Beta_pvals) <- c("DL", "Hartung-Knapp-Sidik-Jonkman",
                                   "ML_Wald", "REML_Wald", "ML_prof_Wilks",
                                   "Skovgaard", "Bartlett", "Permutations")
```

```
  ##############################################################################
  #Return
  ##############################################################################
  return(res_list)
}
```

Appendix

Appendix 2

```
'
Title:       Meta-analysis Jones et al. 2021
Description: Reproduction of the analyis of Jones et al. 2021 (*) and
             integration with advance methods in meta-analysis.
Author:      Enrico Roma

*            Jones, M. D., Wewege, M. A., Hackett, D. A., Keogh,
             J. W., & Hagstrom, A. D. (2021). Sex differences in adaptations
             in muscle strength and size following resistance training in
             older adults: A systematic review and meta-analysis.
             Sports Medicine, 51, 503-517.
'


# Workspace
rm(list=ls())
setwd("C:\\Users\\romae\\Desktop\\Analisi_tesi_Statistica")
image_path <-
  "C:\\Users\\romae\\Desktop\\Analisi_tesi_Statistica\\Grafici_tesi"

# Libraries
# install.packages(c("metafor", "metaLik", "metatest", "esc",
#                    "dplyr", "ggplot2", "readxl",
#                    "wrapr", "tidyverse"))

library(esc)
library(metafor)
library(metaLik)
library(metatest)
library(readxl)
library(dplyr)
library(xtable)
library(tidyverse)

# Hand written function
source("meta_many.R")

#Overall parameter setting
Alpha <- 0.05 #Sign. level
```

```
# Data
df <- read_xlsx("Data_190620.xlsx",
                sheet="Data_older",
                skip = 0,
                col_names=TRUE)

df <- as.data.frame(df)

round_df <- function(df, digits) {
  nums <- vapply(df, is.numeric, FUN.VALUE = logical(1))
  df[,nums] <- round(df[,nums], digits = digits)
  (df)
}

df <- round_df(df, digits = 2)

################################################################################
## Analysis 1a: Relative change upper body muscle strength
################################################################################

# Filter
df %>%
  filter(Age == "Older", Upper_strength == 1) -> df_upper

View(df_upper)

# Studies included
df_upper %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_upper %>%
  summarise(Upper_males = sum(Male_upper_strength_post_number),
            Upper_females = sum(Female_upper_strength_post_number))


# Random effect meta-analysis
(rel_upper_rma <- rma(data = df_upper,
                      m1i = Male_upper_strength_relative_change_mean,
                      sd1i = Male_upper_strength_relative_change_sd,
                      n1i = Male_upper_strength_post_number,
                      m2i = Female_upper_strength_relative_change_mean,
                      sd2i = Female_upper_strength_relative_change_sd,
                      n2i = Female_upper_strength_post_number,
                      weighted = TRUE,
                      method = "REML",
                      measure = "SMD",
                      slab = paste(Study, Reference, sep = " ")))

# Decrease margins so the full space is used
pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_rma.pdf", sep="\\"),
    width = 12, height = 8)
par(mar = c(4,4,1,2))

# forest plot
```

```
forest(rel_upper_rma,
       addfit = FALSE,
       addcred = FALSE,
       showweights = TRUE,
       xlim = c(-14, 6),
       ylim = c(-1, 9.5),
       at = c(-2.5, 0, 2.5),
       rows = 6.5:0.5,
       ilab = round(cbind(
         df_upper$Male_upper_strength_relative_change_mean,
         df_upper$Male_upper_strength_relative_change_sd,
         df_upper$Male_upper_strength_post_number,
         df_upper$Female_upper_strength_relative_change_mean,
         df_upper$Female_upper_strength_relative_change_sd,
         df_upper$Female_upper_strength_post_number),
         digits = 2),
       ilab.xpos = c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(rel_upper_rma, row = -0.5, cex = 0.8, mlab = "")
text(6, -0.8, paste("p =", formatC(rel_upper_rma$pval, digits = 4)),
     pos = 2, cex = 0.8)

rel_upper_fsn <- fsn(yi = rel_upper_rma$yi,
                     vi = rel_upper_rma$vi,
                     type = "Rosenthal",
                     alpha = 0.05)

text(-14, -1, pos = 4, cex = 0.8,
     paste("Fail-safe N =", formatC(rel_upper_fsn$fsnum)))

op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      7.25,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       7.75,
     c("Male", "Female"))
text(-14,                             7.75, "Study",  pos = 4)
text(6,                               7.75, "Hedges' g [95% CI]", pos = 2)
text(0,                               7.75,
     "Relative change in upper body muscle strength")

text(-14, -0.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
                  .(formatC(rel_upper_rma$QE, digits=2,
                            format="f")),
                  ", df = ", .(rel_upper_rma$k - rel_upper_rma$p),
                  ", p = ",.(formatC(rel_upper_rma$QEp, digits=2,
                                     format="f")),
                  "; ", I^2, " = ",.(formatC(rel_upper_rma$I2,digits=1,
                                             format="f")),
                  "%)")))
dev.off()
```

```
# Assumption checks
rel_upper_inf <- influence(rel_upper_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_Normality.pdf", sep="\\"),
    width = 12, height = 8)

qqnorm(rel_upper_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(rel_upper_inf$inf$rstudent)

dev.off()

shapiro.test(rel_upper_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_upper_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(rel_upper_inf$p/(rel_upper_inf$k-rel_upper_inf$p
3 * sqrt(rel_upper_inf$p/(rel_upper_inf$k-rel_upper_inf$p))

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_upper_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (rel_upper_inf$p +1))
qchisq(p = 0.5, df = (rel_upper_inf$p +1))

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_Cook.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_upper_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_upper_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)
```

```
plot(rel_upper_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_upper_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Alternative methods

# Effect sizes calculations
escalc(measure = "SMD",
       data = df_upper,
       m1i = Male_upper_strength_relative_change_mean,
       sd1i = Male_upper_strength_relative_change_sd,
       n1i = Male_upper_strength_post_number,
       m2i = Female_upper_strength_relative_change_mean,
       sd2i = Female_upper_strength_relative_change_sd,
       n2i = Female_upper_strength_post_number, append = TRUE) -> df_upper

# Meta many
set.seed(23091994)
rel_upper_meta_many <- meta_many(y_i = df_upper$yi,
                                 si2 = df_upper$vi,
                                 X = NULL, alpha = Alpha, B = 2000)

# Betas
rel_upper_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
rel_upper_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Betas p-val
rel_upper_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 4)


# Tau2s
```

```
rel_upper_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
rel_upper_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(rel_upper_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_rel_upper_rma <- regtest(rel_upper_rma))

#variance
(Egger_rel_upper_rma_va <- regtest(rel_upper_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(rel_upper_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)
```

```
# se
lines(coef(Egger_rel_upper_rma$fit)[1] +
        coef(Egger_rel_upper_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_rel_upper_rma_va$fit)[1] +
        coef(Egger_rel_upper_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
(rel_upper_trim_fill <- trimfill(rel_upper_rma))

pdf(paste(image_path, "Rel_Upper_Chp2\\rel_upper_TrimFill.pdf", sep="\\"),
    width = 12, height = 8)

funnel(rel_upper_trim_fill,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

##############################################################################
## Analysis 1b: Absolute change upper body muscle strength
##############################################################################

# Random effect meta-analysis
(abs_upper_rma <- rma(data = df_upper,
                      m1i = Male_upper_strength_absolute_change_mean,
                      sd1i = Male_upper_strength_absolute_change_sd,
                      n1i = Male_upper_strength_post_number,
                      m2i = Female_upper_strength_absolute_change_mean,
                      sd2i = Female_upper_strength_absolute_change_sd,
                      n2i = Female_upper_strength_post_number,
                      weighted = TRUE,
                      method = "REML",
                      measure = "SMD",
                      slab = paste(Study, Reference, sep = " ")))

# Decrease margins so the full space is used
pdf(paste(image_path, "Abs_Upper_Chp2\\abs_upper_rma.pdf", sep="\\"),
    width = 12, height = 8)
par(mar = c(4,4,1,2))

# forest plot
```

```
forest(abs_upper_rma,
       addfit = FALSE,
       addcred = FALSE,
       showweights = TRUE,
       xlim = c(-14, 6),
       ylim = c(-1, 9.5),
       at = c(-3, 0, 3),
       rows = 6.5:0.5,
       ilab = round(cbind(df_upper$Male_upper_strength_absolute_change_mean,
                          df_upper$Male_upper_strength_absolute_change_sd,
                          df_upper$Male_upper_strength_post_number,
                          df_upper$Female_upper_strength_absolute_change_mean,
                          df_upper$Female_upper_strength_absolute_change_sd,
                          df_upper$Female_upper_strength_post_number),
                    digits = 2),
       ilab.xpos = c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(abs_upper_rma, row = -0.5, cex = 0.8, mlab = "")
text(6, -0.8, paste("p =", formatC(abs_upper_rma$pval, digits = 4)),
     pos = 2, cex = 0.8)

abs_upper_fsn <- fsn(yi = abs_upper_rma$yi, vi = abs_upper_rma$vi,
                     type = "Rosenthal", alpha = 0.05)
text(-14, -1, pos = 4, cex = 0.8, paste("Fail-safe N =",
                                        formatC(abs_upper_fsn$fsnum)))

op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      7.25,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       7.75, c("Male", "Female"))
text(-14,                             7.75, "Study",  pos = 4)
text(6,                               7.75, "Hedges' g [95% CI]", pos = 2)
text(0,                               7.75,
     "Absolute change in upper body muscle strength")

text(-14, -0.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
                  .(formatC(abs_upper_rma$QE, digits=2, format="f")),
                  ", df = ",
                  .(abs_upper_rma$k - abs_upper_rma$p),
                  ", p = ",
                  .(formatC(abs_upper_rma$QEp, digits=2, format="f")),
                   "; ", I^2, " = ",
                   .(formatC(abs_upper_rma$I2, digits=1, format="f")),
                   "%)")))
dev.off()
# Assumption checks
abs_upper_inf <- influence(abs_upper_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_Normality.pdf", sep="\\"),
    width = 12, height = 8)
```

```
qqnorm(abs_upper_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(abs_upper_inf$inf$rstudent)

dev.off()

shapiro.test(abs_upper_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotinf = 1, cex = 1.5)
dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(abs_upper_inf$p/(abs_upper_inf$k-abs_upper_inf$p
3 * sqrt(abs_upper_inf$p/(abs_upper_inf$k-abs_upper_inf$p))

pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (abs_upper_inf$p +1))
qchisq(p = 0.5, df = (abs_upper_inf$p +1))

pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_Cook.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas
```

```
pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_upper_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Significance removing "Lexell et al."
df_upper %>%
  subset(Study != "Lexell et al.") %>%
  rma(data = .,
      m1i = Male_upper_strength_absolute_change_mean,
      sd1i = Male_upper_strength_absolute_change_sd,
      n1i = Male_upper_strength_post_number,
      m2i = Female_upper_strength_absolute_change_mean,
      sd2i = Female_upper_strength_absolute_change_sd,
      n2i = Female_upper_strength_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Alternative methods

# Effect sizes calculations
escalc(measure = "SMD",
       data = df_upper,
       m1i = Male_upper_strength_absolute_change_mean,
       sd1i = Male_upper_strength_absolute_change_sd,
       n1i = Male_upper_strength_post_number,
       m2i = Female_upper_strength_absolute_change_mean,
       sd2i = Female_upper_strength_absolute_change_sd,
       n2i = Female_upper_strength_post_number, append = TRUE) -> df_upper

# Meta many
set.seed(23091994)
abs_upper_meta_many <- meta_many(y_i = df_upper$yi,
                                 si2 = df_upper$vi,
                                 X = NULL, alpha = Alpha, B = 2000)

# Betas
abs_upper_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
abs_upper_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)
```

```
# Betas p-val
abs_upper_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 4)


# Tau2s
abs_upper_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
abs_upper_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot

pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(abs_upper_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_abs_upper_rma <- regtest(abs_upper_rma))

#variance
(Egger_abs_upper_rma_va <- regtest(abs_upper_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(abs_upper_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
```

```
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)

# se
lines(coef(Egger_abs_upper_rma$fit)[1] +
        coef(Egger_abs_upper_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_abs_upper_rma_va$fit)[1] +
        coef(Egger_abs_upper_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
(abs_upper_trim_fill <- trimfill(abs_upper_rma))

pdf(paste(image_path, "abs_Upper_Chp2\\abs_upper_TrimFill.pdf", sep="\\"),
    width = 12, height = 8)

funnel(abs_upper_trim_fill,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

################################################################################
## Analysis 2a: Relative change lower body muscle strength
################################################################################

# Filter
df %>%
  filter(Age == "Older", Lower_strength == 1) -> df_lower

# Studies included
df_lower %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_lower %>%
  summarise(lower_males = sum(Male_lower_strength_post_number),
```

```
            lower_females = sum(Female_lower_strength_post_number))


# Random effect meta-analysis
(rel_lower_rma <- rma(data = df_lower,
                      m1i = Male_lower_strength_relative_change_mean,
                      sd1i = Male_lower_strength_relative_change_sd,
                      n1i = Male_lower_strength_post_number,
                      m2i = Female_lower_strength_relative_change_mean,
                      sd2i = Female_lower_strength_relative_change_sd,
                      n2i = Female_lower_strength_post_number,
                      weighted = TRUE,
                      method = "REML",
                      measure = "SMD",
                      slab = paste(Study, Reference, sep = " ")))

# Decrease margins so the full space is used
pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_rma.pdf", sep="\\"),
    width = 12, height = 8)
par(mar = c(4,4,1,2))

# forest plot
forest(rel_lower_rma,
       addfit = FALSE,
       showweights = TRUE,
       xlim=c(-15, 6.5),
       ylim = c(-2, 37.5),
       at = c(-3, 0, 3),
       ilab = round(cbind(df_lower$Male_lower_strength_relative_change_mean,
                          df_lower$Male_lower_strength_relative_change_sd,
                          df_lower$Male_lower_strength_post_number,
                          df_lower$Female_lower_strength_relative_change_mean,
                          df_lower$Female_lower_strength_relative_change_sd,
                          df_lower$Female_lower_strength_post_number),
                    digits = 2),
       ilab.xpos=c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
       rows = 34:0,
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(rel_lower_rma, row = -1.5, cex = 0.8, mlab = "")
text(6.5, -2.5, paste("p =", formatC(rel_lower_rma$pval, digits = 1)),
     pos = 2, cex = 0.8)

rel_lower_fsn <- fsn(yi = rel_lower_rma$yi, vi = rel_lower_rma$vi,
                     type = "Rosenthal", alpha = 0.05)
text(-15, -2.5, pos = 4, cex = 0.8, paste("Fail-safe N =",
                                          formatC(rel_lower_fsn$fsnum)))

op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      35,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       36, c("Male", "Female"))
text(-15,                             36, "Study",  pos = 4)
```

```
text(6.5,                              36, "Hedges' g [95% CI]", pos = 2)
text(0,                                36,
     "Relative change in lower body muscle strength")

text(-15, -1.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
     .(formatC(rel_lower_rma$QE, digits=2, format="f")),
     ", df = ",
     .(rel_lower_rma$k - rel_lower_rma$p),
     ", p = ",
     .(formatC(rel_lower_rma$QEp, digits=2, format="f")),
    "; ", I^2, " = ",
    .(formatC(rel_lower_rma$I2, digits=1, format="f")),
    "%)")))

dev.off()

# Assumption checks
rel_lower_inf <- influence(rel_lower_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_Normality.pdf", sep="\\"),
    width = 12, height = 8)

qqnorm(rel_lower_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(rel_lower_inf$inf$rstudent)

dev.off()

shapiro.test(rel_lower_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_lower_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(rel_lower_inf$p/(rel_lower_inf$k-rel_lower_inf$p
3 * sqrt(rel_lower_inf$p/(rel_lower_inf$k-rel_lower_inf$p))

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_lower_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (rel_lower_inf$p +1))
qchisq(p = 0.5, df = (rel_lower_inf$p +1))

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_Cook.pdf", sep="\\"),
    width = 12, height = 8)
```

```
plot(rel_lower_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_lower_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_lower_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_lower_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Check removing "Sood et al." & "Walts et al."
df_lower %>%
  subset(Study != "Sood et al." & Study != "Walts et al.") %>%
  rma(data = .,
      m1i = Male_lower_strength_relative_change_mean,
      sd1i = Male_lower_strength_relative_change_sd,
      n1i = Male_lower_strength_post_number,
      m2i = Female_lower_strength_relative_change_mean,
      sd2i = Female_lower_strength_relative_change_sd,
      n2i = Female_lower_strength_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Alternative methods

# Effect sizes calculations
escalc(measure = "SMD",
       data = df_lower,
       m1i = Male_lower_strength_relative_change_mean,
       sd1i = Male_lower_strength_relative_change_sd,
       n1i = Male_lower_strength_post_number,
```

```
       m2i = Female_lower_strength_relative_change_mean,
       sd2i = Female_lower_strength_relative_change_sd,
       n2i = Female_lower_strength_post_number, append = TRUE) -> df_lower

# Meta many
set.seed(23091994)
rel_lower_meta_many <- meta_many(y_i = df_lower$yi,
                                 si2 = df_lower$vi,
                                 X = NULL, alpha = Alpha, B = 2000)

# Betas
rel_lower_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
rel_lower_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  slice(-length(Estimator)) %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Betas p-val
rel_lower_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 4)


# Tau2s
rel_lower_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
rel_lower_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(rel_lower_rma,
       level=c(90, 95, 99),
```

```
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_rel_lower_rma <- regtest(rel_lower_rma))

#variance
(Egger_rel_lower_rma_va <- regtest(rel_lower_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(rel_lower_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)

# se
lines(coef(Egger_rel_lower_rma$fit)[1] +
        coef(Egger_rel_lower_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_rel_lower_rma_va$fit)[1] +
        coef(Egger_rel_lower_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
(rel_lower_trim_fill <- trimfill(rel_lower_rma))

pdf(paste(image_path, "Rel_Lower_Chp2\\rel_lower_TrimFill.pdf", sep="\\"),
    width = 12, height = 8)

funnel(rel_lower_trim_fill,
```

```
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Meta - regression models

# training type
df_lower %>% group_by(training) %>% summarise(n())


df_lower %>%
  ggplot(aes(x = training, y = yi)) +
  geom_boxplot(alpha = 0.80, outlier.shape = NA) +
  geom_point(aes(size = (1/vi)/sum(1/vi), fill = training),
             shape = 21, position = position_jitterdodge()) +
  xlab("Training type") +
  ylab("SMD") +
  labs(fill = "Training type", size = "CE weights")


ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_training.pdf",
             sep="\\"), dpi = 600)

# Study duration
df_lower %>%
  ggplot(aes(x = duration, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_duration.pdf",
             sep="\\"), dpi = 600)

# Weekly Volume
df_lower %>%
  ggplot(aes(x = metareg_weekly_repetitions, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_volume.pdf",
             sep="\\"), dpi = 600)

# Intensity
df_lower %>%
  ggplot(aes(x = metareg_intensity, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_intensity.pdf",
             sep="\\"), dpi = 600)
```

```
# Frequency
df_lower %>%
  ggplot(aes(x = metareg_frequency, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_frequency.pdf",
             sep="\\"), dpi = 600)

# N_exercise
df_lower %>%
  ggplot(aes(x = metareg_exercises, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_exercise.pdf",
             sep="\\"), dpi = 600)

# Sets (per exercise)
df_lower %>%
  ggplot(aes(x = metareg_sets, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_exp_sets.pdf",
             sep="\\"), dpi = 600)

# Metaregression Model
(rel_lower_metareg_1 <- rma(data = df_lower,
                            mods = ~ training * metareg_weekly_repetitions +
                              duration + metareg_intensity + metareg_frequency,
                            m1i = Male_lower_strength_relative_change_mean,
                            sd1i = Male_lower_strength_relative_change_sd,
                            n1i = Male_lower_strength_post_number,
                            m2i = Female_lower_strength_relative_change_mean,
                            sd2i = Female_lower_strength_relative_change_sd,
                            n2i = Female_lower_strength_post_number,
                            weighted = TRUE,
                            method = "REML",
                            measure = "SMD"))

(rel_lower_metareg_2 <- update(rel_lower_metareg_1,
                               ~ . - training:metareg_weekly_repetitions))
(rel_lower_metareg_3 <- update(rel_lower_metareg_2, ~ . - training))
(rel_lower_metareg_4 <- update(rel_lower_metareg_3, ~ . - metareg_intensity))
(rel_lower_metareg_5 <- update(rel_lower_metareg_4, ~ . - metareg_frequency))
(rel_lower_metareg_6 <- update(rel_lower_metareg_5, ~ . - duration))


# Plot
pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg.pdf",
          sep="\\"),
    width = 12, height = 8)
```

```
regplot(rel_lower_metareg_6, xlab="Weekly repetitions")

dev.off()

# Metareg Assumption Check
rel_lower_metareg_inf <- influence(rel_lower_metareg_4)

# Normality of (external) studentized residuals
pdf(paste(image_path,
          "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_Normality.pdf",
          sep="\\"),
    width = 12, height = 8)

qqnorm(rel_lower_metareg_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(rel_lower_metareg_inf$inf$rstudent)

dev.off()

shapiro.test(rel_lower_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_extres.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(rel_lower_metareg_inf$p/
# (rel_lower_metareg_inf$k-rel_lower_metareg_inf$p
3 * sqrt(rel_lower_metareg_inf$p/
           (rel_lower_metareg_inf$k-rel_lower_metareg_inf$p))

pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_dffits.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (rel_lower_metareg_inf$p +1))
qchisq(p = 0.5, df = (rel_lower_metareg_inf$p +1))

pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_Cook.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
```

```
pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_Covr.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_Tau2L1o.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Rel_Lower_Metareg_Chp2\\rel_lower_metareg_dfbetas.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_lower_metareg_inf, plotdfbs = 1:2, plotinf=FALSE, cex = 1.5)

dev.off()

#Identification of the outliers
df_lower$Study[c(12, 23, 35)]

# Check removing "Sood et al." & "Walts et al."
df_lower %>%
  subset(!(Study %in% df_lower$Study[c(12, 23, 35)])) %>%
  rma(data = .,
      mods = ~ metareg_weekly_repetitions,
      m1i = Male_lower_strength_relative_change_mean,
      sd1i = Male_lower_strength_relative_change_sd,
      n1i = Male_lower_strength_post_number,
      m2i = Female_lower_strength_relative_change_mean,
      sd2i = Female_lower_strength_relative_change_sd,
      n2i = Female_lower_strength_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Meta many metareg
Cov_matrix <- model.matrix(yi ~ metareg_weekly_repetitions,
                           data = df_lower)

Cov_matrix <- Cov_matrix[, -1] #removing intercept for meta_many
set.seed(23091994)

#Attensione intercetta
```

```
rel_lower_metareg_meta_many <- meta_many(y_i = df_lower$yi,
                                         si2 = df_lower$vi,
                                         X = Cov_matrix,
                                         alpha = Alpha, B = 2000)

# Betas
rel_lower_metareg_meta_many$Betas %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  arrange(Var_names) %>%
  xtable(digits = 4)

# Betas CI
rel_lower_metareg_meta_many$Beta_CIs %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
   pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)')%>%
  mutate(Width = UL - LL) %>%
  arrange(Var_names) %>%
  xtable(digits = 4)

# Betas p-val
rel_lower_metareg_meta_many$Beta_pvals %>%
  mutate(Var_names = row.names(rel_lower_metareg_meta_many$Beta_CIs)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(-Var_names,
               names_to = "Method",
               values_to = "p") %>%
  pivot_wider(names_from = Var_names, values_from = p) %>%
  xtable(digits = 4)


# Tau2s
rel_lower_metareg_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
rel_lower_metareg_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
```

```
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

################################################################################
## Analysis 2b: Absolute change lower body muscle strength
################################################################################

# Random effect meta-analysis
(abs_lower_rma <- rma(data = df_lower,
                      m1i = Male_lower_strength_absolute_change_mean,
                      sd1i = Male_lower_strength_absolute_change_sd,
                      n1i = Male_lower_strength_post_number,
                      m2i = Female_lower_strength_absolute_change_mean,
                      sd2i = Female_lower_strength_absolute_change_sd,
                      n2i = Female_lower_strength_post_number,
                      weighted = TRUE,
                      method = "REML",
                      measure = "SMD",
                      slab = paste(Study, Reference, sep = " ")))

## forest plot

# Decrease margins so the full space is used
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_rma.pdf", sep="\\"),
    width = 12, height = 8)

par(mar = c(4,4,1,2))

forest(abs_lower_rma,
       addfit = FALSE,
       showweights = TRUE,
       xlim=c(-15, 7),
       ylim = c(-2, 37.5),
       at = c(-3, 0, 3),
       ilab = round(cbind(df_lower$Male_lower_strength_absolute_change_mean,
                          df_lower$Male_lower_strength_absolute_change_sd,
                          df_lower$Male_lower_strength_post_number,
                          df_lower$Female_lower_strength_absolute_change_mean,
                          df_lower$Female_lower_strength_absolute_change_sd,
                          df_lower$Female_lower_strength_post_number),
                    digits = 2),
       ilab.xpos=c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
       rows = 34:0,
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(abs_lower_rma, row = -1.5, cex = 0.8, mlab = "")
text(7, -2.5, paste("p = 0.000002"), pos = 2, cex = 0.8)

abs_lower_fsn <- fsn(yi = abs_lower_rma$yi, vi = abs_lower_rma$vi,
                     type = "Rosenthal", alpha = 0.05)
text(-15, -2.5, pos = 4, cex = 0.8, paste("Fail-safe N =",
                                          formatC(abs_lower_fsn$fsnum)))
```

```
op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      35,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       36, c("Male", "Female"))
text(-15,                             36, "Study",  pos = 4)
text(7,                               36, "Hedges' g [95% CI]", pos = 2)
text(0,                               36,
     "Absolute change in lower body muscle strength")

text(-15, -1.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
                  .(formatC(abs_lower_rma$QE, digits=2, format="f")),
                  ", df = ",
                  .(abs_lower_rma$k - abs_lower_rma$p),
                  ", p = ",
                  .(formatC(abs_lower_rma$QEp, digits=2, format="f")),
                  "; ", I^2, " = ",
                  .(formatC(abs_lower_rma$I2, digits=1, format="f")),
                  "%)")))
dev.off()

# Assumption checks
abs_lower_inf <- influence(abs_lower_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_Normality.pdf", sep="\\"),
    width = 12, height = 8)

qqnorm(abs_lower_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(abs_lower_inf$inf$rstudent)

dev.off()

shapiro.test(abs_lower_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotinf = 1, cex = 1.5)
dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(abs_lower_inf$p/(abs_lower_inf$k-abs_lower_inf$p
3 * sqrt(abs_lower_inf$p/(abs_lower_inf$k-abs_lower_inf$p))

pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (abs_lower_inf$p +1))
qchisq(p = 0.5, df = (abs_lower_inf$p +1))
```

```
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_Cook.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_lower_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Significance removing "Holviala et al."
df_lower %>%
  subset(Study != "Holviala et al.") %>%
  rma(data = .,
      m1i = Male_lower_strength_absolute_change_mean,
      sd1i = Male_lower_strength_absolute_change_sd,
      n1i = Male_lower_strength_post_number,
      m2i = Female_lower_strength_absolute_change_mean,
      sd2i = Female_lower_strength_absolute_change_sd,
      n2i = Female_lower_strength_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Alternative methods

# Effect sizes calculations
escalc(measure = "SMD",
       data = df_lower,
```

```
       m1i = Male_lower_strength_absolute_change_mean,
       sd1i = Male_lower_strength_absolute_change_sd,
       n1i = Male_lower_strength_post_number,
       m2i = Female_lower_strength_absolute_change_mean,
       sd2i = Female_lower_strength_absolute_change_sd,
       n2i = Female_lower_strength_post_number, append = TRUE) -> df_lower

# Meta many
set.seed(23091994)
abs_lower_meta_many <- meta_many(y_i = df_lower$yi,
                                 si2 = df_lower$vi,
                                 X = NULL, alpha = Alpha, B = 2000)

# Betas
abs_lower_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
abs_lower_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  slice(-length(Estimator)) %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Betas p-val
abs_lower_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 4)


# Tau2s
abs_lower_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
abs_lower_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot

pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)
```

```
par(mar=c(5,4,1,2))
funnel(abs_lower_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_abs_lower_rma <- regtest(abs_lower_rma))

#variance
(Egger_abs_lower_rma_va <- regtest(abs_lower_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(abs_lower_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)

# se
lines(coef(Egger_abs_lower_rma$fit)[1] +
        coef(Egger_abs_lower_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_abs_lower_rma_va$fit)[1] +
        coef(Egger_abs_lower_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
(abs_lower_trim_fill <- trimfill(abs_lower_rma))

pdf(paste(image_path, "Abs_Lower_Chp2\\abs_lower_TrimFill.pdf", sep="\\"),
```

```
    width = 12, height = 8)

funnel(abs_lower_trim_fill,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Meta - regression models

# training type
df_lower %>% group_by(training) %>% summarise(n())


df_lower %>%
  ggplot(aes(x = training, y = yi)) +
  geom_boxplot(alpha = 0.80, outlier.shape = NA) +
  geom_point(aes(size = (1/vi)/sum(1/vi), fill = training),
             shape = 21, position = position_jitterdodge()) +
  xlab("Training type") +
  ylab("SMD") +
  labs(fill = "Training type", size = "CE weights")


ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_training.pdf",
             sep="\\"), dpi = 600)

# Study duration
df_lower %>%
  ggplot(aes(x = duration, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_duration.pdf",
             sep="\\"), dpi = 600)

# Weekly Volume
df_lower %>%
  ggplot(aes(x = metareg_weekly_repetitions, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_volume.pdf",
             sep="\\"), dpi = 600)

# Intensity
df_lower %>%
  ggplot(aes(x = metareg_intensity, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")
```

```
ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_intensity.pdf",
             sep="\\"), dpi = 600)

# Frequency
df_lower %>%
  ggplot(aes(x = metareg_frequency, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_frequency.pdf",
             sep="\\"), dpi = 600)

# N_exercise
df_lower %>%
  ggplot(aes(x = metareg_exercises, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_exercise.pdf",
             sep="\\"), dpi = 600)

# Sets (per exercise)
df_lower %>%
  ggplot(aes(x = metareg_sets, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_exp_sets.pdf",
             sep="\\"), dpi = 600)

# Metaregression Model
(abs_lower_metareg_1 <- rma(data = df_lower,
                            mods = ~ training * metareg_weekly_repetitions +
                              duration + metareg_intensity + metareg_frequency,
                            m1i = Male_lower_strength_absolute_change_mean,
                            sd1i = Male_lower_strength_absolute_change_sd,
                            n1i = Male_lower_strength_post_number,
                            m2i = Female_lower_strength_absolute_change_mean,
                            sd2i = Female_lower_strength_absolute_change_sd,
                            n2i = Female_lower_strength_post_number,
                            weighted = TRUE,
                            method = "REML",
                            measure = "SMD"))

(abs_lower_metareg_2 <- update(abs_lower_metareg_1,
                               ~ . - training:metareg_weekly_repetitions))
(abs_lower_metareg_3 <- update(abs_lower_metareg_2, ~ . - duration))
(abs_lower_metareg_4 <- update(abs_lower_metareg_3, ~ . - training))
(abs_lower_metareg_5 <- update(abs_lower_metareg_4, ~ . - metareg_frequency))
(abs_lower_metareg_6 <- update(abs_lower_metareg_5, ~ . - metareg_intensity))

# Plot
pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg.pdf",
          sep="\\"),
```

```
    width = 12, height = 8)

regplot(abs_lower_metareg_6, xlab="Weekly repetitions")

dev.off()

# Metareg Assumption Check
abs_lower_metareg_inf <- influence(abs_lower_metareg_6)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_Normality.pdf",
          sep="\\"),
    width = 12, height = 8)

qqnorm(abs_lower_metareg_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(abs_lower_metareg_inf$inf$rstudent)

dev.off()

shapiro.test(abs_lower_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_extres.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(abs_lower_metareg_inf$p/
# (abs_lower_metareg_inf$k-abs_lower_metareg_inf$p
3 * sqrt(abs_lower_metareg_inf$p/
           (abs_lower_metareg_inf$k-abs_lower_metareg_inf$p))

pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_dffits.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (abs_lower_metareg_inf$p +1))
qchisq(p = 0.5, df = (abs_lower_metareg_inf$p +1))

pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_Cook.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotinf = 3, cex = 1.5)

dev.off()
```

```
# Covariance ratio
pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_Covr.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_Tau2L1o.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Abs_Lower_Metareg_Chp2\\abs_lower_metareg_dfbetas.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_lower_metareg_inf, plotdfbs = 1:2, plotinf=FALSE, cex = 1.5)

dev.off()

#Identification of the outliers
df_lower$Study[c(12, 23, 35)]

# Check removing "Sood et al." & "Walts et al."
df_lower %>%
  subset(!(Study %in% df_lower$Study[c(23)])) %>%
  rma(data = .,
      mods = ~ metareg_weekly_repetitions,
      m1i = Male_lower_strength_relative_change_mean,
      sd1i = Male_lower_strength_relative_change_sd,
      n1i = Male_lower_strength_post_number,
      m2i = Female_lower_strength_relative_change_mean,
      sd2i = Female_lower_strength_relative_change_sd,
      n2i = Female_lower_strength_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Meta many metareg
Cov_matrix <- model.matrix(yi ~ metareg_weekly_repetitions,
                           data = df_lower)

Cov_matrix <- Cov_matrix[, -1] #removing intercept for meta_many
set.seed(23091994)
```

```
#Attensione intercetta
abs_lower_metareg_meta_many <- meta_many(y_i = df_lower$yi,
                                         si2 = df_lower$vi,
                                         X = Cov_matrix,
                                         alpha = Alpha, B = 2000)

# Betas
abs_lower_metareg_meta_many$Betas %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  arrange(Var_names) %>%
  xtable(digits = 5)

# Betas CI
abs_lower_metareg_meta_many$Beta_CIs %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>%
  arrange(Var_names) %>%
  xtable(digits = 6)

# Betas p-val
abs_lower_metareg_meta_many$Beta_pvals %>%
  mutate(Var_names = row.names(abs_lower_metareg_meta_many$Beta_CIs)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(-Var_names,
               names_to = "Method",
               values_to = "p") %>%
  pivot_wider(names_from = Var_names, values_from = p) %>%
  xtable(digits = 10)


# Tau2s
abs_lower_metareg_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 7)

# Tau2s CI
abs_lower_metareg_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
```

```
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

################################################################################
## Analysis 3a: Relative change muscle size
################################################################################

# Filter
df %>%
  filter(Age == "Older", Muscle_size == 1) -> df_muscle

# Studies included
df_muscle %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_muscle %>%
  summarise(muscle_males = sum(Male_muscle_post_number),
            muscle_females = sum(Female_muscle_post_number))


# Random effect meta-analysis
(rel_muscle_rma <- rma(data = df_muscle,
                       m1i = Male_muscle_relative_change_mean,
                       sd1i = Male_muscle_relative_change_sd,
                       n1i = Male_muscle_post_number,
                       m2i = Female_muscle_relative_change_mean,
                       sd2i = Female_muscle_relative_change_sd,
                       n2i = Female_muscle_post_number,
                       weighted = TRUE,
                       method = "REML",
                       measure = "SMD",
                       slab = paste(Study, Reference, sep = " ")))

# Decrease margins so the full space is used
pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_rma.pdf", sep="\\"),
    width = 12, height = 8)
par(mar = c(4,4,1,2))

# forest plot
forest(rel_muscle_rma,
       addfit = FALSE,
       showweights = TRUE,
       xlim=c(-14, 6),
       ylim = c(-2, 32.5),
       at = c(-3, 0, 3),
       ilab = cbind(df_muscle$Male_muscle_relative_change_mean,
                    df_muscle$Male_muscle_relative_change_sd,
                    df_muscle$Male_muscle_post_number,
                    df_muscle$Female_muscle_relative_change_mean,
                    df_muscle$Female_muscle_relative_change_sd,
                    df_muscle$Female_muscle_post_number),
       ilab.xpos=c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
```

```
       rows = c(29:0),
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(rel_muscle_rma, row = -1.5, cex = 0.8, mlab = "")
text(6, -2.5, paste("p =", formatC(rel_muscle_rma$pval, digits = 4)),
     pos = 2, cex = 0.8)

rel_muscle_fsn <- fsn(yi = rel_muscle_rma$yi, vi = rel_muscle_rma$vi,
                      type = "Rosenthal", alpha = 0.05)
text(-14, -2.5, pos = 4, cex = 0.8,
     paste("Fail-safe N =", formatC(rel_muscle_fsn$fsnum)))

op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      30,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       31, c("Male", "Female"))
text(-14,                             31, "Study",  pos = 4)
text(6,                               31, "Hedges' g [95% CI]", pos = 2)
text(0,                               31,
     "Relative change in muscle size")

text(-14, -1.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
                  .(formatC(rel_muscle_rma$QE, digits=2, format="f")),
                  ", df = ",
                  .(rel_muscle_rma$k - rel_muscle_rma$p),
                  ", p = ",
                  .(formatC(rel_muscle_rma$QEp, digits=2, format="f")),
                  "; ", I^2, " = ",
                  .(formatC(rel_muscle_rma$I2, digits=1, format="f")),
                  "%)")))
dev.off()

# Assumption checks
rel_muscle_inf <- influence(rel_muscle_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_Normality.pdf", sep="\\"),
    width = 12, height = 8)

qqnorm(rel_muscle_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(rel_muscle_inf$inf$rstudent)

dev.off()

shapiro.test(rel_muscle_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotinf = 1, cex = 1.5)

dev.off()
```

```
# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(rel_muscle_inf$p/(rel_muscle_inf$k-rel_muscle_inf$p
3 * sqrt(rel_muscle_inf$p/(rel_muscle_inf$k-rel_muscle_inf$p))

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (rel_muscle_inf$p +1))
qchisq(p = 0.5, df = (rel_muscle_inf$p +1))

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_Cook.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotinf = 5, cex = 1.5)

dev.off()

rel_muscle_inf$inf$slab
# Dfbetas

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Alternative methods

# Effect sizes calculations
escalc(measure = "SMD",
       data = df_muscle,
```

```
       m1i = Male_muscle_relative_change_mean,
       sd1i = Male_muscle_relative_change_sd,
       n1i = Male_muscle_post_number,
       m2i = Female_muscle_relative_change_mean,
       sd2i = Female_muscle_relative_change_sd,
       n2i = Female_muscle_post_number, append = TRUE) -> df_muscle

# Meta many
set.seed(23091994)
rel_muscle_meta_many <- meta_many(y_i = df_muscle$yi,
                                  si2 = df_muscle$vi,
                                  X = NULL, alpha = Alpha, B = 2000)

# Betas
rel_muscle_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
rel_muscle_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Betas p-val
rel_muscle_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 4)


# Tau2s
rel_muscle_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
rel_muscle_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
```

```
funnel(rel_muscle_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_rel_muscle_rma <- regtest(rel_muscle_rma))

#variance
(Egger_rel_muscle_rma_va <- regtest(rel_muscle_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(rel_muscle_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)

# se
lines(coef(Egger_rel_muscle_rma$fit)[1] +
        coef(Egger_rel_muscle_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_rel_muscle_rma_va$fit)[1] +
        coef(Egger_rel_muscle_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
(rel_muscle_trim_fill <- trimfill(rel_muscle_rma))

pdf(paste(image_path, "Rel_Muscle_Chp2\\rel_muscle_TrimFill.pdf", sep="\\"),
    width = 12, height = 8)
```

```
funnel(rel_muscle_trim_fill,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Meta - regression models

# training type
df_muscle %>% group_by(training) %>% summarise(n())


df_muscle %>%
  ggplot(aes(x = training, y = yi)) +
  geom_boxplot(alpha = 0.80, outlier.shape = NA) +
  geom_point(aes(size = (1/vi)/sum(1/vi), fill = training),
             shape = 21, position = position_jitterdodge()) +
  xlab("Training type") +
  ylab("SMD") +
  labs(fill = "Training type", size = "CE weights")


ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_training.pdf",
             sep="\\"), dpi = 600)

# Study duration
df_muscle %>%
  ggplot(aes(x = duration, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_duration.pdf",
             sep="\\"), dpi = 600)

# Weekly Volume
df_muscle %>%
  ggplot(aes(x = metareg_weekly_repetitions, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_volume.pdf",
             sep="\\"), dpi = 600)

# Intensity
df_muscle %>%
  ggplot(aes(x = metareg_intensity, y = yi, col=training)) +
```

```
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_intensity.pdf",
             sep="\\"), dpi = 600)

# Frequency
df_muscle %>%
  ggplot(aes(x = metareg_frequency, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_frequency.pdf",
             sep="\\"), dpi = 600)

# N_exercise
df_muscle %>%
  ggplot(aes(x = metareg_exercises, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_exercise.pdf",
             sep="\\"), dpi = 600)

# Sets (per exercise)
df_muscle %>%
  ggplot(aes(x = metareg_sets, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Rel_Muscle_Metareg_Chp2\\rel_muscle_exp_sets.pdf",
             sep="\\"), dpi = 600)

# Metaregression Model
(rel_muscle_metareg_1 <- rma(data = df_muscle,
                             mods = ~ training * metareg_weekly_repetitions +
                               duration + metareg_frequency + metareg_intensity,
                             m1i = Male_muscle_relative_change_mean,
                             sd1i = Male_muscle_relative_change_sd,
                             n1i = Male_muscle_post_number,
                             m2i = Female_muscle_relative_change_mean,
                             sd2i = Female_muscle_relative_change_sd,
                             n2i = Female_muscle_post_number,
                             weighted = TRUE,
                             method = "REML",
                             measure = "SMD"))
(rel_muscle_metareg_2 <- update(rel_muscle_metareg_1, ~ . - metareg_intensity))
(rel_muscle_metareg_3 <- update(rel_muscle_metareg_2, ~ . - metareg_frequency))
(rel_muscle_metareg_4 <- update(rel_muscle_metareg_3,
                                ~ . - training:metareg_weekly_repetitions))
(rel_muscle_metareg_5 <- update(rel_muscle_metareg_4, ~ . - training))
```

```
# Plot duration
pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_duration.pdf",
          sep="\\"),
    width = 12, height = 8)

regplot(rel_muscle_metareg_5, mod = "duration",  xlab="Weekly repetitions")

dev.off()

# Plot duration
pdf(paste(image_path, "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_reps.pdf",
          sep="\\"),
    width = 12, height = 8)

regplot(rel_muscle_metareg_5, mod= "metareg_weekly_repetitions",
        xlab="Weekly repetitions")

dev.off()


# Metareg Assumption Check
rel_muscle_metareg_inf <- influence(rel_muscle_metareg_4)

# Normality of (external) studentized residuals
pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_Normality.pdf",
          sep="\\"),
    width = 12, height = 8)

qqnorm(rel_muscle_metareg_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(rel_muscle_metareg_inf$inf$rstudent)

dev.off()

shapiro.test(rel_muscle_metareg_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_extres.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(rel_muscle_metareg_inf$p/
# (rel_muscle_metareg_inf$k-rel_muscle_metareg_inf$p
3 * sqrt(rel_muscle_metareg_inf$p/
           (rel_muscle_metareg_inf$k-rel_muscle_metareg_inf$p))

pdf(paste(image_path,
```

```
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_dffits.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (rel_muscle_metareg_inf$p +1))
qchisq(p = 0.5, df = (rel_muscle_metareg_inf$p +1))

pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_Cook.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_Covr.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_Tau2L1o.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path,
          "Rel_Muscle_Metareg_Chp2\\rel_muscle_metareg_dfbetas.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(rel_muscle_metareg_inf, plotdfbs = 1:3, plotinf=FALSE, cex = 1.5)

dev.off()

#Identification of the outliers
df_muscle$Study[c(22, 23, 24)]
```

```
# Check removing "Sood et al." & "Walts et al."
df_muscle %>%
  subset(!(Study %in% df_muscle$Study[c(22, 23, 24)])) %>%
  rma(data = .,
      mods = ~ metareg_weekly_repetitions + duration,
      m1i = Male_muscle_relative_change_mean,
      sd1i = Male_muscle_relative_change_sd,
      n1i = Male_muscle_post_number,
      m2i = Female_muscle_relative_change_mean,
      sd2i = Female_muscle_relative_change_sd,
      n2i = Female_muscle_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

df_muscle %>%
  subset(!(Study %in% df_muscle$Study[c(22, 23, 24)])) %>%
  rma(data = .,
      mods = ~ metareg_weekly_repetitions,
      m1i = Male_muscle_relative_change_mean,
      sd1i = Male_muscle_relative_change_sd,
      n1i = Male_muscle_post_number,
      m2i = Female_muscle_relative_change_mean,
      sd2i = Female_muscle_relative_change_sd,
      n2i = Female_muscle_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))


# Meta many metareg
Cov_matrix <- model.matrix(yi ~ metareg_weekly_repetitions + duration,
                           data = df_muscle)

Cov_matrix <- Cov_matrix[, -1] #removing intercept for meta_many
set.seed(23091994)

#Attensione intercetta
rel_muscle_metareg_meta_many <- meta_many(y_i = df_muscle$yi,
                                          si2 = df_muscle$vi,
                                          X = Cov_matrix,
                                          alpha = Alpha, B = 2000)

# Betas

rel_muscle_metareg_meta_many$Betas %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
```

```
    names_pattern = '(.*)_(\\w+)') %>%
  arrange(match(Var_names, c("intrcpt",
                             "duration",
                             "metareg_weekly_repetitions"))) %>%
  xtable(digits = 4)

# Betas CI
rel_muscle_metareg_meta_many$Beta_CIs %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>%
  arrange(match(Var_names, c("intrcpt",
                             "duration",
                             "metareg_weekly_repetitions")))%>%
  xtable(digits = 3)

# Betas p-val
rel_muscle_metareg_meta_many$Beta_pvals %>%
  mutate(Var_names = row.names(rel_muscle_metareg_meta_many$Beta_CIs)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(-Var_names,
               names_to = "Method",
               values_to = "p") %>%
  pivot_wider(names_from = Var_names, values_from = p) %>%
  xtable(digits = 4)


# Tau2s
rel_muscle_metareg_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 32)

# Tau2s CI
rel_muscle_metareg_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

################################################################################
## Analysis 3b: Absolute change muscle size
################################################################################

# Filter
df_muscle  %>%
  filter(!is.na(Male_muscle_absolute_change_mean)) -> df_muscle_abs
```

```
# Studies included
df_muscle_abs %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_muscle_abs %>%
  summarise(muscle_males = sum(Male_muscle_post_number),
            muscle_females = sum(Female_muscle_post_number))

# Random effect meta-analysis
(abs_muscle_rma <- rma(data = df_muscle_abs,
                       m1i = Male_muscle_absolute_change_mean,
                       sd1i = Male_muscle_absolute_change_sd,
                       n1i = Male_muscle_post_number,
                       m2i = Female_muscle_absolute_change_mean,
                       sd2i = Female_muscle_absolute_change_sd,
                       n2i = Female_muscle_post_number,
                       weighted = TRUE,
                       method = "REML",
                       measure = "SMD",
                       slab = paste(Study, Reference, sep = " ")))

## forest plot

# Decrease margins so the full space is used
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_rma.pdf", sep="\\"),
    width = 12, height = 8)

par(mar = c(4,4,1,2))

forest(abs_muscle_rma,
       addfit = FALSE,
       showweights = TRUE,
       xlim=c(-14, 6.5),
       ylim = c(-2, 30.5),
       at = c(-3, 0, 3),
       ilab = cbind(df_muscle_abs$Male_muscle_absolute_change_mean,
                    df_muscle_abs$Male_muscle_absolute_change_sd,
                    df_muscle_abs$Male_muscle_post_number,
                    df_muscle_abs$Female_muscle_absolute_change_mean,
                    df_muscle_abs$Female_muscle_absolute_change_sd,
                    df_muscle_abs$Female_muscle_post_number),
       ilab.xpos=c(-10, -9, -8, -6, -5, -4), cex = 0.8,
       mlab = "",
       rows = c(27:0),
       psize = 1,
       xlab = "Favours females                    Favours males")

addpoly(abs_muscle_rma, row = -1.5, cex = 0.8, mlab = "")
text(6.5, -2.5, paste("p = 0.00007"), pos = 2, cex = 0.8)

abs_muscle_fsn <- fsn(yi = abs_muscle_rma$yi,
                      vi = abs_muscle_rma$vi,
```

```
                      type = "Rosenthal", alpha = 0.05)
text(-14, -2.5, pos = 4, cex = 0.8,
     paste("Fail-safe N =", formatC(abs_muscle_fsn$fsnum)))

op <- par(cex = 0.8, font = 2)
text(c(-10, -9, -8, -6, -5, -4),      28,
     c("Mean", "SD", "Sample", "Mean", "SD", "Sample"))
text(c(-9, -5),                       29, c("Male", "Female"))
text(-14,                             29, "Study",  pos = 4)
text(6.5,                             29, "Hedges' g [95% CI]", pos = 2)
text(0,                               29, "Absolute change in muscle size")

text(-14, -1.5, pos = 4, cex = 1,
     bquote(paste("Random-effects model (Q = ",
     .(formatC(abs_muscle_rma$QE, digits=2, format="f")),
     ", df = ",
     .(abs_muscle_rma$k - abs_muscle_rma$p),
     ", p = ",
     .(formatC(abs_muscle_rma$QEp, digits=2, format="f")),
     "; ", I^2, " = ",
     .(formatC(abs_muscle_rma$I2, digits=1, format="f")),
     "%)")))

dev.off()

# Assumption checks
abs_muscle_inf <- influence(abs_muscle_rma)

# Normality of (external) studentized residuals
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_Normality.pdf", sep="\\"),
    width = 12, height = 8)

qqnorm(abs_muscle_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(abs_muscle_inf$inf$rstudent)

dev.off()

shapiro.test(abs_muscle_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_extres.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotinf = 1, cex = 1.5)
dev.off()

# L1o standardized predicted average (dffits),
# lines = 3 * sqrt(abs_muscle_inf$p/(abs_muscle_inf$k-abs_muscle_inf$p
3 * sqrt(abs_muscle_inf$p/(abs_muscle_inf$k-abs_muscle_inf$p))

pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_dffits.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotinf = 2, cex = 1.5)
```

```
dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (abs_muscle_inf$p +1))
qchisq(p = 0.5, df = (abs_muscle_inf$p +1))

pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_Cook.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_Covr.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_Tau2L1o.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_dfbetas.pdf", sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_inf, plotdfbs = 1, plotinf=FALSE, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Significance removing "Holviala et al."
df_muscle_abs %>%
  subset(Study != "Holviala et al.") %>%
  rma(data = .,
      m1i = Male_muscle_absolute_change_mean,
      sd1i = Male_muscle_absolute_change_sd,
      n1i = Male_muscle_post_number,
      m2i = Female_muscle_absolute_change_mean,
      sd2i = Female_muscle_absolute_change_sd,
      n2i = Female_muscle_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " "))

# Alternative methods
```

```
# Effect sizes calculations
escalc(measure = "SMD",
       data = df_muscle_abs,
       m1i = Male_muscle_absolute_change_mean,
       sd1i = Male_muscle_absolute_change_sd,
       n1i = Male_muscle_post_number,
       m2i = Female_muscle_absolute_change_mean,
       sd2i = Female_muscle_absolute_change_sd,
       n2i = Female_muscle_post_number, append = TRUE) -> df_muscle_abs

# Meta many
set.seed(23091994)
abs_muscle_meta_many <- meta_many(y_i = df_muscle_abs$yi,
                                  si2 = df_muscle_abs$vi,
                                  X = NULL, alpha = Alpha, B = 2000)

# Betas
abs_muscle_meta_many$Betas %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Betas CI
abs_muscle_meta_many$Beta_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Betas p-val
abs_muscle_meta_many$Beta_pvals %>%
  pivot_longer(everything(),
               names_to = "Method",
               values_to = "p") %>% xtable(digits = 8)


# Tau2s
abs_muscle_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
abs_muscle_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)

# Publication bias
# Contour enhanced funnel plot
```

```
pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_cefunnel.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(abs_muscle_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Egger's regression test for asymmetry
# se
(Egger_abs_muscle_rma <- regtest(abs_muscle_rma))

#variance
(Egger_abs_muscle_rma_va <- regtest(abs_muscle_rma, predictor="vi"))

# Funnel plot with superimposed Egger's regression line

pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_cefunnelreg.pdf", sep="\\"),
    width = 12, height = 8)

par(mar=c(5,4,1,2))
funnel(abs_muscle_rma,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = FALSE)

se <- seq(0,0.6,length=100)

# se
lines(coef(Egger_abs_muscle_rma$fit)[1] +
        coef(Egger_abs_muscle_rma$fit)[2]*se, se)

# Sampling variance
lines(coef(Egger_abs_muscle_rma_va$fit)[1] +
        coef(Egger_abs_muscle_rma_va$fit)[2]*se^2, se, lwd=2, lty="dotted")

legend("topright", inset=.02, lty=c("solid","dotted"),
       lwd=2, cex=0.9, bg="white",
       legend=c("Standard Errors as Predictor",
                "Sampling Variances as Predictor"))

dev.off()

# Trim and fill (and contour enhanced Funel plot )
```

```
(abs_muscle_trim_fill <- trimfill(abs_muscle_rma))

pdf(paste(image_path, "Abs_Muscle_Chp2\\abs_muscle_TrimFill.pdf", sep="\\"),
    width = 12, height = 8)

funnel(abs_muscle_trim_fill,
       level=c(90, 95, 99),
       shade=c("white", "gray", "darkgray"),
       refline = 0,
       addtau2 = TRUE,
       xlab = "Hedges' g",
       cex = 1.5,
       legend = TRUE)

dev.off()

# Meta - regression models

# training type
df_muscle_abs %>% group_by(training) %>% summarise(n())


df_muscle_abs %>%
  ggplot(aes(x = training, y = yi)) +
  geom_boxplot(alpha = 0.80, outlier.shape = NA) +
  geom_point(aes(size = (1/vi)/sum(1/vi), fill = training),
             shape = 21, position = position_jitterdodge()) +
  xlab("Training type") +
  ylab("SMD") +
  labs(fill = "Training type", size = "CE weights")


ggsave(paste(image_path,
             "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_training.pdf",
             sep="\\"), dpi = 600)

# Study duration
df_muscle_abs %>%
  ggplot(aes(x = duration, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
             "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_duration.pdf",
             sep="\\"), dpi = 600)

# Weekly Volume
df_muscle_abs %>%
  ggplot(aes(x = metareg_weekly_repetitions, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")


ggsave(paste(image_path,
             "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_volume.pdf",
```

```
            sep="\\"), dpi = 600)

# Intensity
df_muscle_abs %>%
  ggplot(aes(x = metareg_intensity, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
            "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_intensity.pdf",
            sep="\\"), dpi = 600)

# Frequency
df_muscle_abs %>%
  ggplot(aes(x = metareg_frequency, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
            "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_frequency.pdf",
            sep="\\"), dpi = 600)

# N_exercise
df_muscle_abs %>%
  ggplot(aes(x = metareg_exercises, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path,
            "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_exercise.pdf",
            sep="\\"), dpi = 600)

# Sets (per exercise)
df_muscle_abs %>%
  ggplot(aes(x = metareg_sets, y = yi, col=training)) +
  geom_point(aes(size = (1/vi)/sum(1/vi))) +
  labs(col = "Training type", size = "CE weights")

ggsave(paste(image_path, "Abs_Muscle_Metareg_Chp2\\abs_muscle_exp_sets.pdf",
            sep="\\"), dpi = 600)

# Metaregression Model
(abs_muscle_metareg_1 <- rma(data = df_muscle_abs,
                             mods = ~ training * metareg_weekly_repetitions +
                               duration + metareg_frequency + metareg_intensity,
                             m1i = Male_muscle_absolute_change_mean,
                             sd1i = Male_muscle_absolute_change_sd,
                             n1i = Male_muscle_post_number,
                             m2i = Female_muscle_absolute_change_mean,
                             sd2i = Female_muscle_absolute_change_sd,
                             n2i = Female_muscle_post_number,
                             weighted = TRUE,
                             method = "REML",
                             measure = "SMD"))
```

```
(abs_muscle_metareg_2 <- update(abs_muscle_metareg_1, ~ . - metareg_intensity))
(abs_muscle_metareg_3 <- update(abs_muscle_metareg_2, ~ . - metareg_frequency))

#Presentation of the results

cbind(abs_muscle_metareg_3$beta,
      abs_muscle_metareg_3$se,
      abs_muscle_metareg_3$zval,
      abs_muscle_metareg_3$pval,
      abs_muscle_metareg_3$ci.lb,
      abs_muscle_metareg_3$ci.ub) %>%
  t() %>%
  xtable(digits = 4)

# Plot
#pdf(paste(image_path, "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg.pdf",
#           sep="\\"),
#    width = 12, height = 8)

#regplot(abs_muscle_metareg_4, xlab="Weekly repetitions")

#dev.off()

# Metareg Assumption Check
abs_muscle_metareg_inf <- influence(abs_muscle_metareg_3)

# Normality of (external) studentized residuals
pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_Normality.pdf",
          sep="\\"),
    width = 12, height = 8)

qqnorm(abs_muscle_metareg_inf$inf$rstudent, cex = 1.5, pch = 21, bg="grey")
qqline(abs_muscle_metareg_inf$inf$rstudent)

dev.off()

shapiro.test(abs_muscle_metareg_inf$inf$rstudent)

# "Size" of (external) studentized residuals,  lines = 1.96
pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_extres.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotinf = 1, cex = 1.5)

dev.off()

# L1o standardized predicted average (dffits), lines = 3 * sqrt(abs_muscle_metareg_inf$p/(abs_muscl
3 * sqrt(abs_muscle_metareg_inf$p/
           (abs_muscle_metareg_inf$k-abs_muscle_metareg_inf$p))

pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_dffits.pdf",
```

```
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotinf = 2, cex = 1.5)

dev.off()

# Cook distances, lines = qchisq(p = 0.5, df = (abs_muscle_metareg_inf$p +1))
qchisq(p = 0.5, df = (abs_muscle_metareg_inf$p +1))

pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_Cook.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotinf = 3, cex = 1.5)

dev.off()

# Covariance ratio
pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_Covr.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotinf = 4, cex = 1.5)
abline(h = 1, lty = "dashed")

dev.off()

# Tau2 l1o estimation, line = tau estimated from the model with all the studies
pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_Tau2L1o.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotinf = 5, cex = 1.5)

dev.off()

# Dfbetas

pdf(paste(image_path,
          "Abs_Muscle_Metareg_Chp2\\abs_muscle_metareg_dfbetas.pdf",
          sep="\\"),
    width = 12, height = 8)

plot(abs_muscle_metareg_inf, plotdfbs = 1:5, plotinf=FALSE, cex = 1.5)

dev.off()

#Identification of the outliers
df_muscle_abs$Study[c(9, 20, 22)]

# Check removing the influential studies
```

```
(df_muscle_abs %>%
  subset(!(Study %in% df_muscle_abs$Study[c(9, 20, 22)])) %>%
  rma(data = .,
      mods = ~ training * metareg_weekly_repetitions +
        duration,
      m1i = Male_muscle_absolute_change_mean,
      sd1i = Male_muscle_absolute_change_sd,
      n1i = Male_muscle_post_number,
      m2i = Female_muscle_absolute_change_mean,
      sd2i = Female_muscle_absolute_change_sd,
      n2i = Female_muscle_post_number,
      weighted = TRUE,
      method = "REML",
      measure = "SMD",
      slab = paste(Study, Reference, sep = " ")) -> abs_muscle_metareg_3_noinf)

cbind(abs_muscle_metareg_3_noinf$beta,
      abs_muscle_metareg_3_noinf$se,
      abs_muscle_metareg_3_noinf$zval,
      abs_muscle_metareg_3_noinf$pval,
      abs_muscle_metareg_3_noinf$ci.lb,
      abs_muscle_metareg_3_noinf$ci.ub) %>%
  xtable(digits = 4)

# Meta many metareg
Cov_matrix <- model.matrix(yi ~ training * metareg_weekly_repetitions +
                             duration,
                           data = df_muscle_abs)

Cov_matrix <- Cov_matrix[, -c(1, ncol(Cov_matrix))] #removing intercept
set.seed(23091994)

#Attensione intercetta
abs_muscle_metareg_meta_many <- meta_many(y_i = df_muscle_abs$yi,
                                          si2 = df_muscle_abs$vi,
                                          X = Cov_matrix,
                                          alpha = Alpha, B = 2000)

# Betas
abs_muscle_metareg_meta_many$Betas %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  arrange(match(Var_names, c("intrcpt",
                             "duration",
                             "trainingLower",
                             "trainingUpper",
                             "metareg_weekly_repetitions",
                             "trainingLower:metareg_weekly_repetitions"))) %>%
  xtable(digits = 4)
```

```
# Betas CI
abs_muscle_metareg_meta_many$Beta_CIs %>%
  mutate(Var_names = row.names(.)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(
    -Var_names,
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>%
  arrange(match(Var_names, c("intrcpt",
                             "duration",
                             "trainingLower",
                             "trainingUpper",
                             "metareg_weekly_repetitions",
                             "trainingLower:metareg_weekly_repetitions"))) %>%
  xtable(digits = 4)

# Betas p-val
abs_muscle_metareg_meta_many$Beta_pvals %>%
  mutate(Var_names = row.names(abs_muscle_metareg_meta_many$Beta_CIs)) %>%
  select(c(ncol(.),
           1:(ncol(.)-1))) %>%
  pivot_longer(-Var_names,
               names_to = "Method",
               values_to = "p") %>%
  pivot_wider(names_from = Var_names, values_from = p) %>%
select("Method", "intrcpt", "trainingLower", "trainingUpper",
       "metareg_weekly_repetitions", "duration",
       "trainingLower:metareg_weekly_repetitions") %>%
  xtable(digits = 4)


# Tau2s
abs_muscle_metareg_meta_many$Tau2s %>%
  pivot_longer(everything(), names_to = c("Estimator", '.value'),
               names_pattern = '(.*)_(\\w+)') %>% xtable(digits = 4)

# Tau2s CI
abs_muscle_metareg_meta_many$Tau2_CIs %>%
  pivot_longer(
    everything(),
    cols_vary = "slowest",
    names_to = c("Estimator", ".value"),
    names_pattern = '(.*)_(\\w+)') %>%
  mutate(Width = UL - LL) %>% xtable(digits = 4)
```

Appendix

Appendix 3

```
##############################################################################
# List of function and documentation
##############################################################################

#' L_beta_zeta_sigma_i_BG_mod
#' This function calculates the log-likelihood value following equation 1 in
#' Bellio, Guolo, 2016.

#' Arguments:
#' @param  beta:    Numeric value representing the parameter beta.
#' @param zeta:     Numeric value representing the parameter zeta.
#' @param sigma_i:  Numeric value representing the parameter sigma_i.
#' @param y_i:      Numeric value representing the observed data y_i.
#' @param s2_i:     Numeric value representing the precision measure s2_i.
#' @param n1_i:     Numeric value representing the sample size of group 1.
#' @param n2_i:     Numeric value representing the sample size of group 2.
#' @param hat_beta: Numeric value representing the maximum likelihood
#'                  estimated of beta.

#' Returns:
#' The likelihood value L.

# Function
L_beta_zeta_sigma_i_BG_mod <- function(beta, zeta, sigma_i, y_i, s2_i,
                                       n1_i, n2_i, hat_beta){

  #calculation fi_1
  f_i = n1_i + n2_i -2

  #Reparametrizing
  tau2 = zeta^2 + (hat_beta - beta)^2

  #Likelihood calculation

  l =   -0.5 * log(sigma_i^2 + tau2) - 0.5 * (y_i - beta)^2 /
    (sigma_i^2 + tau2) -
    0.5 * f_i *log (sigma_i^2) - 0.5 * f_i * s2_i / (sigma_i^2)

  L = exp(l) / exp(n1_i + n2_i)
```

```
  return(L)

}



#' g_i_beta_zeta_mod
#' This function calculates the element g_i(\beta, \zeta) of the equation 6 in
#' Bellio, Guolo, 2016.

#' Arguments:
#' @param  beta:      Numeric value representing the parameter beta.
#' @param zeta:       Numeric value representing the parameter zeta.
#' @param y_i:        Numeric value representing the observed data y_i.
#' @param s2_i:       Numeric value representing the precision measure s2_i.
#' @param n1_i:       Numeric value representing the sample size of group 1.
#' @param n2_i:       Numeric value representing the sample size of group 2.
#' @param hat_beta:   Numeric value representing the maximum likelihood
#'                    estimated of beta.

#' Returns:
#' A numeric value representing the result of the calculation.
#' - int_values: A numeric vector of the integrated values.

# Function
g_i_beta_zeta_mod <- function(beta, zeta, y_i, s2_i, n1_i, n2_i, hat_beta){

  #F objective for integration
  f_obv <- function(beta, zeta, sigma_i, y_i, s2_i,
                    n1_i, n2_i, hat_beta){
    L_beta_zeta_sigma_i_BG_mod(beta, zeta,
                               sigma_i, y_i, s2_i, n1_i, n2_i,
                               hat_beta) * 1/sigma_i
  }

  #integrate
  res <- integrate(f = f_obv,
                   lower = 0, upper = Inf,
                   beta = beta, zeta = zeta,
                   y_i = y_i, s2_i = s2_i,
                   n1_i = n1_i, n2_i = n2_i,
                   hat_beta = hat_beta)$value

  #Return
  return(res)
}

#' l_int_beta_mod
#' This function calculate the log likelihood of the equation 6 in
#' Bellio, Guolo, 2016.

#' Arguments:
#' @param  beta:     Numeric value representing the parameter beta.
#' @param y_i:       Numeric value representing the observed data y_i.
#' @param s2_i:      Numeric value representing the precision measure s2_i.
```

```
#' @param n1_i:     Numeric value representing the sample size of group 1.
#' @param n2_i:     Numeric value representing the sample size of group 2.
#' @param hat_beta: Numeric value representing the maximum likelihood
#'                  estimated of beta.

#' Returns:
#' A numeric value representing the result of the integration calculation.

# Function
l_int_beta_mod <- function(beta, y, s2, n1, n2, hat_beta){

  f_obv <-function(beta, zeta, y, s2, n1, n2, hat_beta){

    # matrix
    l = length(y)
    mat_tmp = cbind(rep(beta, l),
                    rep(zeta, l),
                    y, s2, n1, n2,
                    rep(hat_beta, l))

    #g calculation
    res <- apply(mat_tmp, 1, function(t)
      g_i_beta_zeta_mod(beta = t[1], zeta = t[2],
                    y_i = t[3], s2_i = t[4],
                    n1_i = t[5], n2_i = t[6],
                    hat_beta = t[7]))

    #prod
    #p_g <- prod(res) * 1 #\pi(\zeta) = 1
    p_g <- exp(sum(log(res)))  #exp log sum trick #as.bigz()


    #Return
    return(p_g)
  }

  res <-integrate(f = Vectorize(f_obv, vectorize.args = "zeta"),
                  lower = 0,
                  upper = Inf,
                  beta = beta,
                  y = y, s2 = s2,
                  n1 = n1, n2= n2,
                  hat_beta = hat_beta)$value

  #log lik calculation
  res <- log(res) + sum(n1 + n2)

  #Return
  return(res)
}


#' r_int_beta
#'
#' This function calculate the integrated log likelihood ratio statistics
```

```
#' proposed in of the equation 3 in Bellio, Guolo, 2016. It provides confidence
#' intervals, the p-value and a integrated log likelihood plot with a red line
#' corresponding to qnorm(1-alpha/2).
#'
#' @param beta_zero The hypothesized value of Beta for which the
#'                  statistical test is performed.
#' @param y         A numeric vector of observed effect sizes.
#' @param s2        A numeric vector of observed pecisions.
#' @param n1        A numeric vector of observed sample sizes in groups 1.
#' @param n2        A numeric vector of observed sample sizes in groups 2.
#' @param hat_beta  A numeric value representing the ML estimated value of Beta.
#' @param k_opt     A numeric value representing the multiplier for the range
#'                  optimization. Defaults to 1.5.
#' @param alpha     The desired level of confidence for the confidence
#'                  interval. Defaults to 0.05.
#' @param plot      A logical value indicating whether to plot the integrated
#'                  likelihood curve. Defaults to TRUE.
#' @param print     A logical value indicating whether to print the results.
#'                  Defaults to TRUE.
#'
#' @return          A list containing the following results:
#'                  - r_Int: The estimated r_{Int}(\beta) statistis
#'                  - p: The p-value of the hypothesis test.
#'                  - CI_ll: The lower limit of the confidence interval.
#'                  - CI_ul: The upper limit of the confidence interval.
#'@print            - all the values returned plus beta_bar,
#'                  which is the integrated ML estimate

#Function
r_int_beta_mod <- function(beta_zero, y, s2,
                       n1, n2, hat_beta, k_opt = 1.5, alpha = 0.05,
                       plot=TRUE, print =TRUE){

  #range
  if(hat_beta > 0){
    range_opt <- c(hat_beta - k_opt * hat_beta, hat_beta + k_opt * hat_beta)
  } else{
    range_opt <- c(hat_beta + k_opt * hat_beta, hat_beta - k_opt * hat_beta)
  }


  #Optimization
  res <- optimize(f = l_int_beta_mod,
                  maximum = TRUE,
                  interval = range_opt,
                  y = y, s2 = s2,
                  n1 = n1, n2 = n2,
                  hat_beta = hat_beta)

  #Param
  beta_bar <- res$maximum
  llik_int_bbar <- res$objective
  llik_int_b0   <- l_int_beta_mod(beta = beta_zero, y = y, s2 = s2,
                                  n1 = n1, n2 = n2,
                                  hat_beta = hat_beta)
```

```
#Stat
stat <- sign(beta_bar - beta_zero) * sqrt(2 * (llik_int_bbar - llik_int_b0))
pval <- 2*(1-pnorm(abs(stat)))

#Confidence Interval
zero_to_solve <- function(beta, beta_bar, y, s2, n1, n2, hat_beta, alpha){
  llik_1 <- l_int_beta_mod(beta, y = y, s2 = s2,
                        n1 = n1, n2 = n2,
                        hat_beta = hat_beta)

  llik_2 <- l_int_beta_mod(beta_bar, y = y, s2 = s2,
                        n1 = n1, n2 = n2,
                        hat_beta = hat_beta)

  val <-abs(llik_1 - llik_2 + qnorm(1-alpha/2))

  return(val)
}

ll_ci <- optimize(zero_to_solve,
                  lower = beta_bar - 5*k_opt,
                  upper = beta_bar,
                  beta_bar = beta_bar, y = y, s2 =s2,
                  n1 = n1, n2 =n2,
                  hat_beta = hat_beta, alpha=alpha)

ul_ci <- optimize(zero_to_solve,
                  lower = beta_bar,
                  upper = beta_bar + 5*k_opt,
                  beta_bar = beta_bar, y = y, s2 =s2,
                  n1 = n1, n2 =n2,
                  hat_beta = hat_beta, alpha=alpha)

#Plot
if(plot){
  x <- seq(beta_bar - k_opt, beta_bar + k_opt, by=0.1)
  y <- sapply(x, function(t) l_int_beta_mod(t, y = y, s2 = s2,
                                        n1 = n1, n2 = n2,
                                        hat_beta = hat_beta))

  par(mar = c(5, 5, 4, 6))

  plot(x, y, type="l",
       xlab = expression(beta),
       ylab = expression(paste("l(", beta, ")")),
       cex.lab = 1.5)

  abline(h = llik_int_bbar - qnorm(1-alpha/2), col="red")

  par(mar = c(5, 4, 4, 2))
}

if(print){
  cat("beta_bar = ", beta_bar,
```

```
        ", Z = ", stat,
        ", p = ", pval,
        ", CI:(", ll_ci$minimum,",",ul_ci$minimum,")\n")
  }

  #Return
  return(list(r_Int = stat, p = pval,
              CI_ll = ll_ci$minimum, CI_ul = ul_ci$minimum))

}



#Remove the comment sign at line 256 and 327 to run it.
'
################################################################################
# Example
################################################################################

#Setup
if(require(metafor)==F) install.packages("metafor")
if(require(metaLik)==F) install.packages("metaLik")
library(metafor)
library(metaLik)

#Data
col_names <- c("Study",
               "n_cocoa", "mean_delta_DBP_cocoa", "sd_delta_DBP_cocoa",
               "n_control", "mean_delta_DBP_control", "sd_delta_DBP_control")

df_cocoa <- matrix(c(1,13,-1.6,1.4,13,0.2,1.6,
                     2,11,0.9,2.3,10,-0.1,1.9,
                     3,15,-4.1,4.1,15,-0.6,2.1,
                     4,20,-6.2,4.2,20,-0.3,3.1,
                     5,28,-5.0,2.0,13,-1.0,2.0),
                   ncol = 7,
                   byrow = TRUE)

colnames(df_cocoa) <- col_names
df_cocoa <- as.data.frame(df_cocoa)

#Effect size calculation (Mean difference)

escalc(data = df_cocoa,
       m1i = mean_delta_DBP_cocoa,
       sd1i = sd_delta_DBP_cocoa,
       n1i = n_cocoa,
       m2i = mean_delta_DBP_control,
       sd2i = sd_delta_DBP_control,
       n2i = n_control,
       measure = "MD",
       append = TRUE) -> df_cocoa

#Random effect model (DL)
(fit_rma_cocoa <- rma(data = df_cocoa,
```

```
                      yi = yi,
                      vi = vi,
                      weighted = TRUE,
                      method = "REML"))

#Lik based inference

(fit_metalik_cocoa <- metaLik(yi ~ 1, sigma2 = vi, data = df_cocoa))

#Forest plot
forest(fit_rma_cocoa)

#Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta <- rma(data = df_cocoa,
                yi = yi,
                vi = vi,
                weighted = TRUE,
                method = "ML")$beta[1,1]

#Integrated likelihood results
r_int_beta_mod(beta_zero = 0,
               y=df_cocoa$yi,
               s2 = df_cocoa$vi,
               n1 = df_cocoa$n_cocoa,
               n2 = df_cocoa$n_control,
               hat_beta = hat_beta,
               k_opt = 3.5,
               alpha = 0.05,
               plot = TRUE, print = TRUE)
'
```

Appendix

Appendix 4

```
'
Title:        Integrated likelihood analysis
Author:       Enrico Roma
Description:  This script performs an integrated likelihood analysis on the
              the dataset Data_19062020.xlsx. The results are compared to the
              restricted maximum likelihood random-effects meta-analysis,
              proposed by the authors in the original paper.
'

# Workspace
rm(list=ls())
setwd("C:\\Users\\romae\\Desktop\\Analisi_tesi_Statistica")
image_path <- "C:\\Users\\romae\\Desktop\\Analisi_tesi_Statistica\\Grafici_tesi"

# Libraries
# install.packages(c("metafor", "metaLik", "metatest", "esc",
#                    "dplyr", "ggplot2", "readxl",
#                    "wrapr", "tidyverse"))

library(esc)
library(metafor)
library(metaLik)
library(metatest)
library(readxl)
library(dplyr)
library(xtable)
library(tidyverse)

# Hand written function
source("Integrated_likelihood_Cocoa_modified.R")

#Overall parameter setting
Alpha <- 0.05 #Sign. level

# Data
df <- read_xlsx("Data_190620.xlsx",
                sheet="Data_older",
                skip = 0,
                col_names=TRUE)
```

```
df <- as.data.frame(df)

round_df <- function(df, digits) {
  nums <- vapply(df, is.numeric, FUN.VALUE = logical(1))
  df[,nums] <- round(df[,nums], digits = digits)
  (df)
}

df <- round_df(df, digits = 2)


################################################################################
## Analysis 1a: Relative change upper body muscle strength
################################################################################

# Filter
df %>%
  filter(Age == "Older", Upper_strength == 1) -> df_upper

# Studies included
df_upper %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_upper %>%
  summarise(Upper_males = sum(Male_upper_strength_post_number),
            Upper_females = sum(Female_upper_strength_post_number))


# Random effect meta-analysis
escalc(data = df_upper,
       measure = "SMD",
       m1i = Male_upper_strength_relative_change_mean,
       sd1i = Male_upper_strength_relative_change_sd,
       n1i = Male_upper_strength_post_number,
       m2i = Female_upper_strength_relative_change_mean,
       sd2i = Female_upper_strength_relative_change_sd,
       n2i = Female_upper_strength_post_number,
       append = TRUE) -> df_upper

(rel_upper_rma <- rma(data = df_upper,
                      yi = yi,
                      vi = vi,
                      weighted = TRUE,
                      method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_rel_upper <- rma(data = df_upper,
                          yi = yi,
                          vi = vi,
                          weighted = TRUE,
                          method = "ML")$beta[1,1]
```

```
#Integrated likelihood results
pdf(paste(image_path,
          "Integrated_Likelihood_Chp3\\rel_upper_IntLik.pdf", sep="\\"),
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_upper$yi,
               s2 = df_upper$vi,
               n1 = df_upper$Male_upper_strength_post_number,
               n2 = df_upper$Female_upper_strength_post_number,
               hat_beta = hat_beta_rel_upper,
               k_opt = 3.5,
               alpha = Alpha,
               plot = TRUE, print = TRUE) -> rel_upper_intlik_res
dev.off()

rel_upper_intlik_res

##############################################################################
## Analysis 1b: Absolute change upper body muscle strength
##############################################################################

# Random effect meta-analysis
escalc(data = df_upper,
       measure = "SMD",
       m1i = Male_upper_strength_absolute_change_mean,
       sd1i = Male_upper_strength_absolute_change_sd,
       n1i = Male_upper_strength_post_number,
       m2i = Female_upper_strength_absolute_change_mean,
       sd2i = Female_upper_strength_absolute_change_sd,
       n2i = Female_upper_strength_post_number,
       append = TRUE) -> df_upper

(abs_upper_rma <- rma(data = df_upper,
                      yi = yi,
                      vi = vi,
                      weighted = TRUE,
                      method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_abs_upper <- abs_upper_rma$beta[1,1]

#Integrated likelihood results
pdf(paste(image_path,
          "Integrated_Likelihood_Chp3\\abs_upper_IntLik.pdf", sep="\\"),
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_upper$yi,
               s2 = df_upper$vi,
               n1 = df_upper$Male_upper_strength_post_number,
               n2 = df_upper$Female_upper_strength_post_number,
               hat_beta = hat_beta_abs_upper,
```

```
              k_opt = 3.5,
              alpha = Alpha,
              plot = TRUE, print = TRUE) -> abs_upper_intlik_res

dev.off()

abs_upper_intlik_res

################################################################################
## Analysis 2a: Relative change lower body muscle strength
################################################################################

# Filter
df %>%
  filter(Age == "Older", Lower_strength == 1) -> df_lower

#df_lower %>%
#  slice(c(1:31, 33, 35:nrow(df_lower))) -> df_lower

# Studies included
df_lower %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_lower %>%
  summarise(Lower_males = sum(Male_lower_strength_post_number),
            Lower_females = sum(Female_lower_strength_post_number))


# Random effect meta-analysis
escalc(data = df_lower,
       measure = "SMD",
       m1i = Male_lower_strength_relative_change_mean,
       sd1i = Male_lower_strength_relative_change_sd,
       n1i = Male_lower_strength_post_number,
       m2i = Female_lower_strength_relative_change_mean,
       sd2i = Female_lower_strength_relative_change_sd,
       n2i = Female_lower_strength_post_number,
       append = TRUE) -> df_lower

(rel_lower_rma <- rma(data = df_lower,
                      yi = yi,
                      vi = vi,
                      weighted = TRUE,
                      method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_rel_lower <- rel_lower_rma$beta[1,1]

#Integrated likelihood results

pdf(paste(image_path,
          "Integrated_Likelihood_Chp3\\rel_lower_IntLik.pdf", sep="\\"),
```

```
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_lower$yi,
               s2 = df_lower$vi,
               n1 = df_lower$Male_lower_strength_post_number,
               n2 = df_lower$Female_lower_strength_post_number,
               hat_beta = hat_beta_rel_lower,
               k_opt = 3.5,
               alpha = Alpha,
               plot = TRUE, print = TRUE) -> rel_lower_intlik_res

dev.off()

rel_lower_intlik_res

################################################################################
## Analysis 2b: Absolute change lower body muscle strength
################################################################################

# Random effect meta-analysis
escalc(data = df_lower,
       measure = "SMD",
       m1i = Male_lower_strength_absolute_change_mean,
       sd1i = Male_lower_strength_absolute_change_sd,
       n1i = Male_lower_strength_post_number,
       m2i = Female_lower_strength_absolute_change_mean,
       sd2i = Female_lower_strength_absolute_change_sd,
       n2i = Female_lower_strength_post_number,
       append = TRUE) -> df_lower

(abs_lower_rma <- rma(data = df_lower,
                      yi = yi,
                      vi = vi,
                      weighted = TRUE,
                      method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_abs_lower <- abs_lower_rma$beta[1,1]

#Integrated likelihood results

pdf(paste(image_path,
          "Integrated_Likelihood_Chp3\\abs_lower_IntLik.pdf", sep="\\"),
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_lower$yi,
               s2 = df_lower$vi,
               n1 = df_lower$Male_lower_strength_post_number,
               n2 = df_lower$Female_lower_strength_post_number,
               hat_beta = hat_beta_abs_lower,
               k_opt = 3.5,
```

```
               alpha = Alpha,
               plot = TRUE, print = TRUE) -> abs_lower_intlik_res

dev.off()

abs_lower_intlik_res

################################################################################
## Analysis 3a: Relative change lower body muscle size
################################################################################

# Filter
df %>%
  filter(Age == "Older", Muscle_size == 1) -> df_muscle

# Studies included
df_muscle %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_muscle %>%
  summarise(muscle_males = sum(Male_muscle_post_number),
            muscle_females = sum(Female_muscle_post_number))


# Random effect meta-analysis
escalc(data = df_muscle,
       measure = "SMD",
       m1i = Male_muscle_relative_change_mean,
       sd1i = Male_muscle_relative_change_sd,
       n1i = Male_muscle_post_number,
       m2i = Female_muscle_relative_change_mean,
       sd2i = Female_muscle_relative_change_sd,
       n2i = Female_muscle_post_number,
       append = TRUE) -> df_muscle

(rel_muscle_rma <- rma(data = df_muscle,
                       yi = yi,
                       vi = vi,
                       weighted = TRUE,
                       method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_rel_muscle <- rel_muscle_rma$beta[1,1]

#Integrated likelihood results

pdf(paste(image_path,
          "Integrated_Likelihood_Chp3\\rel_muscle_IntLik.pdf", sep="\\"),
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_muscle$yi,
```

```
                s2 = df_muscle$vi,
                n1 = df_muscle$Male_muscle_post_number,
                n2 = df_muscle$Female_muscle_post_number,
                hat_beta = hat_beta_rel_muscle,
                k_opt = 3.5,
                alpha = Alpha,
                plot = TRUE, print = TRUE) -> rel_muscle_intlik_res

dev.off()

rel_muscle_intlik_res

##############################################################################
## Analysis 3b: Absolute change muscle body muscle size
##############################################################################

# Filter
df_muscle  %>%
  filter(!is.na(Male_muscle_absolute_change_mean)) -> df_muscle_abs


# Studies included
df_muscle_abs %>%
  summarise(studies_included = length(unique(Study)))

# Participant in each group
df_muscle_abs %>%
  summarise(muscle_males = sum(Male_muscle_post_number),
            muscle_females = sum(Female_muscle_post_number))

# Random effect meta-analysis
escalc(data = df_muscle_abs,
       measure = "SMD",
       m1i = Male_muscle_absolute_change_mean,
       sd1i = Male_muscle_absolute_change_sd,
       n1i = Male_muscle_post_number,
       m2i = Female_muscle_absolute_change_mean,
       sd2i = Female_muscle_absolute_change_sd,
       n2i = Female_muscle_post_number,
       append = TRUE) -> df_muscle_abs

(abs_muscle_rma <- rma(data = df_muscle_abs,
                       yi = yi,
                       vi = vi,
                       weighted = TRUE,
                       method = "REML"))

# Integrated likelihood

#Calculation of maximum likelihood estimate for \beta
hat_beta_abs_muscle <- abs_muscle_rma$beta[1,1]

#Integrated likelihood results

pdf(paste(image_path,
```

```
          "Integrated_Likelihood_Chp3\\abs_muscle_IntLik.pdf", sep="\\"),
    width = 10, height = 8)

r_int_beta_mod(beta_zero = 0,
               y=df_muscle_abs$yi,
               s2 = df_muscle_abs$vi,
               n1 = df_muscle_abs$Male_muscle_post_number,
               n2 = df_muscle_abs$Female_muscle_post_number,
               hat_beta = hat_beta_abs_muscle,
               k_opt = 3.5,
               alpha = Alpha,
               plot = TRUE, print = TRUE) -> abs_muscle_intlik_res

abs_muscle_intlik_res

dev.off()
```

Bibliography

Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Macmillan.

Barndorff-Nielsen, O. E. (1986). Infereni on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307–322.

Bellio, R. & Guolo, A. (2016). Integrated likelihood inference in small sample meta-analysis for continuous outcomes. *Scandinavian Journal of Statistics* **43**, 191–201.

Bishop, P., Cureton, K. & Collins, M. (1987). Sex difference in muscular strength in equally-trained men and women. *Ergonomics* **30**, 675–687.

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C. & Biggeri, A. (2002). Some general points in estimating heterogeneity variance with the dersimonian–laird estimator. *Biostatistics* **3**, 445–457.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R. et al. (2020). World health organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine* **54**, 1451–1462.

Carpenter, J. & Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine* **19**, 1141–1164.

Casella, G. & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Chalmers, I., Hedges, L. V. & Cooper, H. (2002). A brief history of research synthesis. *Evaluation & the health professions* **25**, 12–37.

Cheng, S.-J., Yang, Y.-R., Cheng, F.-Y., Chen, I.-H. & Wang, R.-Y. (2014). The changes of muscle strength and functional activities during aging in male and female populations. *International Journal of Gerontology* **8**, 197–202.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society* **4**, 102–118.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10**, 101–129.

Cook, R. D. & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials* **7**, 177–188.

Di Gessa, G. (2008). *Simple strategies for variance uncertainty in meta-analysis*. Ph.D. thesis, University of Glasgow.

Dreyer, H. C., Fujita, S., Glynn, E. L., Drummond, M. J., Volpi, E. & Rasmussen, B. B. (2010). Resistance exercise increases leg muscle protein synthesis and mtor signalling independent of sex. *Acta physiologica* **199**, 71–81.

Du, Y., Wang, X., Xie, H., Zheng, S., Wu, X., Zhu, X., Zhang, X., Xue, S., Li, H., Hong, W. et al. (2019). Sex differences in the prevalence and adverse outcomes of sarcopenia and sarcopenic obesity in community dwelling elderly in east china using the awgs criteria. *BMC endocrine disorders* **19**, 1–11.

Egger, M., Smith, G. D., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj* **315**, 629–634.

Elwood, P. (2006). The first randomized trial of aspirin for heart attack and the advent of systematic overviews of trials. *Journal of the Royal Society of Medicine* **99**, 586–588.

Faigenbaum, A. D., Kraemer, W. J., Blimkie, C. J., Jeffreys, I., Micheli, L. J., Nitka, M. & Rowland, T. W. (2009). Youth resistance training: updated position statement paper from the national strength and conditioning association. *The Journal of Strength & Conditioning Research* **23**, S60–S79.

Feinstein, A. R. (1995). Meta-analysis: statistical alchemy for the 21st century. *Journal of clinical epidemiology* **48**, 71–79.

Fiatarone, M. A., Marks, E. C., Ryan, N. D., Meredith, C. N., Lipsitz, L. A. & Evans, W. J. (1990). High-intensity strength training in nonagenarians: effects on skeletal muscle. *Jama* **263**, 3029–3034.

Fielding, R. A., Vellas, B., Evans, W. J., Bhasin, S., Morley, J. E., Newman, A. B., van Kan, G. A., Andrieu, S., Bauer, J., Breuille, D. et al. (2011). Sarcopenia: an undiagnosed condition in older adults. current consensus definition: prevalence, etiology, and consequences. international working group on sarcopenia. *Journal of the American Medical Directors Association* **12**, 249–256.

Fisher, R. (1932). *Statistical Methods for research workers*. London: Oliver and Boyd.

Fisher, R. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Flores, D. F., Gentil, P., Brown, L. E., Pinto, R. S., Carregaro, R. L. & Bottaro, M. (2011). Dissociated time course of recovery between genders after resistance exercise. *The Journal of Strength & Conditioning Research* **25**, 3039–3044.

Follmann, D. A. & Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics* **55**, 732–737.

Fragala, M. S., Cadore, E. L., Dorgo, S., Izquierdo, M., Kraemer, W. J., Peterson, M. D. & Ryan, E. D. (2019). Resistance training for older adults: position statement from the national strength and conditioning association. *The Journal of Strength & Conditioning Research* **33**.

Frontera, W. R., Hughes, V. A., Fielding, R. A., Fiatarone, M. A., Evans, W. J. & Roubenoff, R. (2000). Aging of skeletal muscle: a 12-yr longitudinal study. *Journal of applied physiology* **88**, 1321–1326.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher* **5**, 3–8.

Goldberger, J. (1907). Typhoid bacillus-carriers. *Rosenau, Report on the Origin* **167**.

Guolo, A. (2012). Higher-order likelihood inference in meta-analysis and meta-regression. *Statistics in Medicine* **31**, 313–327.

Haizlip, K., Harrison, B. & Leinwand, L. (2015). Sex-based differences in skeletal muscle kinetics and fiber-type composition. *Physiology* **30**, 30–39.

Hardy, R. J. & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in medicine* **15**, 619–629.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American statistical association* **72**, 320–338.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *journal of Educational Statistics* **6**, 107–128.

Higgins, J. P. & Deeks, J. (2011). *Cochrane handbook for systematic reviews of interventions*, chap. 16.1.3.2. The Cochrane Collaboration. `https://training.cochrane.org/handbook/archive/v5.1/`, (visited 2022-11-01).

Higgins, J. P. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* **21**, 1539–1558.

Higgins, J. P. & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in medicine* **23**, 1663–1682.

Huizenga, H. M., Visser, I. & Dolan, C. V. (2011). Testing overall and moderator effects in random effects meta-regression. *British Journal of Mathematical and Statistical Psychology* **64**, 1–19.

Jackson, D. (2009). The significance level of the standard test for a treatment effect in meta-analysis. *Statistics in Biopharmaceutical Research* **1**, 92–100.

Jackson, D. & White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal* **60**, 1040–1058.

Johnson, B. T. & Huedo-Medina, T. B. (2013). Meta-analytic statistical inferences for continuous measure outcomes as a function of effect size metric and other assumptions. *Review from Agency for Healthcare Research and Quality (US), Rockville (MD)* .

Jones, M. D., Wewege, M. A., Hackett, D. A., Keogh, J. W. & Hagstrom, A. D. (2021). Sex differences in adaptations in muscle strength and size following resistance training in older adults: A systematic review and meta-analysis. *Sports Medicine* **51**, 503–517.

Kacker, R. N. (2004). Combining information from interlaboratory evaluations using a random effects model. *Metrologia* **41**, 132.

Lv, H. (1985). Olkin i. statistical methods for meta-analysis.

Malzahn, U., Böhning, D. & Holling, H. (2000). Nonparametric estimation of heterogeneity variance for the standardised difference used in meta-analysis. *Biometrika* **87**, 619–632.

Molenaar, J. P., McNeil, C. J., Bredius, M. S. & Gandevia, S. C. (2013). Effects of aging and sex on voluntary activation and peak relaxation rate of human elbow flexors studied with motor cortical stimulation. *Age* **35**, 1327–1337.

Morton, R. W., Sato, K., Gallaugher, M. P., Oikawa, S. Y., McNicholas, P. D., Fujita, S. & Phillips, S. M. (2018). Muscle androgen receptor content but not systemic hormones is associated with resistance training-induced skeletal muscle hypertrophy in healthy, young men. *Frontiers in physiology* **9**, 1373.

Murphy, S. A. & Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95**, 449–465.

of Sports Medicine, A. C. et al. (2009). American college of sports medicine position stand. progression models in resistance training for healthy adults. *Medicine and science in sports and exercise* **41**, 687–708.

O'rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine* **100**, 579–582.

Patterson, H. D. & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554.

Paule, R. C. & Mandel, J. (1982). Consensus values and weighting factors. *Journal of research of the National Bureau of Standards* **87**, 377.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British medical journal* **2**, 1243–1246.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R. & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of clinical epidemiology* **61**, 991–996.

Peto, R. (1987). Why do we need systematic overviews of randomized trials?(transcript of an oral presentation, modified by the editors). *Statistics in medicine* **6**, 233–240.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W. & Kahaner, D. K. (2012). *Quadpack: a subroutine package for automatic integration*, vol. 1. Springer Science & Business Media.

Porta, M. (2014). *A dictionary of epidemiology*. Oxford university press.

Rencher, A. C. & Schaalje, G. B. (2008). *Linear models in statistics*. John Wiley & Sons.

Rice, K., Higgins, J. P. & Lumley, T. (2018). A re-evaluation of fixed effect (s) meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **181**, 205–227.

Roberts, B. M., Nuckols, G. & Krieger, J. W. (2020). Sex differences in resistance training: a systematic review and meta-analysis. *The Journal of Strength & Conditioning Research* **34**, 1448–1460.

Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H. & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics* **12**, 122–142.

Schmid, C. H., Stijnen, T. & White, I. (2020). *Handbook of Meta-analysis*. CRC Press.

Schoenfeld, B. J. (2010). The mechanisms of muscle hypertrophy and their application to resistance training. *The Journal of Strength & Conditioning Research* **24**, 2857–2872.

Scott, B. R., Duthie, G. M., Thornton, H. R. & Dascombe, B. J. (2016). Training monitoring for resistance exercise: theory and applications. *Sports medicine* **46**, 687–698.

Senn, S. (2016). Mastering variation: variance components and personalised medicine. *Statistics in medicine* **35**, 966–977.

Severini, T. A. (2000). *Likelihood methods in statistics*. Oxford University Press.

Severini, T. A. (2007). Integrated likelihood functions for non-bayesian inference. *Biometrika* **94**, 529–542.

Severini, T. A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika* **97**, 481–496.

Sharma, G. & Mathew, T. (2011). Higher order inference for the consensus mean in inter-laboratory studies. *Biometrical Journal* **53**, 128–136.

Skovgaard, I. M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2**, 145–165.

Steele, J., Fisher, J., Giessing, J. & Gentil, P. (2017). Clarity in reporting terminology and definitions of set endpoints in resistance training. *Muscle & nerve* **56**, 368–374.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association* **54**, 30–34.

Stijnen, T., Hamza, T. H. & Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in medicine* **29**, 3046–3067.

Stupka, N., Lowther, S., Chorneyko, K., Bourgeois, J., Hogben, C. & Tarnopolsky, M. (2000). Gender differences in muscle inflammation after eccentric exercise. *Journal of applied physiology* **89**, 2325–2332.

Suleman, M., Clark, M., Goldacre, M. & Burton, M. (2010). Exploring the variation in paediatric tonsillectomy rates between english regions: a 5-year nhs and independent sector data analysis. *Clinical Otolaryngology* **35**, 111–117.

Van Den Noortgate, W. & Onghena, P. (2005). Parametric and nonparametric bootstrap methods for meta-analysis. *Behavior Research Methods* **37**, 11–22.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D. & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods* **7**, 55–79.

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in medicine* **26**, 37–52.

Viechtbauer, W. (2010a). Conducting meta-analyses in r with the metafor package. *Journal of statistical software* **36**, 1–48.

Viechtbauer, W. (2010b). Metafor: meta-analysis package for r. *R package version* **2010**, 1–0.

Viechtbauer, W. & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research synthesis methods* **1**, 112–125.

Winkelstein, W. (1998). The first use of meta-analysis? *American journal of epidemiology* **147**, 717–717.

Yates, F. & Cochran, W. G. (1938). The analysis of groups of experiments. *The Journal of Agricultural Science* **28**, 556–580.

Yeung, S. S., Reijnierse, E. M., Pham, V. K., Trappenburg, M. C., Lim, W. K., Meskers, C. G. & Maier, A. B. (2019). Sarcopenia and its association with falls and fractures in older adults: a systematic review and meta-analysis. *Journal of cachexia, sarcopenia and muscle* **10**, 485–500.

Zucker, D. M., Lieberman, O. & Manor, O. (2000). Improved small sample inference in the mixed linear model: Bartlett correction and adjusted likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 827–838.