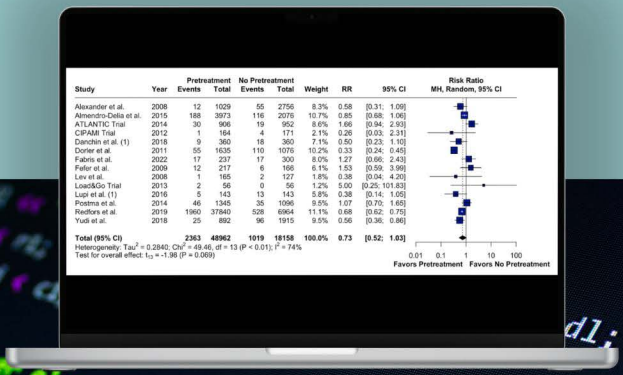
Atena Editora
Year 2026

MASTERING

PAIRWISE META-ANALYSIS

in RStudio R



AN INTUITIVE AND STRAIGHTFORWARD GUIDE

DOUGLAS MESADRI GEWEHR
EMILTON LIMA JUNIOR

Editor-in-chief
Prof. Dr. Antonella Carvalho de Oliveira
Executive Editor
Natalia Oliveira Scheffer

Atena Editora is firmly committed to editorial integrity at all stages of the publication process, ensuring that ethical and academic standards are strictly adhered to. It adopts policies to prevent and combat practices such as plagiarism, manipulation or falsification of data and results, as well as any undue interference from financial or institutional interests.

Any suspicion of scientific misconduct is treated with the utmost seriousness and will be investigated in accordance with the highest standards of academic rigor, transparency, and ethics.

The content of the work and its data, in terms of form, accuracy, and reliability, are the sole responsibility of the author and do not necessarily represent the official position of Atena Editora. The download, sharing, adaptation, and reuse of this work are permitted for any purpose, provided that the appropriate authorship and reference to the publisher are attributed, in accordance with the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0).

National works were submitted to blind peer review by members of the publisher's Editorial Board, while international works were reviewed by external reviewers. All were approved for publication based on criteria of neutrality and academic impartiality.

Mastering pairwise meta-analysis in RStudio: a straightforward guide

| Organizers:
Douglas Mesadri Gewehr
Emilton Lima Junior

| Review:
The authors

| Layout and Cover:
Douglas Mesadri Gewehr
Emilton Lima Junior

International Cataloging-in-Publication Data (CIP)

G396 Gewehr, Douglas Mesadri
Mastering pairwise meta-analysis in RStudio: a straightforward guide / Douglas Mesadri Gewehr. – Ponta Grossa - PR: Atena, 2026.

Format: PDF
System requirements: Adobe Acrobat Reader
Access mode: World Wide Web
Includes bibliography
ISBN 978-65-258-4045-1
DOI: https://doi.org/10.22533/at.ed.451260203

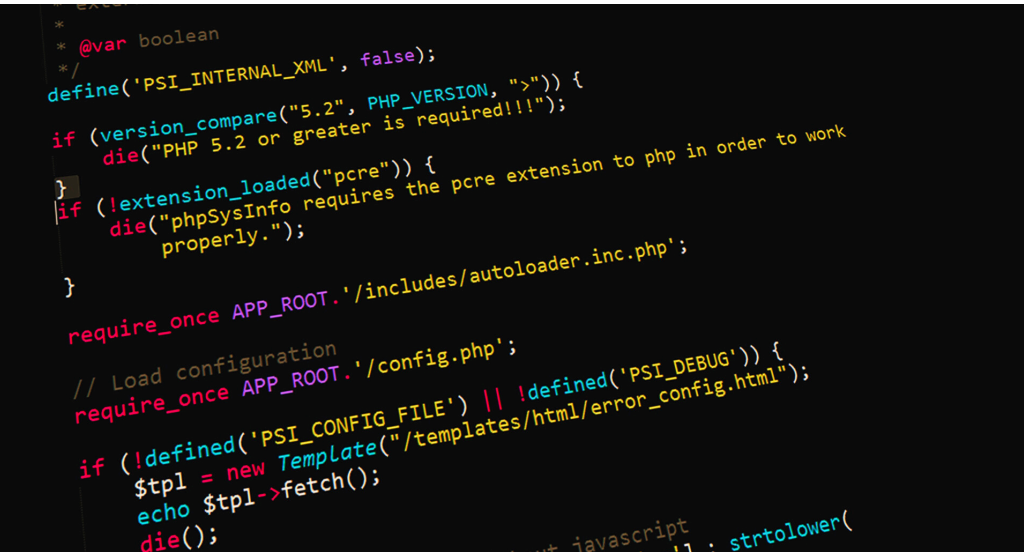
1. Meta-analysis. 2. Statistical computing. I. Gewehr, Douglas Mesadri. II. Título.
CDD 519.5

Prepared by Librarian Janaina Ramos – CRB-8/9166

Atena Editora
+55 (42) 3323-5493
+55 (42) 99955-2866
www.atenaeditora.com.br
contato@atenaeditora.com.br

EDITORIAL BOARD

SUMMARY

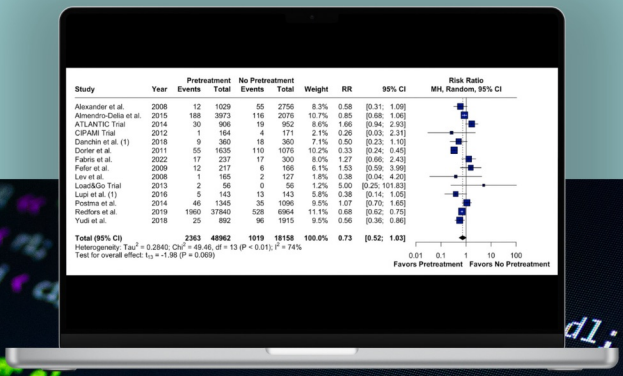


MASTERING

PAIRWISE META-ANALYSIS

CHAPTER 1: BINARY OUTCOME DATA

in RStudio

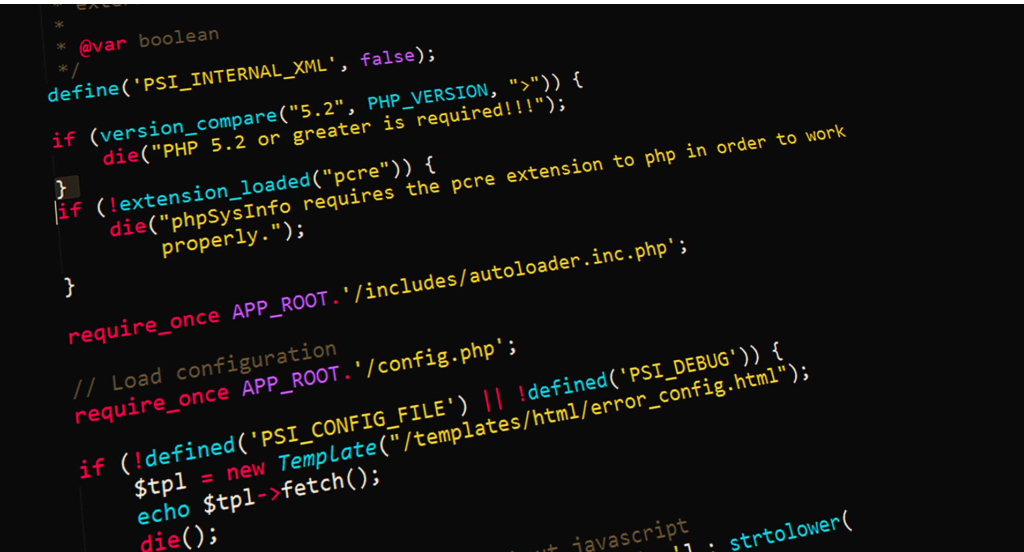


AN INTUITIVE AND STRAIGHTFORWARD GUIDE

DOUGLAS MESADRI GEWEHR

EMILTON LIMA JUNIOR

SUMMARY



I – Essential R Packages

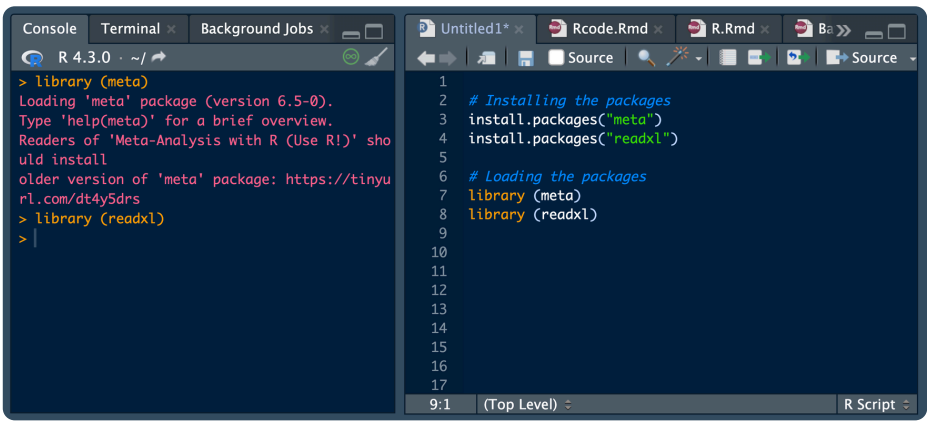
readxl package

A package that provides functions to read data from Excel files into R. It supports both ".xls" and ".xlsx" file formats, and can read data from individual worksheets or the entire workbook.

meta package

This package is a comprehensive package for meta-analysis, including functions for estimating and plotting effects, performing tests for heterogeneity, adjusting meta-regression models, and conducting subgroup analyses.

First, you'll need to **install** and **load** the packages as follows:



Once you install the packages with the **install.packages()** function, you need to load ("activate") the packages, using the **library()** function every time you start RStudio!

QUICK TIPS

II – Preparing your Dataset

Before you begin your meta-analysis, you need to ensure that your data has been organized and structured correctly!

This will help ensure that the data is consistent and compatible with the functions and methods used in the meta-analysis, and can save time and effort in the analysis process.

Here are some **Dos and Don'ts** when preparing your data in Excel:

1. It's very important how you name the columns of your spreadsheet. If you already named the columns of your sheet appropriately in **Excel**, you can save a lot of time later because your data does not have to be transformed using R.

2. Column names should not contain any spaces. An alternative is to use underline sign "_".

3. It does not matter how columns are ordered in your Excel spreadsheet. They just have to be labeled correctly.

4. It is also important to know that the import may distort special characters like ä, ü, ö, and so forth. You might want to transform them into "normal" letters before you proceed.

5. If you have one or several empty rows or columns which used to contain data, make sure to delete those columns/rows completely because R could think that these columns contain (missing) data and import them also.

II – Preparing your Dataset

Structure your dataset following the example below:

N. events experimental group

Sample size experimental group

Subgroups

Study Identification

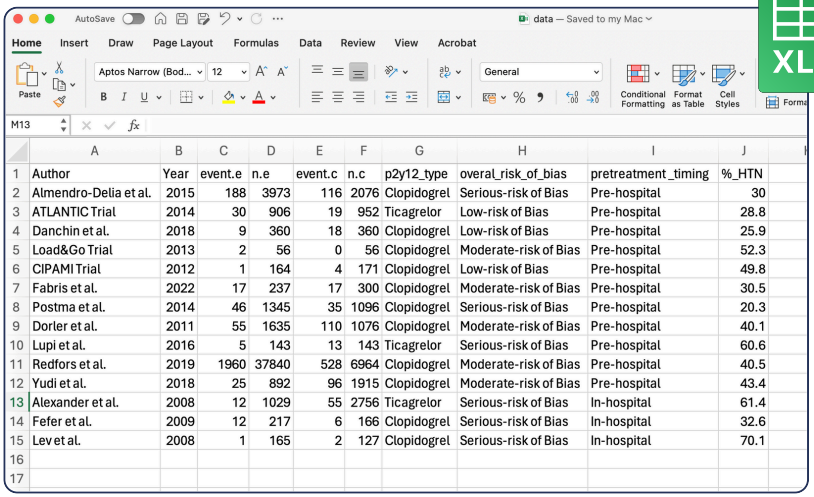
Author	Year	event.e	n.e	event.c	n.c	p2y12_type	overal_risk_of_bias	pretreatment _timing	%_HTN
Almendro-Delia et al.	2015	188	3973	116	2076	Clopidogrel	Serious-risk of Bias	Pre-hospital	30.0
ATLANTIC Trial	2014	30	906	19	952	Ticagrelor	Low-risk of Bias	Pre-hospital	28.8
Danchin et al.	2018	9	360	18	360	Clopidogrel	Low-risk of Bias	Pre-hospital	25.9
Load&Go Trial	2013	2	56	0	56	Clopidogrel	Moderate-risk of Bias	Pre-hospital	52.3
CIPAMI Trial	2012	1	164	4	171	Clopidogrel	Low-risk of Bias	Pre-hospital	49.8
Fabris et al.	2022	17	237	17	300	Clopidogrel	Moderate-risk of Bias	Pre-hospital	30.5
Postma et al.	2014	46	1345	35	1096	Clopidogrel	Serious-risk of Bias	Pre-hospital	20.3
Dorler et al.	2011	55	1635	110	1076	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.1
Lupi et al.	2016	5	143	13	143	Ticagrelor	Serious-risk of Bias	Pre-hospital	60.6
Redfors et al.	2019	1960	37840	528	6964	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.5
Yudi et al.	2018	25	892	96	1915	Clopidogrel	Moderate-risk of Bias	Pre-hospital	43.4
Alexander et al.	2008	12	1029	55	2756	Ticagrelor	Serious-risk of Bias	In-hospital	61.4
Fefer et al.	2009	12	217	6	166	Clopidogrel	Serious-risk of Bias	In-hospital	32.6
Lev et al.	2008	1	165	2	127	Clopidogrel	Serious-risk of Bias	In-hospital	70.1

Covariates for meta-regression

Sample size control group

N. events control group

XLSX



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	event.e	n.e	event.c	n.c	p2y12_type	overal_risk_of_bias	pretreatment_timing	%_HTN
2	Almendro-Delia et al.	2015	188	3973	116	2076	Clopidogrel	Serious-risk of Bias	Pre-hospital	30
3	ATLANTIC Trial	2014	30	906	19	952	Ticagrelor	Low-risk of Bias	Pre-hospital	28.8
4	Danchin et al.	2018	9	360	18	360	Clopidogrel	Low-risk of Bias	Pre-hospital	25.9
5	Load&Go Trial	2013	2	56	0	56	Clopidogrel	Moderate-risk of Bias	Pre-hospital	52.3
6	CIPAMI Trial	2012	1	164	4	171	Clopidogrel	Low-risk of Bias	Pre-hospital	49.8
7	Fabris et al.	2022	17	237	17	300	Clopidogrel	Moderate-risk of Bias	Pre-hospital	30.5
8	Postma et al.	2014	46	1345	35	1096	Clopidogrel	Serious-risk of Bias	Pre-hospital	20.3
9	Dorler et al.	2011	55	1635	110	1076	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.1
10	Lupi et al.	2016	5	143	13	143	Ticagrelor	Serious-risk of Bias	Pre-hospital	60.6
11	Redfors et al.	2019	1960	37840	528	6964	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.5
12	Yudi et al.	2018	25	892	96	1915	Clopidogrel	Moderate-risk of Bias	Pre-hospital	43.4
13	Alexander et al.	2008	12	1029	55	2756	Ticagrelor	Serious-risk of Bias	In-hospital	61.4
14	Fefer et al.	2009	12	217	6	166	Clopidogrel	Serious-risk of Bias	In-hospital	32.6
15	Lev et al.	2008	1	165	2	127	Clopidogrel	Serious-risk of Bias	In-hospital	70.1
16										
17										

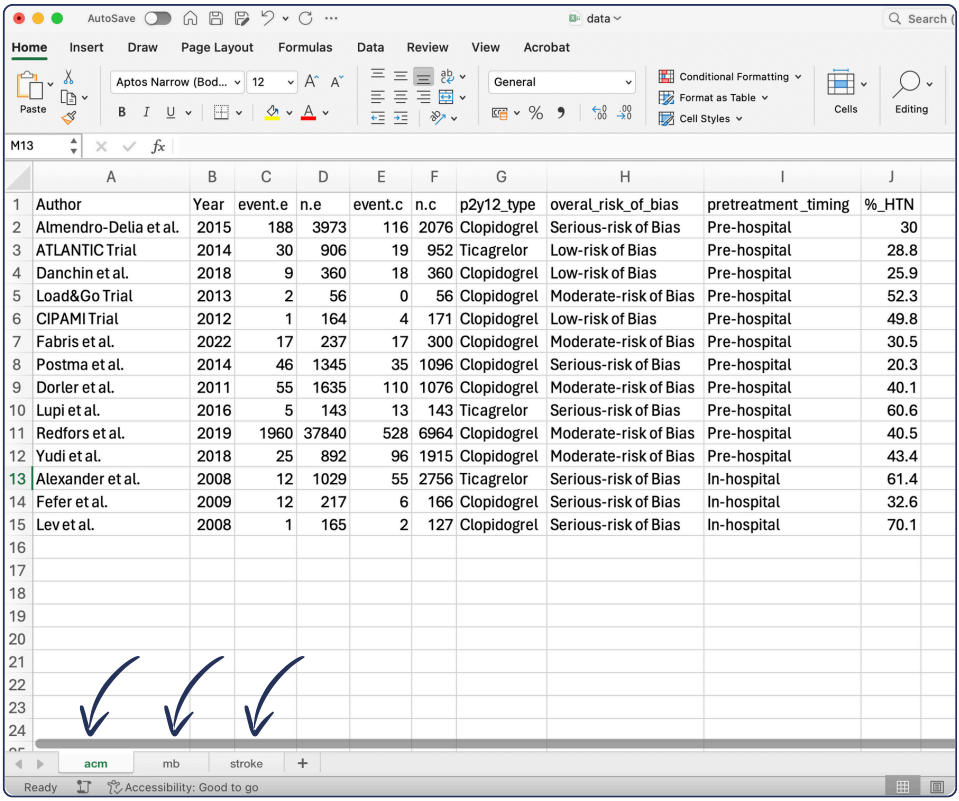
Here's how your Excel spreadsheet should look

II – Preparing your Dataset

Create a spreadsheet for each outcome of your meta-analysis in separate worksheets in your Excel workbook!

This approach saves time by allowing you to import **ALL** your endpoints at once into RStudio.

Here is an example of a dataset with 3 endpoints, each one in its own worksheet: All-cause Mortality (acm), Major Bleeding (mb), and Stroke (stroke).



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	event.e	n.e	event.c	n.c	p2y12_type	overal_risk_of_bias	pretreatment_timing	%_HTN
2	Almendro-Delia et al.	2015	188	3973	116	2076	Clopidogrel	Serious-risk of Bias	Pre-hospital	30
3	ATLANTIC Trial	2014	30	906	19	952	Ticagrelor	Low-risk of Bias	Pre-hospital	28.8
4	Danchin et al.	2018	9	360	18	360	Clopidogrel	Low-risk of Bias	Pre-hospital	25.9
5	Load&Go Trial	2013	2	56	0	56	Clopidogrel	Moderate-risk of Bias	Pre-hospital	52.3
6	CIPAMI Trial	2012	1	164	4	171	Clopidogrel	Low-risk of Bias	Pre-hospital	49.8
7	Fabris et al.	2022	17	237	17	300	Clopidogrel	Moderate-risk of Bias	Pre-hospital	30.5
8	Postma et al.	2014	46	1345	35	1096	Clopidogrel	Serious-risk of Bias	Pre-hospital	20.3
9	Dorler et al.	2011	55	1635	110	1076	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.1
10	Lupi et al.	2016	5	143	13	143	Ticagrelor	Serious-risk of Bias	Pre-hospital	60.6
11	Redfors et al.	2019	1960	37840	528	6964	Clopidogrel	Moderate-risk of Bias	Pre-hospital	40.5
12	Yudi et al.	2018	25	892	96	1915	Clopidogrel	Moderate-risk of Bias	Pre-hospital	43.4
13	Alexander et al.	2008	12	1029	55	2756	Ticagrelor	Serious-risk of Bias	In-hospital	61.4
14	Fefer et al.	2009	12	217	6	166	Clopidogrel	Serious-risk of Bias	In-hospital	32.6
15	Lev et al.	2008	1	165	2	127	Clopidogrel	Serious-risk of Bias	In-hospital	70.1

III – Importing your Dataset

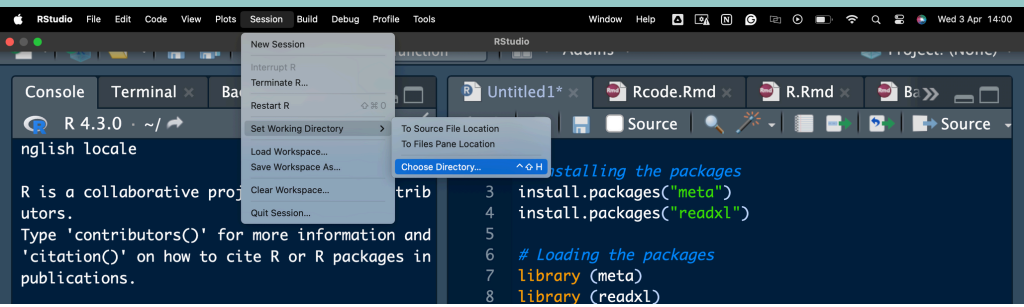
1ST

Once you have your Excel workbook with all endpoints of your meta-analysis in its own worksheets ready, save the **.xlsx** file in a designated folder, commonly referred to as the '**working directory**' (WD), or any other preferred name.

For that, you need to use the **setwd()** function and add between quotes the directory of the folder WD in your computer that you've created:

```
#Setting your WD folder in R Studio:
setwd("/Users/douglasmesadrigewehr/Desktop/WD RStudio")
```

The other way is to manually configure the WD folder, as shown below:



2

The following R code creates a new object with the name of your choice (in this example I am using '**ma**', containing all the outcomes in a data frame object.

```
sheet_names <- excel_sheets("data.xlsx")
sheet_names

ma <- lapply(sheet_names, function(x) {
  as.data.frame(read_excel("data.xlsx", sheet = x)) } )

names(ma) <- sheet_names
```

ATTENTION PLEASE!

When using this code, you just need to rename the **.xlsx** file (arrows) that contains your dataset, located in your WD folder.

IV – Visualizing your Data

Now, we can have a look at each of the lists (our endpoints) of our data frame object.

The object **ma,** we've just imported, has **3** lists of data:

1. "**acm**" for All-cause Mortality data
2. "**mb**" for Major Bleeding data
3. "**stroke**" for Stroke data

To print one specific endpoint, you can do so by adding the dollar sign operator "**$**" after "**ma**" inside de **View()** function:

```
View(ma$acm) #to visualize All-cause Mortality data
View(ma$mb)  #to visualize Major Bleeding data
View(ma$stroke)  #to visualize Stroke data
```

This is how your endpoint dataset will appear in RStudio.
See the example below:

EXAMPLE

Script Meta-Analysis.R | Untitled1* | ma$acm

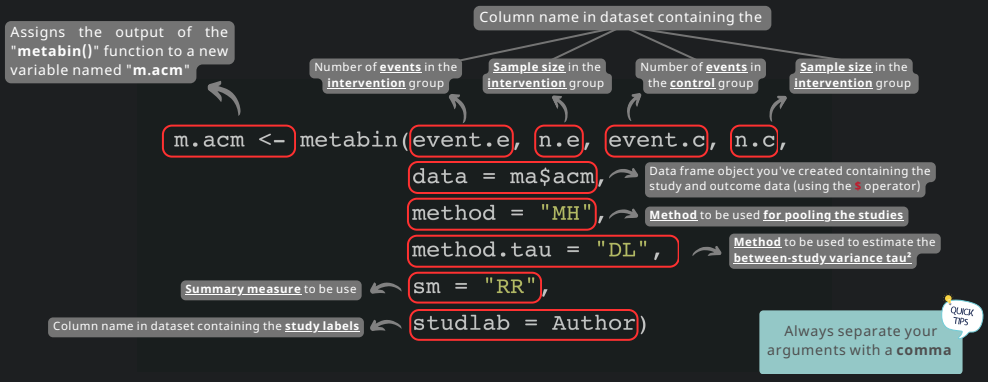
Filter

	Author	Year	event.e	n.e	event.c	n.c	follow_up	risk_of_bias	definition
1	Danchin et al.	2018	10	360	7	360	In-hospital	Low-risk of Bias	Pre-hospital
2	Almendro-Delia et al.	2015	26	3973	16	2076	In-hospital	Serious-risk of Bias	Pre-hospital
3	Fabris et al.	2022	5	237	6	30	In-hospital	Moderate-risk of Bias	Pre-hospital
4	CIPAMI Trial	2012	15	164	14	171	In-hospital	Low-risk of Bias	Pre-hospital
5	ATLANTIC Trial	2014	12	908	12	950	30-day	Low-risk of Bias	Pre-hospital
6	Load&Go Trial	2013	1	56	0	56	30-day	Moderate-risk of Bias	Pre-hospital
7	Postma et al.	2014	11	1364	16	1075	30-day	Serious-risk of Bias	Pre-hospital
8	Dorler et al.	2011	16	1635	15	1076	In-hospital	Moderate-risk of Bias	Pre-hospital
9	Lupi et al.	2016	18	143	19	143	In-hospital	Serious-risk of Bias	Pre-hospital
10	Redfors et al.	2019	966	37840	238	6964	In-hospital	Moderate-risk of Bias	Pre-hospital
11	Yudi et al.	2018	32	892	75	1915	In-hospital	Moderate-risk of Bias	Pre-hospital
12	Alexander et al.	2008	98	1029	306	2756	In-hospital	Serious-risk of Bias	In-hospital
13	Fefer et al.	2009	3	217	1	166	In-hospital	Serious-risk of Bias	In-hospital

V – Performing the Meta-Analysis

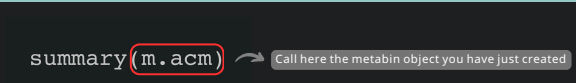
To perform the meta-analysis of binary outcome data we will use the **metabin ()** function from the '**meta**' package

We need to provide some **instructions** for the **metabin ()** function. These instructions are known as **arguments**, which can be one, two, three, or more inputs that the function uses to perform its task. The main arguments of the **metabin()** function are:



Over the next pages, we will detail the following aspects of binary outcome meta-analysis:
(1) methods for **pooling the studies**;
(2) methods to estimate the **between-study variance tau²**; and
(3) the **summary measures**.

To visualize the results of a meta-analysis conducted with the **metabin ()** function, we will use the **summary()** function. It will generate a summary of the analysis



The **summary()** function will display the estimated treatment effect (i.e., the summary measure), the corresponding confidence interval, the between-study variance, and other relevant statistics.

MORE ON THE OTHER SIDE

V – Performing the meta-analysis

Here is an example of the output of the **summary()** function in the **RStudio Console**

This section contains:
(1) the individual studies;
(2) their effect sizes and confidence intervals;
(3) their weights (%) in common (i.e. fixed-) and random-effects

Console | Terminal | Background Jobs

R 4.2.2 · ~/Desktop/WD RStudio/

```
> summary(m.acm)
```

	RR	95%-CI	%W(common)	%W(random)
Almendro-Delia et al.	0.8469	[0.6761; 1.0608]	11.0	13.0
ATLANTIC Trial	1.6591	[0.9407; 2.9262]	1.3	8.3
Danchin et al.	0.5000	[0.2277; 1.0981]	1.3	5.9
Load&Go Trial	5.0000	[0.2455; 101.8321]	0.0	0.6
CIPAMI Trial	0.2607	[0.0294; 2.3078]	0.3	1.2
Fabris et al.	1.2658	[0.6606; 2.4255]	1.1	7.3
Postma et al.	1.0710	[0.6951; 1.6502]	2.8	10.1
Dorler et al.	0.3291	[0.2403; 0.4506]	9.6	11.8
Lupi et al.	0.3846	[0.1408; 1.0508]	0.9	4.3
Redfors et al.	0.6832	[0.6227; 0.7495]	64.4	14.2
Yudi et al.	0.5591	[0.3626; 0.8619]	4.4	10.1
Alexander et al.	0.5844	[0.3142; 1.0867]	2.2	7.6
Fefer et al.	1.5300	[0.5864; 3.9914]	0.5	4.6
Lev et al.	0.3848	[0.0353; 4.1970]	0.2	1.0

```
Number of studies combined: k = 14
Number of observations: o = 67120
Number of events: e = 3382
```

The **total number of studies** in our meta-analysis
The **total number of patients** in our meta-analysis
The **total number of events** in our meta-analysis

	RR	95%-CI	z	p-value
Common effect model	0.6887	[0.6380; 0.7433]	-9.57	< 0.0001
Random effects model	0.7274	[0.5673; 0.9327]	-2.51	0.0121

The POOLED effect size

```
Quantifying heterogeneity:
 tau^2 = 0.1108; tau = 0.3328; I^2 = 73.7% [55.4%; 84.5%]; H = 1.95 [1.50; 2.54]

Test of heterogeneity:
     Q d.f.  p-value
 49.46   13 < 0.0001
```

Results concerning the **between-study heterogeneity**

```
Details on meta-analytical method:
- Mantel-Haenszel method
- DerSimonian-Laird estimator for tau^2
- Mantel-Haenszel estimator used in calculation of Q and tau^2 (like RevMan 5)
- Continuity correction of 0.5 in studies with zero cell frequencies
```

Method to be used **for pooling the studies**

Method to be used to estimate the **between-study variance tau^2**

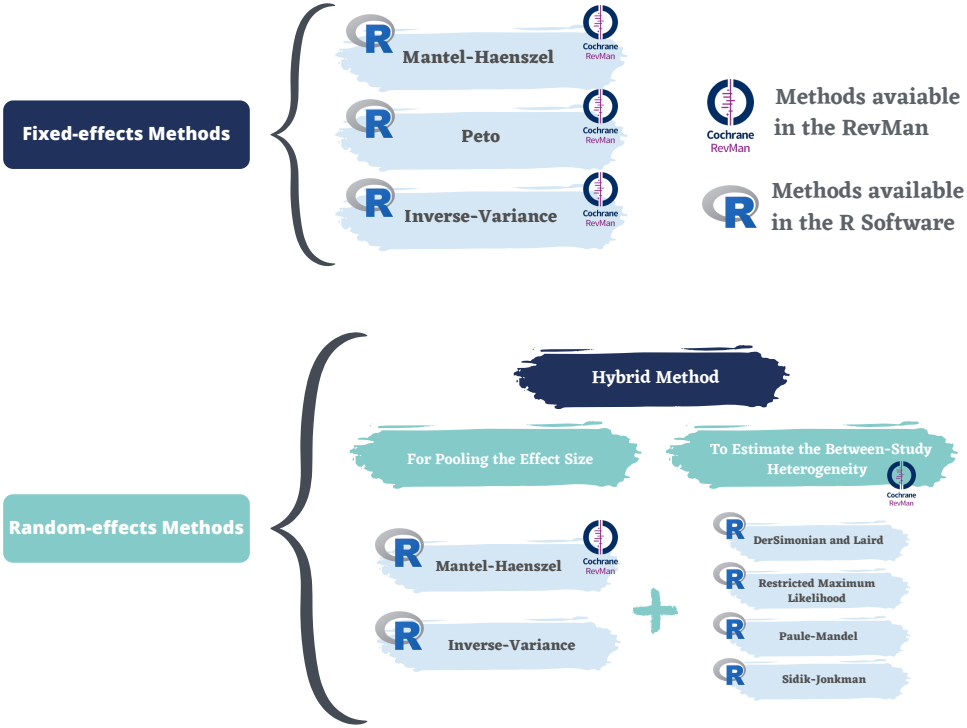
LET'S DIVE DEPPER into the **methods** for **pooling studies** with **binary data**

MORE ON THE OTHER SIDE

V – Performing the meta-analysis

OVERVIEW

Methods for pooling the effect sizes
Methods to estimate the between-study variance



LEARNING R SYNTAX

Methods for pooling the effect sizes

NEXT

```
method = "MH" # the Mantel-Haenszel method is used
method = "Inverse" # the Generic Inverse-Variance method is used
method = "Peto" # the Peto method is used
method = "SSW" # the Bakbergenuly-Sample method is used
```

Always use “ ” when referring to these methods

QUICK TIPS

V – Performing the meta-analysis

METHODS FOR POOLING THE EFFECT SIZES

Which one to choose?

Mantel-Haenszel Method ("MH")

RECOMMENDED

- **Default** method in **metabin()** and **recommended** approach (by the Cochrane) for binary data
- This method uses the number of events and non-events in the treatment and control groups to determine a study's weight.
- It is more suitable than the **Inverse-Variance** method for determining pooling weights when dealing with **small event number** and **sparse data**.

Generic Inverse-Variance Method ("Inverse")

- This approach is **suboptimal** for **binary outcome** data, since we are usually dealing with sparse data.
- When we are dealing with sparse data, the number of events or the total sample sizes of a study is small, the calculated standard error may not be a good estimator of the precision of binary effect size.

Peto Method ("Peto")

- This method only works well when:

1. The desired effect size is **Odds Ratio (OR)**;
2. The between-study **heterogeneity** is **very low** or **zero**, because it estimates **OR** only by **fixed-effects** model;
3. The **number of events** in the treatment and control are **balanced** (i.e. similar);
4. The treatment effect is not overly large.

RECAP

According to Cochrane, the **Peto** may be used when the **OR** is the **desired** effect size metric, and when the **event** of interest is expected to be **rare**.

Bakbergenuly-Sample Method is a fairly new approach, meaning that is not well studied as the other two methods.

V – Performing the meta-analysis

LEARNING R SYNTAX

Methods to estimate the between-study variance

```
method.tau = "REML" # Restricted maximum-likelihood estimator
method.tau = "ML" # Maximum-likelihood estimator
method.tau = "PM" # Paule-Mandel estimator
method.tau = "DL" # DerSimonian-Laird estimator
method.tau = "SJ" # Sidik-Jonkman estimator
method.tau = "HE" # Hedges estimator
method.tau = "EB" # Empirical Bayes estimator
```

Always use “ ” when referring to these methods

QUICK TIPS

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

There are no strict guidelines determining the choice of estimator. In many cases, there are only slight variations in results among different estimators. So, you **shouldn't** worry **TOO** much about it!

If doubt arise, you can always rerun your analyses using different estimators, and check if it changes the interpretation of your results.

In the next page you will find a **tentative algorithm** to support the choice of an estimator based on the **data scenario**

NEXT

V – Performing the meta-analysis

SIMPLIFIED ALGORITHM

Which estimator should we use for binary data?

1. Binary outcome?

   Yes

2. Is the **heterogeneity** of effects in your sample **very large**? Is **avoiding** false-positives a **priority**?

   No

   Yes → Sidik-Jonkman (**"SJ"**)

3. Is there **extreme variation** (small and large studies) in the **sample sizes** among included studies?

   No

   Yes → DerSimonian-Laird (**"DL"**)

   or

   Restricted Maximum Likelihood (**"REML"**)

BEST CHOICE **Paule-Mandel** (**"PM"**)

DerSimonian-Laird (**"DL"**) estimator can be **biased**, particularly when the number of studies is **small** and **heterogeneity** is **high.**

There's no one-size-fits-all solution!

V – Performing the meta-analysis

FINALLY ...

WE NEED TO DECIDE IF WE NEED TO APPLY THE SO-CALLED

KNAPP-HARTUNG

ADJUSTMENTS

What effect does it have on meta-analysis?

These adjustments affect the way the **standard error** (and thus the **confidence intervals**) of our pooled effect size is calculated.

Usually leading to **WIDER confidence intervals** of the pooled effect

How does this method do it?

Knapp-Hartung method uses quantiles of the **t-distribution** rather than the standard normal distribution in the more conventional method when computing a confidence interval (CI) for the average effect.

Why to apply this method?

The Knapp-Hartung adjustments try to **control for the uncertainty** in between-study heterogeneity → **widening** the **confidence interval** → **reducing** the chance of **false-positives**

In which situations is it most appropriate to use this method?

1) When working with random-effects model **AND**

2) We include a small number of studies our meta-analysis **AND**

3) The included studies have similar sample sizes

WARNING Extra caution is needed when there are ⩽ 5 studies of very unequal sizes

NEXT

LERN HOW TO

TO APPLY THE

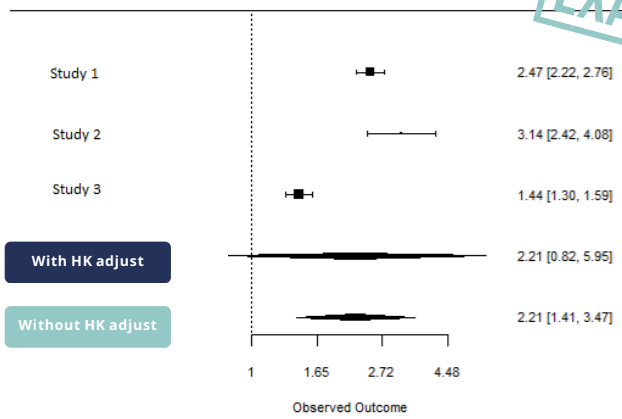
KNAPP-HARTUNG

ADJUSTMENTS

```
m.acm <- metabin(event.e, n.e, event.c, n.c,
                 data = ma$acm,
                 method = "MH",
                 method.tau = "DL",
                 sm = "RR",
                 hakn = TRUE,
                 studlab = Author)
```

To apply the **Knapp-Hartung** adjustments when working with random effects model, you just need to add the argument highlighted in **red** in the **metabin()** function

EXAMPLE

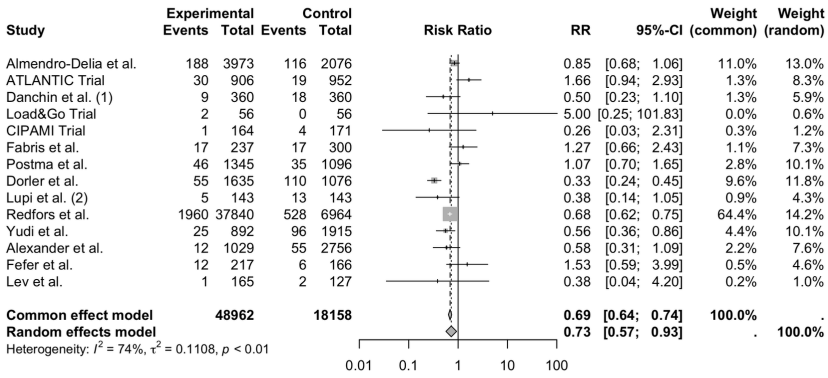


VI – Creating the Forest Plot

You can produce a forest plot for **metabin()** object using the **forest()** function from the '**meta**' package

Usually, these forest plots do not have enough quality for publication by default! In the example below, we plot the '**m.acm**' object that we also used in the previous example:

```
# Creating a forest plot for 'm.acm' metabin object
forest(m.acm)
```



However, the **forest()** function also has countless additional arguments to further customize the appearance of the forest plot.

All these arguments are detailed and described in the function documentation. You can accesses this documentation by running the following code:

```
help(forest)
```

VI – Creating the Forest Plot

But don't worry. I've put together a code that brings together the main arguments for generating **elegant** and **high-quality** forest plots for publication.

Now, let's understand the most important arguments to customize your forest plot!

This is how your code looks in the end:

MORE ON THE OTHER SIDE

VI – Creating the Forest Plot

This is how our forest plot looks in the end:

Study	Year	Experimental (2) Events	Total	Control (3) Events	Total	Weight	RR	95% CI
CIPAMI Trial	2012	1	164	4	171	1.2%	0.26	[0.03; 2.31]
Dorler et al.	2011	55	1635	110	1076	11.8%	0.33	[0.24; 0.45]
Lupi et al. (2)	2016	5	143	13	143	4.3%	0.38	[0.14; 1.05]
Lev et al.	2008	1	165	2	127	1.0%	0.38	[0.04; 4.20]
Danchin et al. (1)	2018	9	360	18	360	5.9%	0.50	[0.23; 1.10]
Yudi et al.	2018	25	892	96	1915	10.1%	0.56	[0.36; 0.86]
Alexander et al.	2008	12	1029	55	2756	7.6%	0.58	[0.31; 1.09]
Redfors et al.	2019	1960	37840	528	6964	14.2%	0.68	[0.62; 0.75]
Almendro-Delia et al.	2015	188	3973	116	2076	13.0%	0.85	[0.68; 1.06]
Postma et al.	2014	46	1345	35	1096	10.1%	1.07	[0.70; 1.65]
Fabris et al.	2022	17	237	17	300	7.3%	1.27	[0.66; 2.43]
Fefer et al.	2009	12	217	6	166	4.6%	1.53	[0.59; 3.99]
ATLANTIC Trial	2014	30	906	19	952	8.3%	1.66	[0.94; 2.93]
Load&Go Trial	2013	2	56	0	56	0.6%	5.00	[0.25; 101.83]
Random effects model (4)		**2363** (6)	**48962** (6)	**1019** (6)	**18158** (6)	**100.0%**	**0.73**	**[0.57; 0.93]**
Prediction interval (7)								[0.33; 1.58] (7)

(8) Heterogeneity: $Tau^2 = 0.1108$; $Chi^2 = 49.46$, df = 13 (P < 0.01); $I^2 = 74\%$

(5) Test for overall effect: Z = -2.51 (P = 0.012)

(1) All-cause Mortality

0.01 0.1 1 10 100

(3) Favors Experimental (4) Favors Control

Study	Year	Experimental Events	Total	Control Events	Total	Weight	RR	95% CI	Risk Ratio MH, Random, 95% CI
CIPAMI Trial	2012	1	164	4	171	1.2%	0.26	[0.03; 2.31]	
Dorler et al.	2011	55	1635	110	1076	11.8%	0.33	[0.24; 0.45]	
Lupi et al. (2)	2016	5	143	13	143	4.3%	0.38	[0.14; 1.05]	
Lev et al.	2008	1	165	2	127	1.0%	0.38	[0.04; 4.20]	
Danchin et al. (1)	2018	9	360	18	360	5.9%	0.50	[0.23; 1.10]	
Yudi et al.	2018	25	892	96	1915	10.1%	0.56	[0.36; 0.86]	
Alexander et al.	2008	12	1029	55	2756	7.6%	0.58	[0.31; 1.09]	
Redfors et al.	2019	1960	37840	528	6964	14.2%	0.68	[0.62; 0.75]	
Almendro-Delia et al.	2015	188	3973	116	2076	13.0%	0.85	[0.68; 1.06]	
Postma et al.	2014	46	1345	35	1096	10.1%	1.07	[0.70; 1.65]	
Fabris et al.	2022	17	237	17	300	7.3%	1.27	[0.66; 2.43]	
Fefer et al.	2009	12	217	6	166	4.6%	1.53	[0.59; 3.99]	
ATLANTIC Trial	2014	30	906	19	952	8.3%	1.66	[0.94; 2.93]	
Load&Go Trial	2013	2	56	0	56	0.6%	5.00	[0.25; 101.83]	
Random effects model		**2363**	**48962**	**1019**	**18158**	**100.0%**	**0.73**	**[0.57; 0.93]**	
Prediction interval								[0.33; 1.58]	

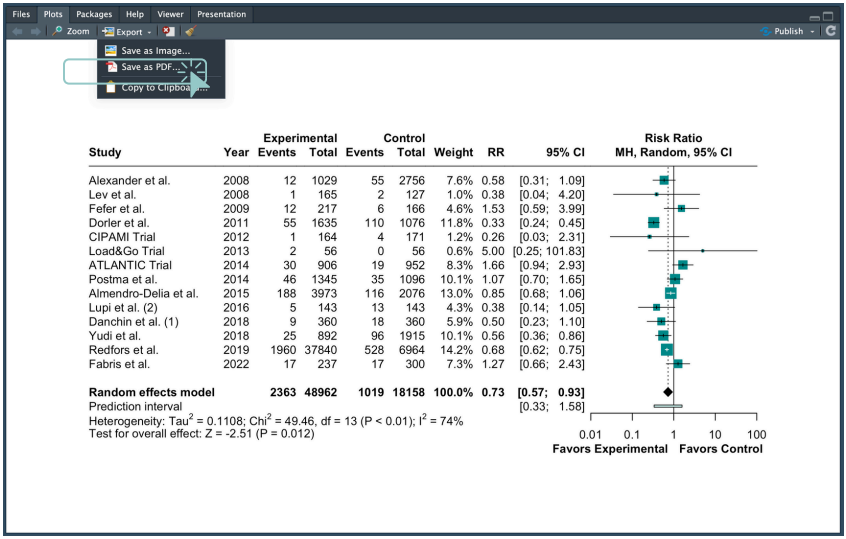
Heterogeneity: $Tau^2 = 0.1108$; $Chi^2 = 49.46$, df = 13 (P < 0.01); $I^2 = 74\%$

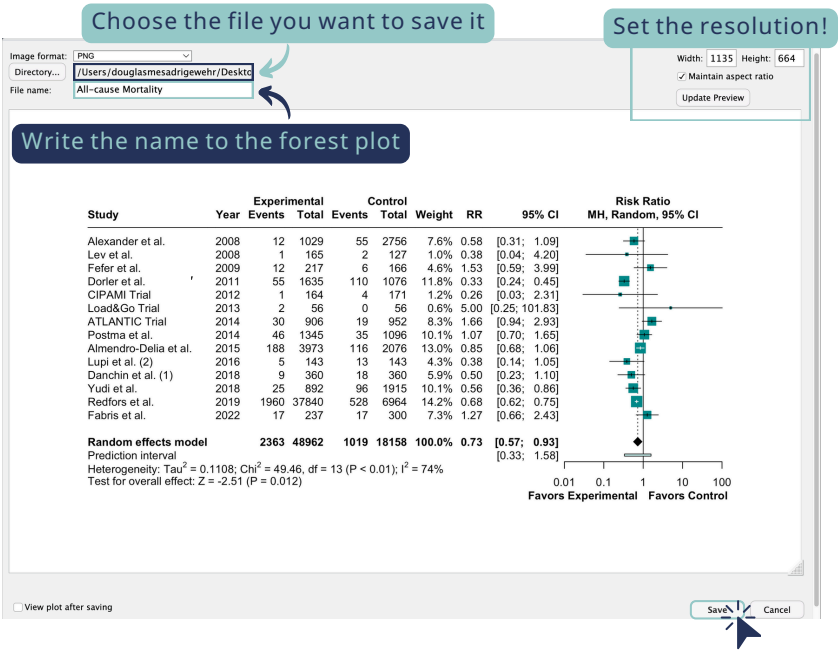
Test for overall effect: Z = -2.51 (P = 0.012)

0.01 0.1 1 10 100

Favors Experimental Favors Control

THE EASIER METHOD





HOW TO DEAL WITH

CHAPTER 1: BINARY OUTCOME DATA

RARE EVENTS

IN META-ANALYSIS

Individual studies often **lack the power** to detect differences in **rare outcomes**, but a meta-analysis of multiple studies can provide enough power to assess the impact of interventions.

However, many meta-analysis methods **rely on large-sample approximations**, which are **unsuitable** for r**are events**.

Authors should **carefully** choose methods appropriate for analyzing rare outcomes.

DEFINITION OF RARE EVENTS

There is no strict threshold to define ‘rare’ events, but risks of **1 in 1000 are certainly rare**, and risks of 1 in 100 are often considered similarly.

However, **even risks as high as 1 in 10** may be affected by the issues discussed here. A common characteristic of rare-event meta-analyses is that many studies report no events in one or more study arms.

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

WHEN USING INVERSE-VARIANCE METHOD ....

When should problematic zero counts be corrected?
When is it appropriate to apply a continuity correction?

In meta-analyses of rare events using the **Inverse Variance** method, studies with **zero events** in **one** or **both arms** can cause **computational errors** due to division by zero.

To handle this, the **metabin()** function in the meta package for R applies a **continuity correction automatically**. It adds a fixed value (**typically 0.5**) to **all cells in the 2×2 table** of **studies with zero counts**, ensuring the calculations can proceed without errors.

```
Number of studies: k = 5
Number of observations: o = 1125 (o.e = 565, o.c = 560)
Number of events: e = 41

                          RR           95%-CI     z  p-value
Common effect model  0.1493 [0.0583; 0.3823] -3.96 < 0.0001
Random effects model 0.1493 [0.0583; 0.3823] -3.96 < 0.0001

Quantifying heterogeneity:
 tau^2 = 0 [0.0000; 2.8280]; tau = 0 [0.0000; 1.6817]
 I^2 = 0.0% [0.0%; 79.2%]; H = 1.00 [1.00; 2.19]

Test of heterogeneity:
    Q d.f. p-value
 1.57    4  0.8135

Details on meta-analytical method:
- Inverse variance method
- Paule-Mandel estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Continuity correction of 0.5 in studies with zero cell frequencies
```

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

WHEN USING INVERSE-VARIANCE METHOD ....

How? It works

Original 2×2 Table (Without Continuity Correction):

Group	Deaths	No Deaths	Total
Experimental	0	50	50
Control	5	45	50

2×2 Table After Continuity Correction (Adding 0.5):

Group	Deaths	No Deaths	Total
Experimental	0.5	49.5	50
Control	5.5	44.5	50

REMINDER

This correction is **automatically** handled by the **metabin()** function.

The **only way** to perform a meta-analysis using the **Inverse Variance** method when studies have **zero counts** is by applying a **continuity correction**.

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

WHEN USING MANTEL-HAENSZEL METHOD ....

When should problematic zero counts be corrected?
When is it appropriate to apply a continuity correction?

Unlike the Inverse Variance (IV) method, the **Mantel-Haenszel** method only requires continuity corrections when **the same cell is zero across all included studies**.

Below are several examples to analyze and determine whether a continuity correction is necessary when using **MH method**.

EXAMPLE 1

MORE ON THE OTHER SIDE

Apply continuity correction

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	0	50	50
	Control	5	45	50
Study 2	Experimental	0	40	40
	Control	2	38	40
Study 3	Experimental	0	30	30
	Control	3	27	30

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

EXAMPLE 2

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	0	50	50
	Control	5	45	50
Study 2	Experimental	1	39	40
	Control	2	38	40
Study 3	Experimental	0	30	30
	Control	3	27	30

Do not apply continuity correction

WHEN USING MANTEL-HAENSZEL METHOD ....

EXAMPLE 3

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	0	50	50
	Control	0	50	50
Study 2	Experimental	1	39	40
	Control	0	40	40
Study 3	Experimental	3	27	30
	Control	0	30	30

Apply continuity correction

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

EXAMPLE 4

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	50	0	50
	Control	45	5	50
Study 2	Experimental	49	1	40
	Control	40	10	40
Study 3	Experimental	30	0	30
	Control	28	2	30

Do not apply continuity correction

WHEN USING MANTEL-HAENSZEL METHOD ....

EXAMPLE 5

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	50	0	50
	Control	45	5	50
Study 2	Experimental	49	0	40
	Control	40	10	40
Study 3	Experimental	30	0	30
	Control	28	2	30

Apply continuity correction

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

EXAMPLE 4

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	50	0	50
	Control	45	5	50
Study 2	Experimental	49	1	40
	Control	40	10	40
Study 3	Experimental	30	0	30
	Control	28	2	30

Do not apply continuity correction

WHEN USING MANTEL-HAENSZEL METHOD ....

EXAMPLE 5

Study	Group	Deaths	No Deaths	Total
Study 1	Experimental	50	0	50
	Control	45	5	50
Study 2	Experimental	49	0	40
	Control	40	10	40
Study 3	Experimental	30	0	30
	Control	28	2	30

Apply continuity correction

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

EXAMPLE 6

Do not apply continuity correction

WHEN USING MANTEL-HAENSZEL METHOD ....

Study ID	Group	Deaths	No Deaths	Total
Study 1	Experimental	0	50	50
Study 1	Control	0	50	50
Study 2	Experimental	0	50	50
Study 2	Control	0	50	50
Study 3	Experimental	3	47	50
Study 3	Control	2	48	50
Study 4	Experimental	0	50	50
Study 4	Control	0	50	50
Study 5	Experimental	0	50	50
Study 5	Control	0	50	50
Study 6	Experimental	0	50	50
Study 6	Control	0	50	50
Study 7	Experimental	0	50	50
Study 7	Control	0	50	50
Study 8	Experimental	0	50	50
Study 8	Control	0	50	50
Study 9	Experimental	0	50	50
Study 9	Control	0	50	50

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

SUMMARIZING ...

FOR INVERSE VARIANCE METHDO

Continuity correction is **always necessary** when **any cell has zero events** (ai; bi, ci, di), **no matter how many studies** report this.

FOR MANTEL-HAENSZEL METHDO

Continuity correction is needed **only if the same** (ai; bi, ci, di) **cell has zero events across all studies**; otherwise, it is not required.

Table 1 Data provided by study *i*, in the form of a 2×2 table

Study *i*	Events	No events
Experimental	a_i	b_i
Control	c_i	d_i

However, the **metabin()** function applies continuity correction to any cell with zero events, **even when it may not be strictly necessary for the MH method**, potentially **leading to excess bias**.

```
Details on meta-analytical method:
- Mantel-Haenszel method
- DerSimonian-Laird estimator for tau^2
- Mantel-Haenszel estimator used in calculation of Q and tau^2 (like RevMan 5)
- Continuity correction of 0.5 in studies with zero cell frequencies
```

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

DISABLING CONTINUITY CORRECTION IN MANTEL-HAENSZEL ANALYSES

When there is no true need for continuity correction, you can prevent the **metabin()** function from applying it by including the argument '**MH.exact = TRUE**'. This ensures that the **Mantel-Haenszel** calculations are performed without introducing unnecessary bias from artificial adjustments.

```
m.acm<-metabin(event.e, n.e, event.c, n.c,
               data = ma$ACM,
               method = "Inverse",
               method.tau = "PM",
               sm = "RR",
               MH.exact = TRUE,
               studlab = Author)
```

Cochrane RevMan

HOWEVER, THIS FEATURE IS NOT AVAILABLE IN REVMAN.

MORE ON THE OTHER SIDE

HOW TO DEAL WITH RARE EVENTS IN META-ANALYSIS

ANOTHER OPTION IS TO PERFORM A META-ANALYSIS USING THE PETO METHOD, WHICH ALMOST NEVER REQUIRES CONTINUITY CORRECTION.

HOWEVER .....

Before applying, you must **make sure**:

- Can **ONLY** be used to estimate **OR**;
- Estimate OR only by **fixed-effects**;
- Treatment groups **NEED** to be **BALANCED**;
- The effects **CAN'T** be very large;

**Exclude double-zero studies

If these conditions do not hold

IX- Reporting the Methods

It is highly advised to specify as the aspects of the model you used in the methods section of your meta-analysis report, including:

(1) the summary measure;
(2) the method to pool the studies;
(3) the between-study variance estimator;
(4) and other details you applied in your analyses.

EXAMPLE

Statistical Analysis

We summarized binary endpoints using the Mantel-Haenszel random-effects model, with risk ratio (RR) and 95% confidence interval (CI) as a measure of effect size. The Paule-Mandel estimator was used to calculate the between-study variance t^2.[1] We assessed heterogeneity with Cochrane's Q statistic and Higgins and Thompson's I^2 statistic, with $p \leq 0.10$ indicating statistical significance. We determined the consistency of the studies based on I^2 values of 0%, ≤ 25%, ≤ 50%, and > 50%, indicating no observed, low, moderate, and substantial heterogeneity, respectively.[2] All tests were two-tailed, and a p-value of < 0.05 was considered statistically significant. We used R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) and the extension package "meta" for all calculations and graphics.[3,4]

References

1. Paule RC, Mandel J. Consensus Values and Weighting Factors. *Journal of Research of the National Bureau of Standards*.1982;87(5):377-385.
2. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine.* 2002;21(11):1539-1558.
3. Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
4. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health*. 22nd ed.; 2019:153-160.

X – Bibliography

Between-study variance estimators

Langan D, Higgins JP, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*. 2005;10(1):83–98.

Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*.2016;7(1):55–79.

Paule-Mandel estimator:
Paule RC, Mandel J. Consensus Values and Weighting Factors. *Journal of Research of the National Bureau of Standards*.1982;87(5):377-385.

Restricted maximum likelihood estimator:
Viechtbauer, W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*. 2005;30(3):261-293.

Sidik-Jonkman estimator:
Sidik K, Jonkman JN. Simple Heterogeneity Variance Estimation for Meta-Analysis. *Journal of the Royal Statistical Society Series C: Applied Statistics. 2005;*54(2):367–384.

DerSimonian-Laird estimator:
DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177–188.

Heterogeneity Assessment

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 2002;21(11):1539-1558.

Meta-analysis of rare events

Efthimiou O. Practical guide to the meta-analysis of rare events. Evid Based Ment Health. 2018 May;21(2):72-76.

R, RStudio, and R packages

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

meta package:
Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health*. 22nd ed.; 2019:153-160.

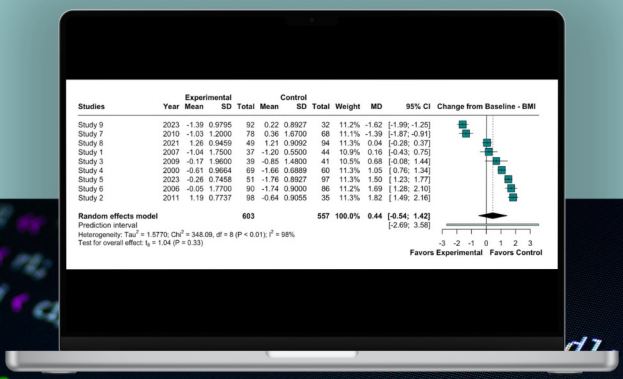
readxl package:
Wickham H, Bryan J (2023). readxl: Read Excel Files. https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.

MASTERING

PAIRWISE META-ANALYSIS

CHAPTER 2: CONTINUOUS OUTCOME DATA

in RStudio

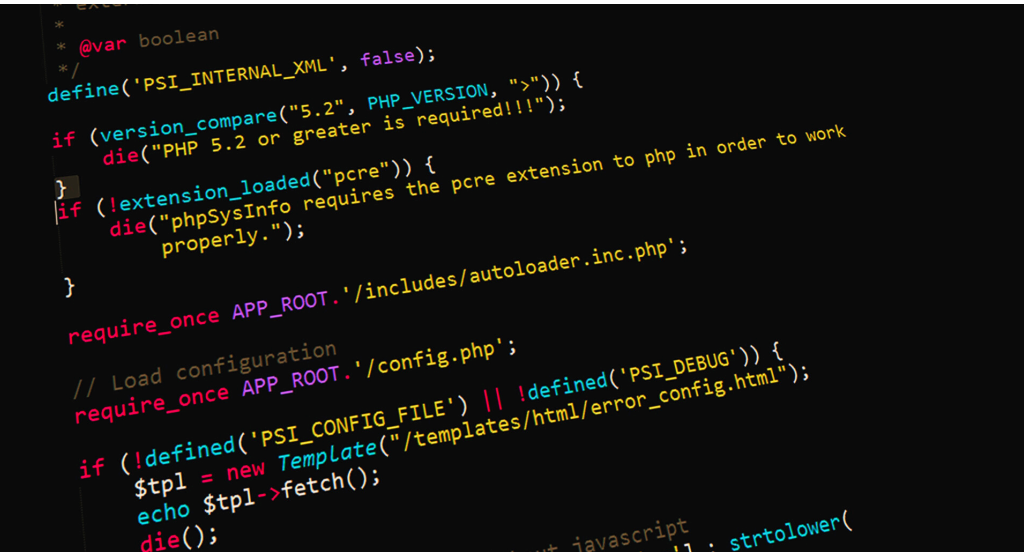


AN INTUITIVE AND STRAIGHTFORWARD GUIDE

DOUGLAS MESADRI GEWEHR

EMILTON LIMA JUNIOR

SUMMARY



I – Essential R Packages

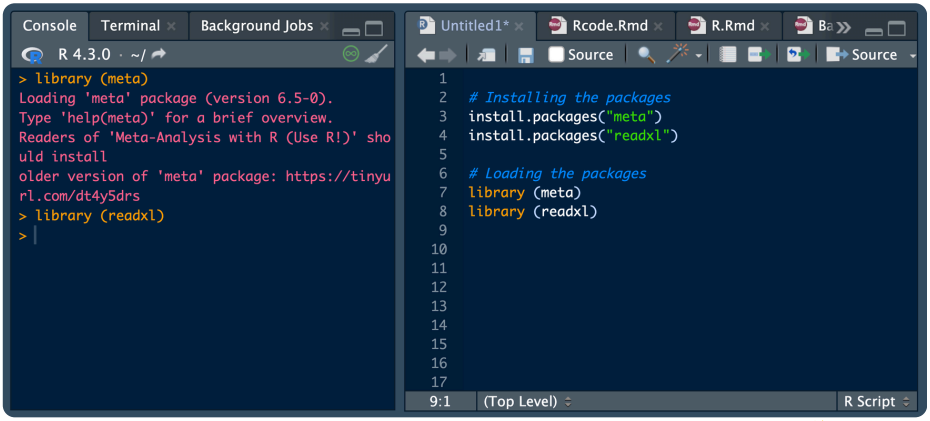
readxl package

A package that provides functions to read data from Excel files into R. It supports both ".xls" and ".xlsx" file formats, and can read data from individual worksheets or the entire workbook.

meta package

This package is a comprehensive package for meta-analysis, including functions for estimating and plotting effects, performing tests for heterogeneity, adjusting meta-regression models, and conducting subgroup analyses.

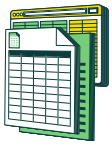
First, you'll need to **install** and **load** the packages as follows:



Once you install the packages with the **install.packages()** function, you need to load ("activate") the packages, using the **library()** function every time you start RStudio!

QUICK TIPS

II – Preparing your Dataset



Before you begin your meta-analysis, you need to ensure that your data has been organized and structured correctly!

This will help ensure that the data is consistent and compatible with the functions and methods used in the meta-analysis, and can save time and effort in the analysis process.

Here are some **Dos and Don'ts** when preparing your data in Excel:

1. It's very important how you name the columns of your spreadsheet. If you already named the columns of your sheet appropriately in **Excel**, you can save a lot of time later because your data does not have to be transformed using R.

2. Column names should not contain any spaces. An alternative is to use underline sign "_".

3. It does not matter how columns are ordered in your Excel spreadsheet. They just have to be labeled correctly.

4. It is also important to know that the import may distort special characters like ä, ü, ö, and so forth. You might want to transform them into "normal" letters before you proceed.

5. If you have one or several empty rows or columns which used to contain data, make sure to delete those columns/rows completely because R could think that these columns contain (missing) data and import them also.

II – Preparing your Dataset

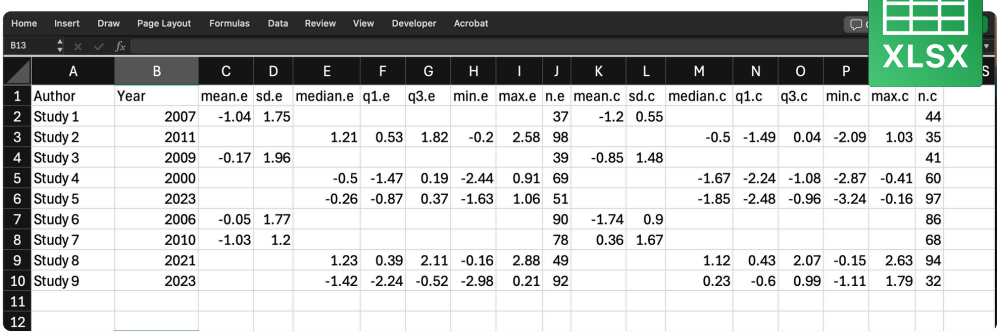
Structure your dataset following the example below:

Mean - experimental group
Standard Deviation - experimental group
Median - experimental group
Quartile 1 and 3 - experimental group
Minimum and Maximum - experimental group
Sample Size - experimental group
Study Identification

Author	Year	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c
Study 1	2007	-1.04	1.75						37	-1.2	0.55						44
Study 2	2011			1.21	0.53	1.82	-0.2	2.58	98			-0.5	-1.49	0.04	-2.09	1.03	35
Study 3	2009	-0.17	1.96						39	-0.85	1.48						41
Study 4	2000			-0.5	-1.47	0.19	-2.44	0.91	69			-1.67	-2.24	-1.08	-2.87	-0.41	60
Study 5	2023			-0.26	-0.87	0.37	-1.63	1.06	51			-1.85	-2.48	-0.96	-3.24	-0.16	97
Study 6	2006	-0.05	1.77						90	-1.74	0.9						86
Study 7	2010	-1.03	1.2						78	0.36	1.67						68
Study 8	2021			1.23	0.39	2.11	-0.16	2.88	49			1.12	0.43	2.07	-0.15	2.63	94
Study 9	2023			-1.42	-2.24	-0.52	-2.98	0.21	92			0.23	-0.6	0.99	-1.11	1.79	32

Mean - control group
Standard Deviation - control group
Median - control group
Quartile 1 and 3 - control group
Minimum and Maximum - control group
sample Size - control group

XLSX



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Author	Year	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c
2	Study 1	2007	-1.04	1.75						37	-1.2	0.55						44
3	Study 2	2011			1.21	0.53	1.82	-0.2	2.58	98			-0.5	-1.49	0.04	-2.09	1.03	35
4	Study 3	2009	-0.17	1.96						39	-0.85	1.48						41
5	Study 4	2000			-0.5	-1.47	0.19	-2.44	0.91	69			-1.67	-2.24	-1.08	-2.87	-0.41	60
6	Study 5	2023			-0.26	-0.87	0.37	-1.63	1.06	51			-1.85	-2.48	-0.96	-3.24	-0.16	97
7	Study 6	2006	-0.05	1.77						90	-1.74	0.9						86
8	Study 7	2010	-1.03	1.2						78	0.36	1.67						68
9	Study 8	2021			1.23	0.39	2.11	-0.16	2.88	49			1.12	0.43	2.07	-0.15	2.63	94
10	Study 9	2023			-1.42	-2.24	-0.52	-2.98	0.21	92			0.23	-0.6	0.99	-1.11	1.79	32
11																		
12																		

Here's how your Excel spreadsheet should look

II – Preparing your Dataset

Why Include These Extra Columns?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Author	Year	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c	
2	Study 1	2007	-1.04	1.75						37	-1.2	0.55						44	
3	Study 2	2011			1.21	0.53	1.82	-0.2	2.58	98			-0.5	-1.49	0.04	-2.09	1.03	35	
4	Study 3	2009	-0.17	1.96						39	-0.85	1.48						41	
5	Study 4	2000			-0.5	-1.47	0.19	-2.44	0.91	69			-1.67	-2.24	-1.08	-2.87	-0.41	60	
6	Study 5	2023			-0.26	-0.87	0.37	-1.63	1.06	51			-1.85	-2.48	-0.96	-3.24	-0.16	97	
7	Study 6	2006	-0.05	1.77						90	-1.74	0.9						86	
8	Study 7	2010	-1.03	1.2						78	0.36	1.67						68	
9	Study 8	2021			1.23	0.39	2.11	-0.16	2.88	49			1.12	0.43	2.07	-0.15	2.63	94	
10	Study 9	2023			-1.42	-2.24	-0.52	-2.98	0.21	92			0.23	-0.6	0.99	-1.11	1.79	32	
11																			
12																			

For a meta-analysis of continuous outcomes using **Mean Difference**, you typically need:

Mean

Measure of central tendency

Standard Deviation

Measure of dispersion

However, many studies report **alternative measures**, particularly when dealing with skewed data :

Median

Measure of central tendency

Q1 **Q3** **Min** **Max**

Measures of dispersion

By organizing your dataset to include columns for **Median**, **Q1**, **Q3**, **Min**, and **Max**, you can easily estimate missing values when needed. This ensures flexibility without losing valuable studies.

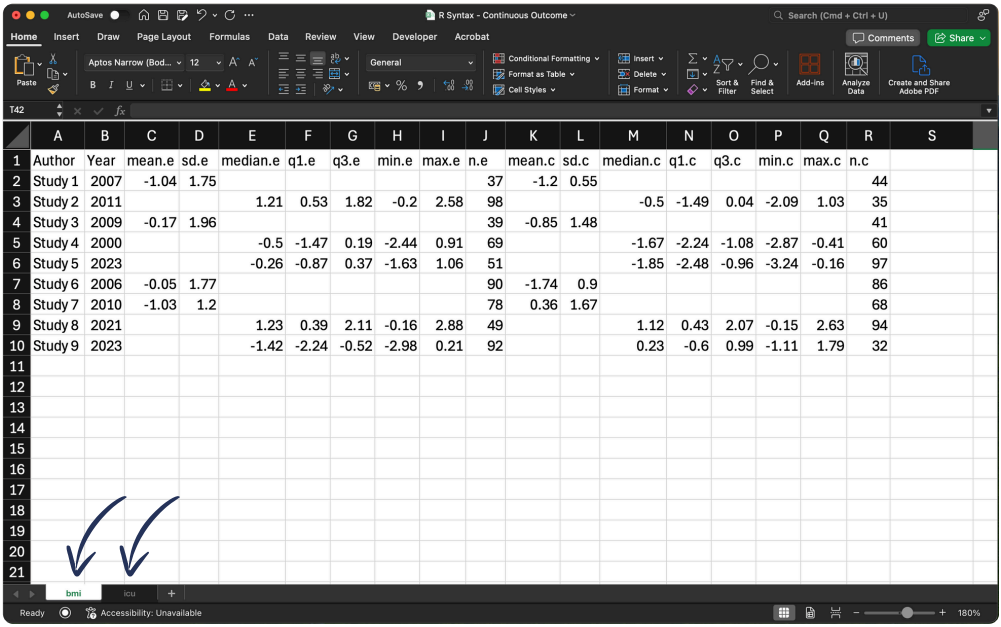
You need at least **one pair** (Q1-Q3 or Min-Max) of dispersion measures. **The more** measures you have, **the more accurate** the estimation will be.

II – Preparing your Dataset

Create a spreadsheet for each outcome of your meta-analysis in separate worksheets in your Excel workbook!

This approach saves time by allowing you to import **ALL** your endpoints at once into RStudio.

Here is an example of a dataset with 2 continuous endpoints, each one in its own worksheet: BMI (bmi) and ICU length of stay (icu)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Author	Year	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c	
2	Study 1	2007	-1.04	1.75						37	-1.2	0.55						44	
3	Study 2	2011			1.21	0.53	1.82	-0.2	2.58	98			-0.5	-1.49	0.04	-2.09	1.03	35	
4	Study 3	2009	-0.17	1.96						39	-0.85	1.48						41	
5	Study 4	2000			-0.5	-1.47	0.19	-2.44	0.91	69			-1.67	-2.24	-1.08	-2.87	-0.41	60	
6	Study 5	2023			-0.26	-0.87	0.37	-1.63	1.06	51			-1.85	-2.48	-0.96	-3.24	-0.16	97	
7	Study 6	2006	-0.05	1.77						90	-1.74	0.9						86	
8	Study 7	2010	-1.03	1.2						78	0.36	1.67						68	
9	Study 8	2021			1.23	0.39	2.11	-0.16	2.88	49			1.12	0.43	2.07	-0.15	2.63	94	
10	Study 9	2023			-1.42	-2.24	-0.52	-2.98	0.21	92			0.23	-0.6	0.99	-1.11	1.79	32	
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			

III – Importing your Dataset

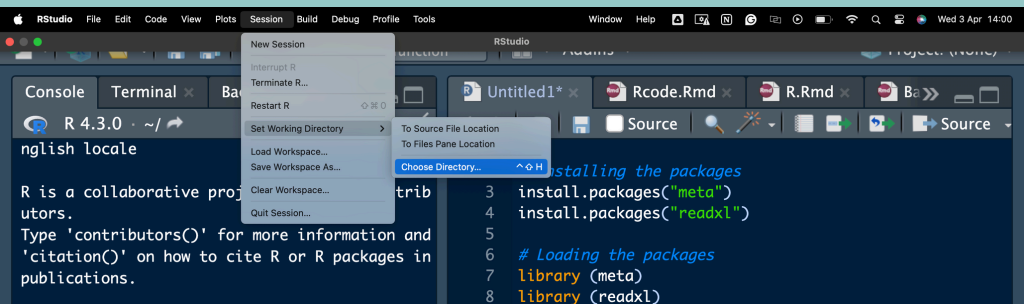
1ST

Once you have your Excel workbook with all endpoints of your meta-analysis in its own worksheets ready, save the **.xlsx** file in a designated folder, commonly referred to as the '**working directory**' (WD), or any other preferred name.

For that, you need to use the **setwd()** function and add between quotes the directory of the folder WD in your computer that you've created:

```
#Setting your WD folder in R Studio:
setwd("/Users/douglasmesadrigewehr/Desktop/WD RStudio")
```

The other way is to manually configure the WD folder, as shown below:



2

The following R code creates a new object with the name of your choice (in this example I am using '**ma**', containing all the outcomes in a data frame object.

```
sheet_names <- excel_sheets("data.xlsx")
sheet_names

ma <- lapply(sheet_names, function(x) {
  as.data.frame(read_excel("data.xlsx", sheet = x)) } )

names(ma) <- sheet_names
```

ATTENTION PLEASE!

When using this code, you just need to rename the **.xlsx** file (arrows) that contains your dataset, located in your WD folder.

IV – Visualizing your Data

Now, we can have a look at each of the lists (our endpoints) of our data frame object.

The object **ma,** we've just imported, has **2** lists of data:

1. "**bmi**" for Body Mass Index data
2. "**icu**" for Intensive Care Unit Length of Stay data

To print one specific endpoint, you can do so by adding the dollar sign operator "**$**" after "**ma**" inside de **View()** function:

```
View(ma$bmi) #to visualize Body Mass Index data
View(ma$icu) #to visualize Intensive Care Unit Length of Stay data
```

This is how your endpoint dataset will appear in RStudio.
See the example below:

Untitled1* | ma$bmi | R Script Statistical Code .R

Filter

	Author	Year	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c
1	Study 1	2007	-1.04	1.75	NA	NA	NA	NA	NA	37	-1.20	0.55	NA	NA	NA	NA	NA	44
2	Study 2	2011	NA	NA	1.21	0.53	1.82	-0.20	2.58	98	NA	NA	-0.50	-1.49	0.04	-2.09	1.03	35
3	Study 3	2009	-0.17	1.96	NA	NA	NA	NA	NA	39	-0.85	1.48	NA	NA	NA	NA	NA	41
4	Study 4	2000	NA	NA	-0.50	-1.47	0.19	-2.44	0.91	69	NA	NA	-1.67	-2.24	-1.08	-2.87	-0.41	60
5	Study 5	2023	NA	NA	-0.26	-0.87	0.37	-1.63	1.06	51	NA	NA	-1.85	-2.48	-0.96	-3.24	-0.16	97
6	Study 6	2006	-0.05	1.77	NA	NA	NA	NA	NA	90	-1.74	0.90	NA	NA	NA	NA	NA	86
7	Study 7	2010	-1.03	1.20	NA	NA	NA	NA	NA	78	0.36	1.67	NA	NA	NA	NA	NA	68
8	Study 8	2021	NA	NA	1.23	0.39	2.11	-0.16	2.88	49	NA	NA	1.12	0.43	2.07	-0.15	2.63	94
9	Study 9	2023	NA	NA	-1.42	-2.24	-0.52	-2.98	0.21	92	NA	NA	0.23	-0.60	0.99	-1.11	1.79	32

Showing 1 to 9 of 9 entries, 18 total columns

V – Performing the Meta-Analysis

To perform the meta-analysis of binary outcome data we will use the **metacont()** function from the '**meta**' package

We need to provide some **instructions** for the **metacont()** function. These instructions are known as **arguments**, which can be one, two, three, or more inputs that the function uses to perform its task. The main arguments of the metacont() function are:

Scenario 1

Mean and SD available

Assigns the output of the "**metacont()**" function to a new variable named "**m.bmi**"

Column name in dataset containing the **Mean** in the **intervention** ('sd.e') and **control** ('sd.c') groups

SD in the **intervention** ('sd.e') and **control** ('sd.c') groups

Sample size in the **intervention** ('n.e') and **control** ('n.c') groups

```
m.bmi<-metacont(mean.e=mean.e, sd.e=sd.e, n.e=n.e,
                mean.c=mean.c, sd.c=sd.c, n.c=n.c,
                data = ma$bmi,
                method.tau = "REML",
                sm = "MD",
                studlab = Author)
```

Data frame object you've created containing the study and outcome data (using the **$** operator)

Method to be used to estimate the **between-study variance tau²**

Summary measure to be use

Column name in dataset containing the **study labels**

Scenario 2

Some Studies with Missing Mean and SD

Median ('median.e'), **first quartile** ('q1.e'), **third quartile** ('q3.e'), **minimum value** ('min.e'), and **maximum value** ('max.e') in the **intervention** group

```
m.bmi<-metacont(mean.e=mean.e, sd.e=sd.e, n.e=n.e,
                median.e=median.e, q1.e=q1.e, q3.e=q3.e,
                min.e=min.e, max.e=max.e,
                mean.c=mean.c, sd.c=sd.c, n.c=n.c,
                median.c=median.c, q1.c=q1.c, q3.c=q3.c,
                min.c=min.c, max.c=max.c,
                data = ma$bmi,
                method.tau = "REML",
                sm = "MD",
                studlab = Author)
```

QUICK TIPS

Always separate your arguments with a **comma**

Median ('median.c'), **first quartile** ('q1.c'), **third quartile** ('q3.c'), **minimum value** ('min.c'), and **maximum value** ('max.c') in the **control** group

To handle missing data, we use statistical methods to estimate **Mean** and **SD** based on other reported measures. Include the additional aforementioned arguments in the **metacont()** function.

V – Performing the Meta-Analysis

Over the next pages, we will detail the following aspects of continuous outcome meta-analysis:
(1) methods for **pooling the studies**;
(2) methods to estimate the **between-study variance tau²**; and
(3) the **summary measures**.

To visualize the results of a meta-analysis conducted with the **metacont()** function, we will use the **summary()** function. It will generate a summary of the analysis

```
summary(m.bmi)
```

Call here the **metacont()** object you have just created

The **summary()** function will display the estimated treatment effect (i.e., the summary measure), the corresponding confidence interval, the between-study variance, and other relevant statistics. See the example below:

Console | Terminal | Background Jobs

R 4.3.0 · ~/Desktop/WD RStudio/

This section contains:
(1) the individual studies;
(2) their effect sizes and confidence intervals;
(3) their weights (%) in common (i.e. fixed-) and random-effects

```
> summary(m.bmi)
                      MD             95%-CI %W(common) %W(random)
Study 1  0.1600 [-0.4268;  0.7468]        4.4       10.9
Study 2  1.8248 [ 1.4879;  2.1616]       13.3       11.3
Study 3  0.6800 [-0.0840;  1.4440]        2.6       10.5
Study 4  1.0505 [ 0.7635;  1.3375]       18.3       11.3
Study 5  1.5002 [ 1.2292;  1.7712]       20.5       11.3
Study 6  1.6900 [ 1.2778;  2.1022]        8.9       11.2
Study 7 -1.3900 [-1.8680; -0.9120]        6.6       11.1
Study 8  0.0447 [-0.2777;  0.3670]       14.5       11.3
Study 9 -1.6167 [-1.9851; -1.2483]       11.1       11.2

Number of studies: k = 9
Number of observations: o = 1160 (o.e = 603, o.c = 557)

                         MD            95%-CI     z  p-value
Common effect model  0.6513 [ 0.5286; 0.7740] 10.40 < 0.0001
Random effects model 0.4417 [-0.3921; 1.2756]  1.04   0.2991

Quantifying heterogeneity:
 tau^2 = 1.5770 [0.6922; 5.8535]; tau = 1.2558 [0.8320; 2.4194]
 I^2 = 97.7% [96.8%; 98.3%]; H = 6.60 [5.60; 7.77]

Test of heterogeneity:
     Q d.f.  p-value
 348.09    8 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
```

The **total number of studies** in our meta-analysis

The **total number of patients** in our meta-analysis

The POOLED effect size

Results concerning the **between-study heterogeneity**

Method to be used **for pooling the studies**

Method to be used to estimate the **between-study variance tau²**

V – Performing the meta-analysis

WHICH SUMMARY MEASURE SHOULD I USE?

Mean Difference ("MD")

When to use MD?

- **Use MD** if all studies measure the outcome on the **same scale** and in the same units (e.g., all measure blood pressure in mmHg).

How to interpret?

- MD preserves the original scale, allowing you to interpret the effect size in familiar units.

Standardized Mean Difference ("SMD")

When to use SMD?

- If studies measure the same underlying concept but **use different units or scales** (e.g., one study measures depression with a 20-point scale, another with a 50-point scale), **use SMD**.

How to interpret?

- The SMD is expressed in standard deviation units rather than the original measurement units. Interpret the size of the effect by how many standard deviations separate the groups.
- This approach assumes that differences in SDs across studies are due to variations in measurement scales, not differences in outcome reliability or population variability.

V – Performing the meta-analysis

METHODS FOR POOLING THE EFFECT SIZES

Which method should I use?

The **inverse-variance** method is the standard approach for pooling **continuous** outcomes in meta-analysis.

This method is recommended by the **Cochrane Collaboration** and is the **default** in the **metacont()** function.

You **do not need** to **specify** this method explicitly when using **metacont()**, as it is automatically applied.

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

Restricted Maximum Likelihood ("REML")

VS

DerSimonian-Laird ("DL") estimator can be **biased**, particularly when the number of studies is **small** and **heterogeneity** is **high.**

This can lead to **overconfident** results with **narrower** confidence intervals than warranted.

False-positive results
Type 1 error

V – Performing the meta-analysis

OVERVIEW

Methods for pooling the effect sizes
Methods to estimate the between-study variance

LEARNING R SYNTAX

Methods to estimate the between-study variance

```
method.tau = "REML" # Restricted maximum-likelihood estimator
method.tau = "DL" # DerSimonian-Laird estimator
```

Always use " " when referring to these methods

QUICK TIPS

RATIONALE
BEHIND
MEAN AND SD ESTIMATION

When dealing with skewed data, it is crucial to use methods validated by authoritative sources like the **Cochrane** Handbook for Systematic Reviews of Interventions. This ensures accurate and reliable estimation of missing means and standard deviations (SDs).

The **metacont()** function **automatically** applies the **Wan and Luo method**, recommended by **Cochrane**, to estimate means and SDs when medians, interquartile ranges (IQR), and/or, less frequently, ranges are provided. This method is **particularly accurate for skewed data**, enabling the inclusion of more studies in your meta-analysis.

V – Performing the meta-analysis

LERN HOW TO

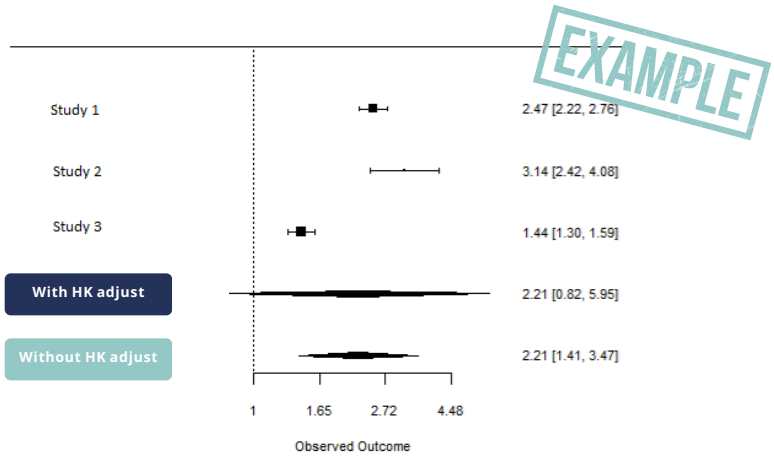
TO APPLY THE

KNAPP-HARTUNG

ADJUSTMENTS

```
m.bmi<-metacont(mean.e=mean.e, sd.e=sd.e, n.e=n.e,
                mean.c=mean.c, sd.c=sd.c, n.c=n.c,
                data = ma$bmi,
                method.tau = "REML",
                sm = "MD",
                hakn = TRUE,
                studlab = Author)
```

To apply the **Knapp-Hartung** adjustments when working with random effects model, you just need to add the argument highlighted in **red** in the **metacont()** function

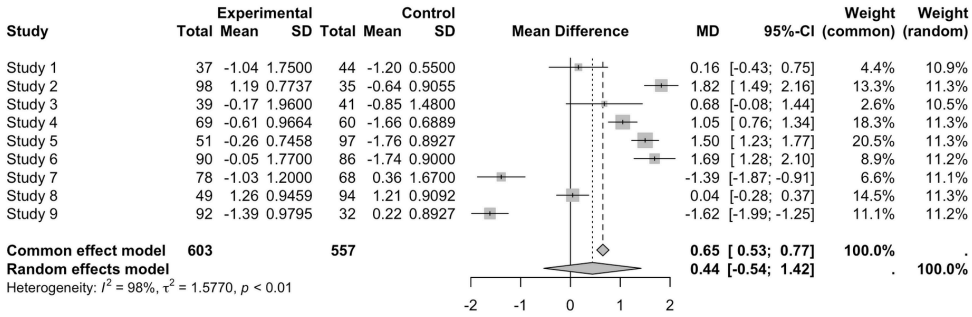


VI – Creating the Forest Plot

You can produce a forest plot for **metacont()** object using the **forest()** function from the '**meta**' package

Usually, these forest plots do not have enough quality for publication by default! In the example below, we plot the '**m.bmi**' object that we also used in the previous example:

```
#Creating a forest plot for 'm.bmi' metacont object
forest(m.bmi)
```



However, the **forest()** function also has countless additional arguments to further customize the appearance of the forest plot.

All these arguments are detailed and described in the function documentation. You can accesses this documentation by running the following code:

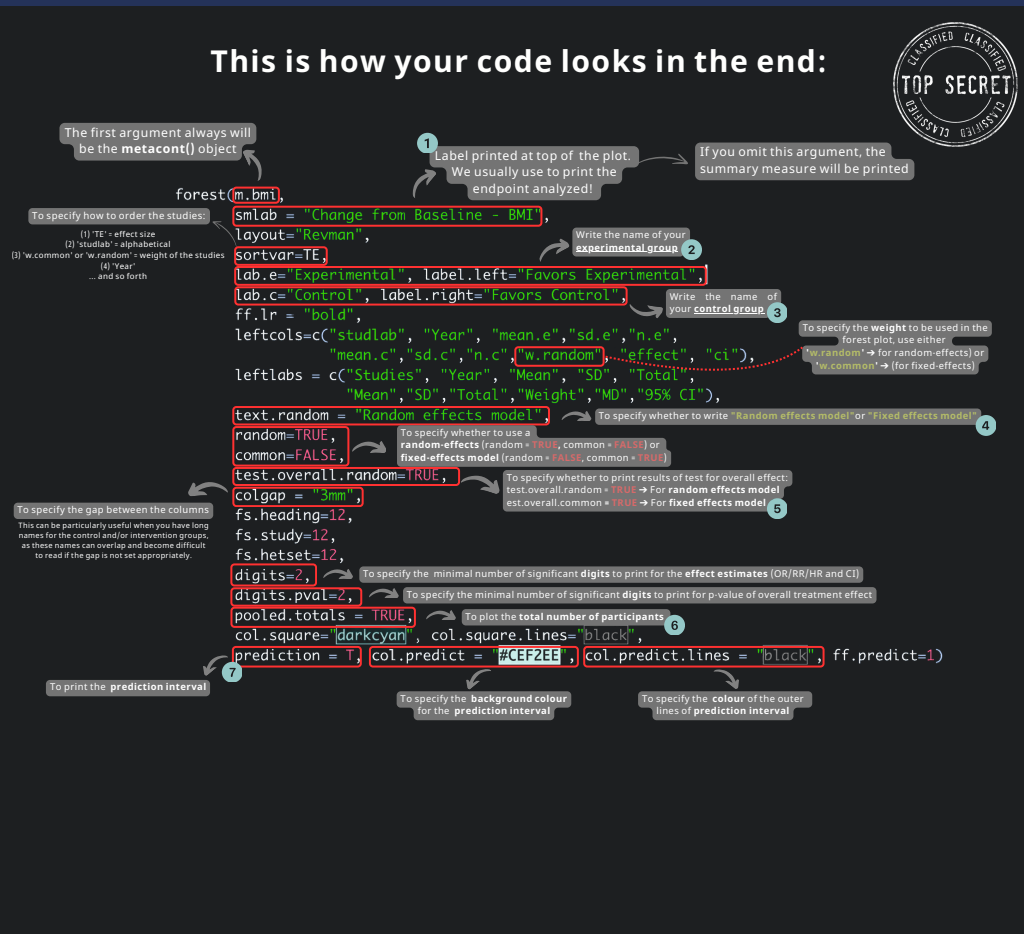
```
help(forest)
```

VI – Creating the Forest Plot

But don't worry. I've put together a code that brings together the main arguments for generating **elegant** and **high-quality** forest plots for publication.

Now, let's understand the most important arguments to customize your forest plot!

This is how your code looks in the end:



MORE ON THE OTHER SIDE

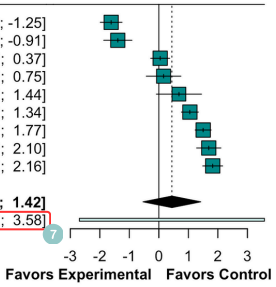
VI – Creating the Forest Plot

This is how our forest plot looks in the end:

Studies	Year	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	MD	95% CI
Study 9	2023	-1.39	0.9795	92	0.22	0.8927	32	11.2%	-1.62	[-1.99; -1.25]
Study 7	2010	-1.03	1.2000	78	0.36	1.6700	68	11.1%	-1.39	[-1.87; -0.91]
Study 8	2021	1.26	0.9459	49	1.21	0.9092	94	11.3%	0.04	[-0.28; 0.37]
Study 1	2007	-1.04	1.7500	37	-1.20	0.5500	44	10.9%	0.16	[-0.43; 0.75]
Study 3	2009	-0.17	1.9600	39	-0.85	1.4800	41	10.5%	0.68	[-0.08; 1.44]
Study 4	2000	-0.61	0.9664	69	-1.66	0.6889	60	11.3%	1.05	[0.76; 1.34]
Study 5	2023	-0.26	0.7458	51	-1.76	0.8927	97	11.3%	1.50	[1.23; 1.77]
Study 6	2006	-0.05	1.7700	90	-1.74	0.9000	86	11.2%	1.69	[1.28; 2.10]
Study 2	2011	1.19	0.7737	98	-0.64	0.9055	35	11.3%	1.82	[1.49; 2.16]
Random effects model				**603**			**557**	**100.0%**	**0.44**	**[-0.54; 1.42]**
Prediction interval										[-2.69; 3.58]

Heterogeneity: $Tau^2 = 1.5770$; $Chi^2 = 348.09$, df = 8 (P < 0.01); $I^2 = 98\%$

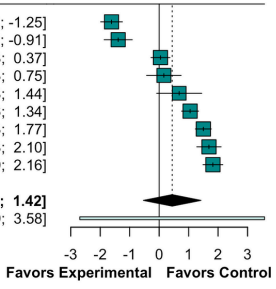
Test for overall effect: $t_8 = 1.04$ (P = 0.33)



Studies	Year	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	MD	95% CI
Study 9	2023	-1.39	0.9795	92	0.22	0.8927	32	11.2%	-1.62	[-1.99; -1.25]
Study 7	2010	-1.03	1.2000	78	0.36	1.6700	68	11.1%	-1.39	[-1.87; -0.91]
Study 8	2021	1.26	0.9459	49	1.21	0.9092	94	11.3%	0.04	[-0.28; 0.37]
Study 1	2007	-1.04	1.7500	37	-1.20	0.5500	44	10.9%	0.16	[-0.43; 0.75]
Study 3	2009	-0.17	1.9600	39	-0.85	1.4800	41	10.5%	0.68	[-0.08; 1.44]
Study 4	2000	-0.61	0.9664	69	-1.66	0.6889	60	11.3%	1.05	[0.76; 1.34]
Study 5	2023	-0.26	0.7458	51	-1.76	0.8927	97	11.3%	1.50	[1.23; 1.77]
Study 6	2006	-0.05	1.7700	90	-1.74	0.9000	86	11.2%	1.69	[1.28; 2.10]
Study 2	2011	1.19	0.7737	98	-0.64	0.9055	35	11.3%	1.82	[1.49; 2.16]
Random effects model				**603**			**557**	**100.0%**	**0.44**	**[-0.54; 1.42]**
Prediction interval										[-2.69; 3.58]

Heterogeneity: $Tau^2 = 1.5770$; $Chi^2 = 348.09$, df = 8 (P < 0.01); $I^2 = 98\%$

Test for overall effect: $t_8 = 1.04$ (P = 0.33)



VI – Creating the Forest Plot

DON'T FORGET

TO **ADJUST** THE **LABELS** IN YOUR FOREST PLOT

It's critical to correctly adjust the "**Favors Experimental**" and "**Favors Control**" labels in forest plots. The interpretation of these labels depends on whether an **increase** or **decrease** in the (standardized) **mean difference** favors the experimental group. **Mislabeling** can mislead readers and **compromise the clarity** of your analysis.

SCENARIO 1 HIGHER SCORE OR VALUE = BETTER OUTCOME

When a **higher score or value** represents a **better** outcome, an **increase** in the **MD** favors the **experimental group**.

EXAMPLE

- **Quality of Life:** Higher scores indicate improved well-being.
- **Cognitive Function Tests:** Higher scores represent better performance.
- **Physical Activity Levels:** Higher values mean better results.

Always **check** the **clinical meaning** of an **increase** or **decrease** in the outcome.

SCENARIO 2 LOWER SCORE OR VALUE = BETTER OUTCOME

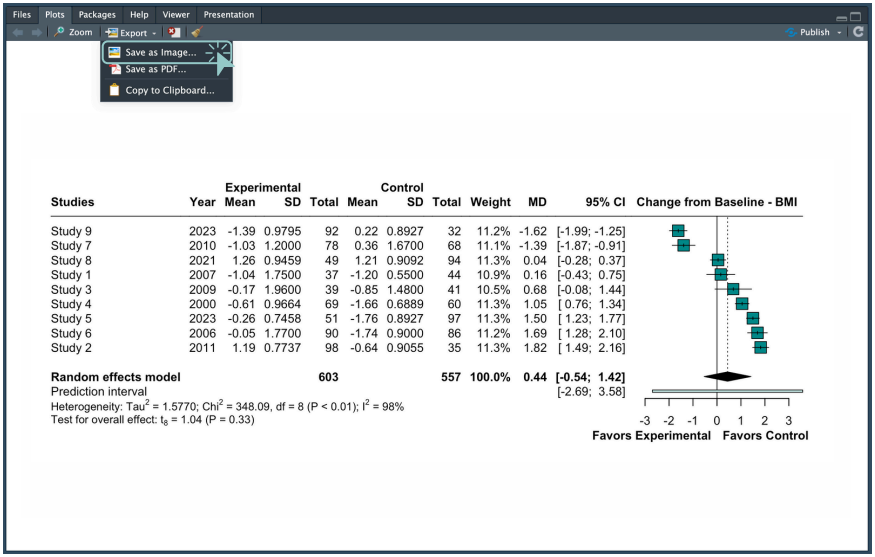
When a **lower score or value** represents a **better** outcome, a **decrease** in the **MD** favors the **experimental group**.

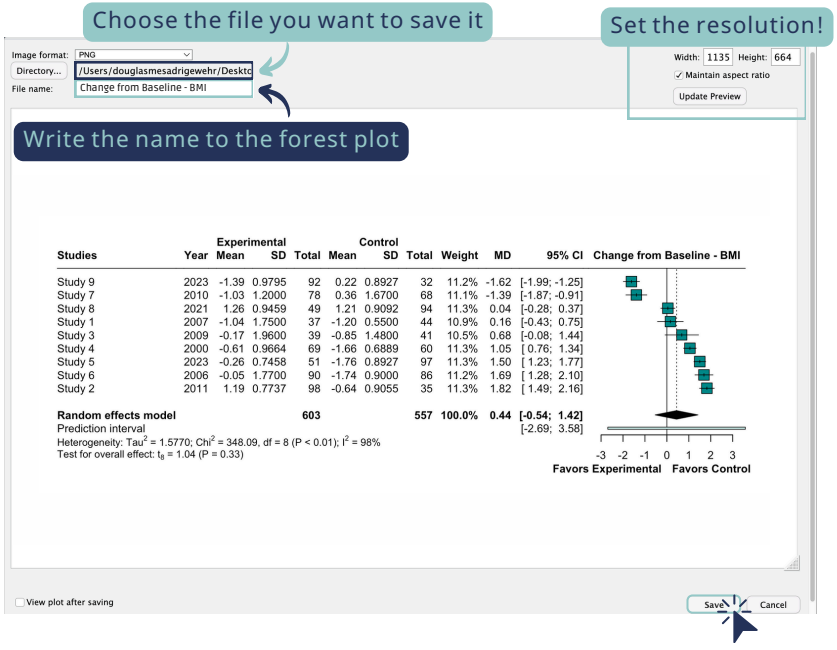
EXAMPLE

- **Blood Pressure:** Lower values reduce cardiovascular risk.
- **Cholesterol Levels:** Lower levels are beneficial.
- **Body Mass Index:** Lower BMI values indicate successful weight-loss interventions.

Always **check** the **clinical meaning** of an **increase** or **decrease** in the outcome.

THE EASIER METHOD





ESSTENTIAL PRECAUTIONS

AVOIDING

COMMON PITFALLS

ALIGNING SCALE DIRECTIONS

- When using the SMD method, it's crucial to ensure that **all scales point are in the same direction**.
- This is important because the SMD method **does not** automatically **correct** for **differences in the direction** of scales.
- If scales measure outcomes in opposite directions, results may be **misleading**.

EXAMPLE

- **Study 1:** Uses the **Beck Depression Inventory (BDI)**, where higher scores indicate greater severity of depression.
- **Study 2:** Uses a different scale where higher scores indicate less severe depression (e.g., **mental health quality of life scales**).

HOW TO ADJUST FOR CONSISTENCY?

1. **First,** identify scales where higher scores indicate less severity.
2. Adjust these scales so higher scores indicate greater severity by

BEST CHOICE

OPTION 1

Multiply the **mean** by **-1**

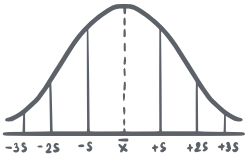
Example: If the mean score is 10, the adjusted mean becomes -10.

OPTION 2

Subtract the **mean** score from the **maximum** possible **score** of the scale.

Example: If the mean score is 10 and the maximum score is 20, the adjusted mean becomes 20 - 10 = 10.

WARNING **You do not need to adjust the standard deviation (SD).**



DEALING WITH MISSING STANDARD DEVIATIONS

MISSING Standard deviations can be obtained from:

Standard Error or **Confidence Intervals**

or or

t-statistics or **P-values**

RevMan Calculator becomes even more powerful with its integrated tools. For example, it includes a link to an online calculator developed by the **Cochrane Collaboration**, designed to estimate missing standard deviations using alternative measures reported in individual studies.

Scan the QR code below to **download** the **Excel calculator** and streamline your meta-analysis workflow.

LET'S **DIVE DEPPER** into the **methods** for **calculating** missing SDs

MORE ON THE OTHER SIDE

DEALING WITH MISSING STANDARD DEVIATIONS

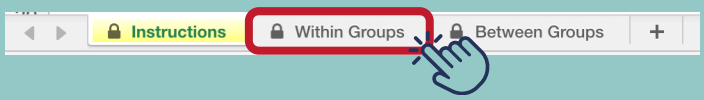
WE HAVE  **KEY SCENARIOS** FOR CALCULATING **MISSING SDs**

SCENARIO 1 ESTIMATING SD OF A SINGLE GROUP

NEXT

Inside the sheet called "**Within groups**", you can estimate the SD of a mean of a single group using the corresponding

Standard Error or **Confidence Intervals**

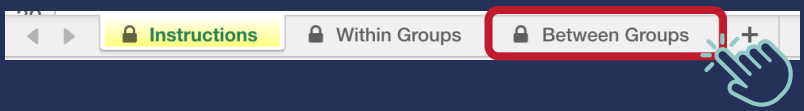


SCENARIO 2 ESTIMATING SD FROM THE MEAN DIFFERENCE AND RELATED STATISTICS

Inside the sheet called "**Between groups**", you can obtain the SD from a

Standard Error or **Confidence Intervals** or **t-statistics** or **P-value**

that relates to a difference between means in two groups



NEXT

DEALING WITH MISSING STANDARD DEVIATIONS

SCENARIO 1 ESTIMATING SD OF A SINGLE GROUP

EXAMPLE

Group	Mean	SD	SE	Lower CI	Upper CI	n	p-value
Intervention	24.68	2.99	0.19	24.31	25.05	250	0.00171
Control	27.17	3.32	0.22	26.73	27.61	220	

Suppose the studies did not report the **SD**, and it is **missing**. We can estimate the SD using the available data:

OPTION 1 USING THE STANDARD ERROR (SE) OF A MEAN

Finding the Standard Deviation using the Standard Error:
When making this transformation, standard errors must be of means calculated from within an intervention group and not standard errors of the difference in means computed between study groups.

GROUP 1		GROUP 2	
Standard Error =	0.19	Standard Error =	0.22
Number of participants =	250	Number of participants =	220
Standard Deviation =	3.0041638	Standard Deviation =	3.263127334

Calculation used: SD = SE x √N

OPTION 2 USING THE CONFIDENCE INTERVAL (CI) OF A MEAN

Finding the Standard Deviation using Confidence Intervals (CI):
This transformation is for Confidence Intervals for mean values calculated within a study group and not for estimates of differences between study groups.
It is important to check that the confidence interval is symmetrical about the mean (the distance between the lower limit and the mean is the same as the distance between the mean and the upper limit). If this is not the case, the confidence interval may have been calculated on transformed values.

GROUP 1		GROUP 2	
What % was the CI? (for 95% CI type 0.95):	0.95	What % was the CI? (for 95% CI type 0.95):	0.95
Number of participants =	250	Number of participants =	220
Upper Confidence Interval limit =	25.05	Upper Confidence Interval limit =	27.61
Lower Confidence Interval limit =	24.31	Lower Confidence Interval limit =	26.73
Mean =	24.68	Mean =	27.17
The Confidence Interval is	3.92 standard errors wide.	The Confidence Interval is	3.92 standard errors wide.
Standard Deviation =	2.9848029	Standard Deviation =	3.32972177

DEALING WITH MISSING STANDARD DEVIATIONS

SCENARIO 2 ESTIMATING SD FROM THE MEAN DIFFERENCE AND RELATED STATISTICS

EXAMPLE

MD	Lower CI	Upper CI	n intervention	n control	P-value
1.23	0.31	2.15	250	250	0.0091

OPTION 1 USING THE CONFIDENCE INTERVAL (CI) OF A MEAN DIFFERENCE

Using the p-value of a t-test to find the SD.

First you must derive the corresponding t-value:

p-value = 0.0091

No. of participants in Group 1 = 250

No. of participants in Group 2 = 250

t-value = 2.6185292

Now go to the calculation box for t-values.

Using the t-value to find the SD

First you must derive the corresponding Standard Error:

t-value = 2.6185292

Difference in means (between groups) = 1.23

Standard Error = 0.4693878

Now go to the calculation box for Standard Errors.

Using the Confidence Intervals to find the SD

First you must derive the Standard Error:

Upper CI limit = 2.15
Lower CI limit = 0.31
% CI used (enter 0.95 for 95%) = 0.95
No. of participants in Group 1 = 250
No. of participants in Group 2 = 250

(divisor = 3.92)

Standard Error = 0.4693878

Now go to the calculation box for Standard Errors.

Using the Standard Error to find the SD

Standard Error of the difference in means = 0.4693878

No. of participants in Group 1 = 250

No. of participants in Group 2 = 250

Estimated SD for each group = 5.2479151

FIRST STEP
USE CI TO FIND SE

SECOND STEP
USE SE TO FIND SD

IMPORTANT

An assumption that the SDs of outcome measurements are the same in both groups is required in all cases.
The same SD is then used for both intervention groups.

DEALING WITH MISSING STANDARD DEVIATIONS

SCENARIO 2 ESTIMATING SD FROM THE MEAN DIFFERENCE AND RELATED STATISTICS

EXAMPLE

MD	Lower CI	Upper CI	n intervention	n control	P-value
1.23	0.31	2.15	250	250	0.0091

OPTION 2 USING THE P-VALUE OF A MEAN DIFFERENCE

Using the p-value of a t-test to find the SD. 1

First you must derive the corresponding t-value:

p-value = 0.0091

No. of participants in Group 1 = 250

No. of participants in Group 2 = 250

t-value = 2.6185292

Now go to the calculation box for t-values.

Using the t-value to find the SD 2

First you must derive the corresponding Standard Error:

t-value = 2.6185292

Difference in means (between groups) = 1.23

Standard Error = 0.46972934

Now go to the calculation box for Standard Errors.

Using the Confidence Intervals to find the SD

First you must derive the Standard Error:

Upper CI limit = 2.15

Lower CI limit = 0.31

% CI used (enter 0.95 for 95%) = 0.95

No. of participants in Group 1 = 250

No. of participants in Group 2 = 250

(divisor = 3.92)

Standard Error = 0.4693878

Now go to the calculation box for Standard Errors.

Using the Standard Error to find the SD 3

Standard Error of the difference in means = 0.4693878

No. of participants in Group 1 = 250

No. of participants in Group 2 = 250

Estimated SD for each group = 5.2479151

When studies report levels of significance as ranges (e.g., P<0.05, P<0.01, or P=NS), rather than exact P-values, interpretation and calculations can become challenging. A conservative approach is to use the upper limit of the reported range. For example:

For **P<0.05**, assume **P=0.05** | For **P<0.01**, assume **P=0.01** | For **P<0.001**, assume **P=0.001**.

However, this method does not provide a solution for results reported as P=NS or P>0.05, as these values lack specificity and cannot be precisely interpreted for further statistical calculations.

DEALING WITH MISSING STANDARD DEVIATIONS

LAST RESORT

IN CASE WHERE DATA TO CALCULATE SD ARE NOT AVAILABLE

STANDARD DEVIATION IMPUTATION

Sometimes, the only option for handling missing SDs is to **borrow** them from other studies.

If **multiple SDs** are available, you might use their **average**, the **highest value**, or a "**reasonably high**" value.

However, since all imputation methods rely on assumptions, they should be used sparingly. When **most studies** in a meta-analysis **have missing SDs, imputation is not advisable**.

Study ID	Group	Mean	SD	Sample Size (n)
Study 1	Experimental	25.4	4.5	50
Study 1	Control	27.1	5.0	50
Study 2	Experimental	23.6	4.2	45
Study 2	Control	26.8	5.3	45
Study 3	Experimental	24.0	(Missing)	40
Study 3	Control	27.5	(Missing)	40

- **Average SD =** (4.5 + 5.0 + 4.2 + 5.3) / 4 = 4.75
- **Highest SD =** 5.3
- **Reasonably High SD =** 5.075 - Imputed using the 75th percentile

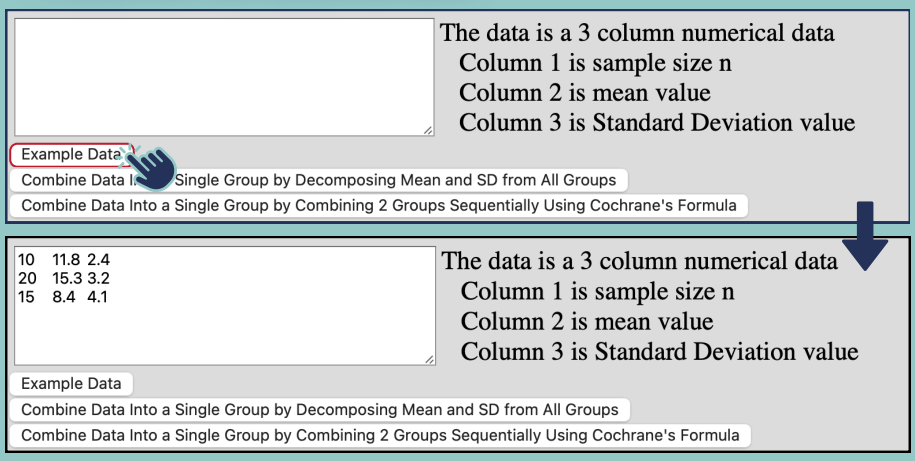
COMBINING SUBGROUPS INTO A SINGLE GROUP

Sometimes it is desirable to combine two or more reported subgroups **into a single group**.

For example, '**Group 1**' and '**Group 2**' may refer to two **slightly different variants** of an intervention to which participants were randomized, such as **different doses** of the **same drug**.

Scan the QR code below to visit the StatsToDo website and streamline your meta-analysis workflow with its user-friendly tools.

1. CLICK ON THE EXAMPLE DATA



COMBINING SUBGROUPS INTO A SINGLE GROUP

2. COMPLETE THE TABLE WITH YOUR DATA

```
10   12.8 1.5
20   13.3 2.2
15   8.2  5.5
```

The data is a 3 column numerical data
Column 1 is sample size n
Column 2 is mean value
Column 3 is Standard Deviation value

Example Data
Combine Data Into a Single Group by Decomposing Mean and SD from All Groups
Combine Data Into a Single Group by Combining 2 Groups Sequentially Using Cochrane's Formula

3. COMBINE THE RESULTS

```
10   12.8 1.5
20   13.3 2.2
15   8.2  5.5
```

The data is a 3 column numerical data
Column 1 is sample size n
Column 2 is mean value
Column 3 is Standard Deviation value

Example Data
Combine Data Into a Single Group by Decomposing Mean and SD from All Groups
Combine Data Into a Single Group by Combining 2 Groups Sequentially Using Cochrane's Formula

Combine groups of means and Standard Deviations into a Single Group by Repeating Cochrane's formula

	Individual Groups			Combined with previous Group		
	n	Mean	Standard Deviation	n	Mean	Standard Deviation
Row 1	10	12.8	1.5	10	12.8	1.5
Row 2	20	13.3	2.2	30	13.1333	1.9816
Row 3	15	8.2	5.5	45	11.4889	4.2124

Pooled Mean: 11.49

Pooled Standard Deviation: 4.21

THIS IS THE OUTPUT

IX – Reporting the Methods

It is highly advised to specify as the aspects of the model you used in the methods section of your meta-analysis report, including:

(1) the summary measure;
(2) the method to pool the studies;
(3) the between-study variance estimator;
(4) and other details you applied in your analyses.

EXAMPLE

Statistical Analysis

We summarized continuous endpoints using the inverse-variance random-effects model, with mean differences (MD) and 95% confidence interval (CI) as a measure of effect size. We estimated missing means and standard deviations using the formula proposed by Wan and Luo which derives these values from medians, interquartile ranges, and ranges as recommended by the Cochrane Collaboration.[1,2] The restricted-maximum likelihood (REML) estimator was used to calculate the between-study variance t^2.[3] We assessed heterogeneity with Cochrane's Q statistic and Higgins and Thompson's I^2 statistic, with $p \leq 0.10$ indicating statistical significance. We determined the consistency of the studies based on I^2 values of 0%, ≤ 25%, ≤ 50%, and > 50%, indicating no observed, low, moderate, and substantial heterogeneity, respectively.[4] All tests were two-tailed, and a p-value of < 0.05 was considered statistically significant. We used R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) and the extension package "meta" for all calculations and graphics.[5,6]

References

1. Viechtbauer, W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics. 2005;30(3):261-293.
2. Luo D, Wan X, Liu J, Tong T (2018): Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Statistical Methods in Medical Research, 27, 1785–805.
3. Wan X, Wang W, Liu J, Tong T (2014): Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology, 14, 135.
4. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine.* 2002;21(11):1539-1558.
5. Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
6. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health*. 22nd ed.; 2019:153-160.

X – Bibliography

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Between-study variance estimators

Langan D, Higgins JP, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*. 2005;10(1):83–98.

Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*.2016;7(1):55–79.

Restricted maximum likelihood estimator:
Viechtbauer, W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*. 2005;30(3):261-293.

DerSimonian-Laird estimator:
DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*. 1986;7(3):177–188.

Heterogeneity Assessment

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 2002;21(11):1539-1558.

Methods for Estimating Means and SD

Luo D, Wan X, Liu J, Tong T (2018): Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Statistical Methods in Medical Research, 27, 1785–805

Wan X, Wang W, Liu J, Tong T (2014): Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology, 14, 135

R, RStudio, and R packages

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

meta package:
Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health*. 22nd ed.; 2019:153-160.

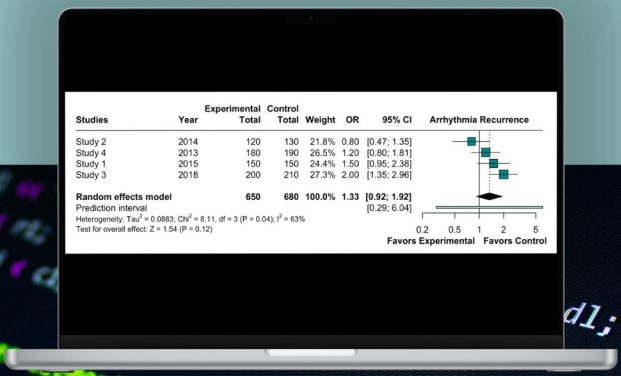
readxl package:
Wickham H, Bryan J (2023). readxl: Read Excel Files. https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.

MASTERING

PAIRWISE META-ANALYSIS

CHAPTER 3: PRE-CALCULATED EFFECT SIZE DATA

in RStudio

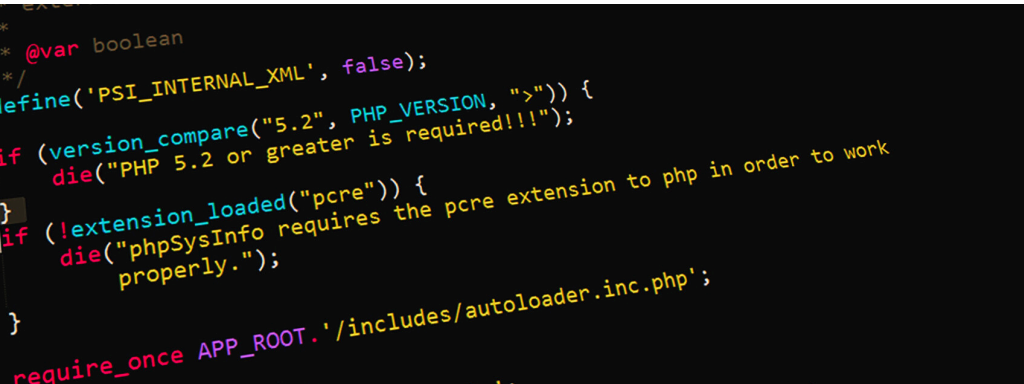


AN INTUITIVE AND STRAIGHTFORWARD GUIDE

DOUGLAS MESADRI GEWEHR

EMILTON LIMA JUNIOR

SUMMARY



I – Essential R Packages

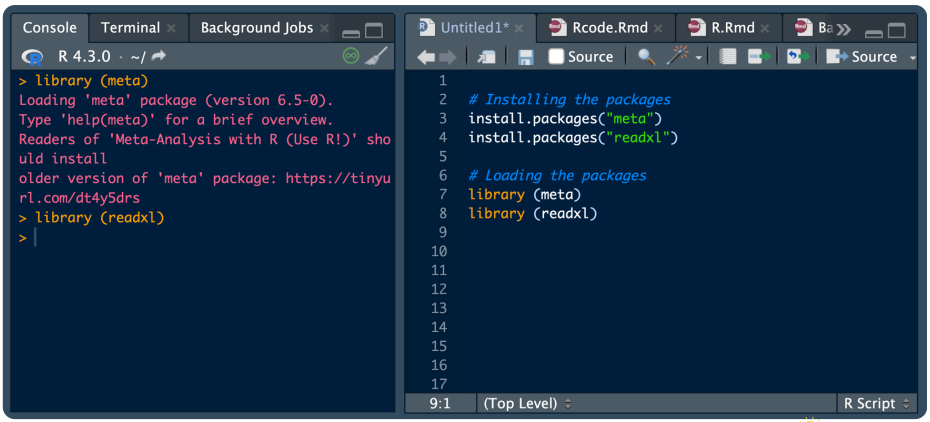
readxl package

A package that provides functions to read data from Excel files into R. It supports both ".xls" and ".xlsx" file formats, and can read data from individual worksheets or the entire workbook.

meta package

This package is a comprehensive package for meta-analysis, including functions for estimating and plotting effects, performing tests for heterogeneity, adjusting meta-regression models, and conducting subgroup analyses.

First, you'll need to **install** and **load** the packages as follows:



Once you install the packages with the **install.packages()** function, you need to load ("activate") the packages, using the **library()** function every time you start RStudio!

QUICK TIPS

II- Importing your Dataset

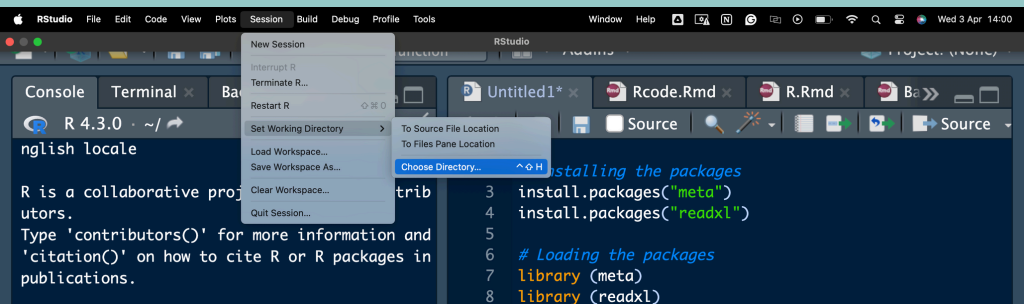
1st

Once you have your Excel workbook with all endpoints of your meta-analysis in its own worksheets ready, save the **.xlsx** file in a designated folder, commonly referred to as the '**working directory**' (WD), or any other preferred name.

For that, you need to use the **setwd()** function and add between quotes the directory of the folder WD in your computer that you've created:

```
#Setting your WD folder in R Studio:
setwd("/Users/douglasmesadrigewehr/Desktop/WD RStudio")
```

The other way is to manually configure the WD folder, as shown below:



2

The following R code creates a new object with the name of your choice (in this example I am using '**ma**', containing all the outcomes in a data frame object.

```
sheet_names <- excel_sheets("data.xlsx")
sheet_names

ma <- lapply(sheet_names, function(x) {
  as.data.frame(read_excel("data.xlsx", sheet = x)) } )

names(ma) <- sheet_names
```

ATTENTION PLEASE!

When using this code, you just need to rename the **.xlsx** file (arrows) that contains your dataset, located in your WD folder.

III – Preparing your Dataset

Before you begin your meta-analysis, you need to ensure that your data has been organized and structured correctly!

This will help ensure that the data is consistent and compatible with the functions and methods used in the meta-analysis, and can save time and effort in the analysis process.

Here are some **Dos and Don'ts** when preparing your data in Excel:

1. It's very important how you name the columns of your spreadsheet. If you already named the columns of your sheet appropriately in **Excel**, you can save a lot of time later because your data does not have to be transformed using R.

2. Column names should not contain any spaces. An alternative is to use underline sign "_".

3. It does not matter how columns are ordered in your Excel spreadsheet. They just have to be labeled correctly.

4. It is also important to know that the import may distort special characters like ä, ü, ö, and so forth. You might want to transform them into "normal" letters before you proceed.

5. If you have one or several empty rows or columns which used to contain data, make sure to delete those columns/rows completely because R could think that these columns contain (missing) data and import them also.

IV

RATIO MEASURES META-ANALYSIS

ODDS RATIO

RISK RATIO

HAZARD RATIO

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 1 COMPLETE DATA REPORT USING **OR**

ALL studies reporting **OR**

along with their **CI** or **STANDARD ERROR (SE)**

Structure your dataset following the example below:

Calculated effect sizes = **OR**

Standard Error

Sample size in experimental ('n.e') and control (n.c) groups

Subgroups

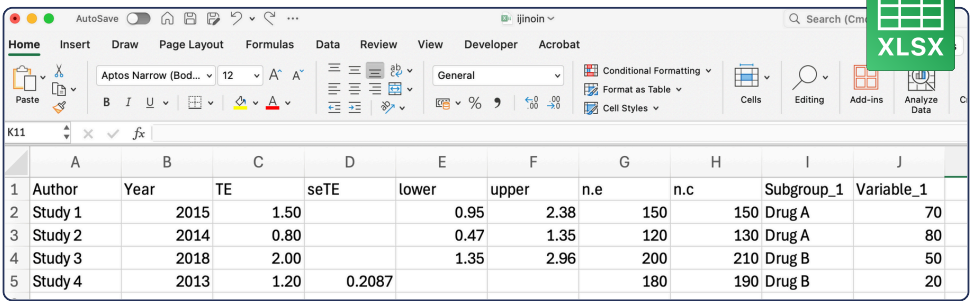
Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
Study 1	2015	1.36		0.95	2.38	150	150	Drug A	70
Study 2	2014	0.80		0.47	1.35	120	130	Drug A	80
Study 3	2018	2.03		1.35	2.96	200	210	Drug B	50
Study 4	2013	1.21	0.2087			180	190	Drug B	20

Meta-regression

Study Identification

Lower and upper limit of confidence interval

XLSX



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	1.50		0.95	2.38	150	150	Drug A	70
3	Study 2	2014	0.80		0.47	1.35	120	130	Drug A	80
4	Study 3	2018	2.00		1.35	2.96	200	210	Drug B	50
5	Study 4	2013	1.20	0.2087			180	190	Drug B	20

If the study reports the **SE**, input the **SE** directly into the spreadsheet. If the study provides a **CI**, enter the **upper** and **lower** limits of the **CI** into the respective columns of the spreadsheet.

If the study reports both the **SE** and the **CI** (rare!), choose one to report. **SE** is preferred in such cases, as the **CI** would typically be used to estimate and calculate the SE.

This ensures that the data are standardized and ready for meta-analysis without additional processing.

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 2 MIXED DATA REPORT USING OR

SOME studies reporting **OR + CI** or **SE**

WHILE OTHERS provide only **RAW DATA**

1st Step Structure your dataset following the example below:

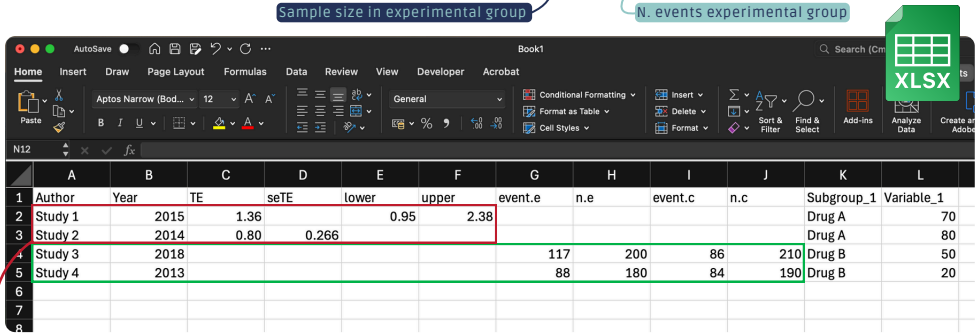
N. events experimental group

Sample size in experimental group

Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
Study 1	2015	1.36		0.95	2.38					Drug A	70
Study 2	2014	0.80	0.2660							Drug A	80
Study 3	2018					117	200	86	210	Drug B	50
Study 4	2013					88	180	84	190	Drug B	20

Sample size in experimental group

N. events experimental group



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	1.36		0.95	2.38					Drug A	70
3	Study 2	2014	0.80	0.266							Drug A	80
4	Study 3	2018					117	200	86	210	Drug B	50
5	Study 4	2013					88	180	84	190	Drug B	20
6												
7												
8												

These studies require **no additional preparation**. The provided values for **OR** and either **CI** or **SE** can be directly used in the **metagen()** function imported into R.

Identify the studies that reported raw data **2nd Step**

And then calculate the **OR** and **SE(log[OR])** using the appropriate formulas!

MORE ON THE OTHER SIDE

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 2 MIXED DATA REPORT USING OR

SOME studies reporting **OR + CI** or **SE**

WHILE OTHERS provide only **RAW DATA**

3rd Step Calculate the **OR** and **SE** from raw data:

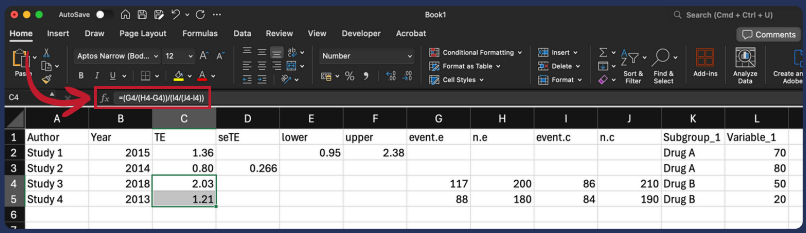
CALCULATING THE ODDS RATIO

$$OR = \frac{(\text{Experimental Events} \,/\, \text{Experimental Non-Events})}{(\text{Control Events} \,/\, \text{Control Non-Events})}$$

In the column labeled **TE**, insert the following formula to calculate the **OR**:

```
=(event.e / (n.e - event.e)) / (event.c / (n.c - event.c))
```

Replace **event.e**, **n.e**, **event.c**, and **n.c** with their respective cell references (e.g., G4, H4, I4, J4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	1.36		0.95	2.38					Drug A	70
3	Study 2	2014	0.80	0.266							Drug A	80
4	Study 3	2018	2.03				117	200	86	210	Drug B	50
5	Study 4	2013	1.21				88	180	84	190	Drug B	20
6												

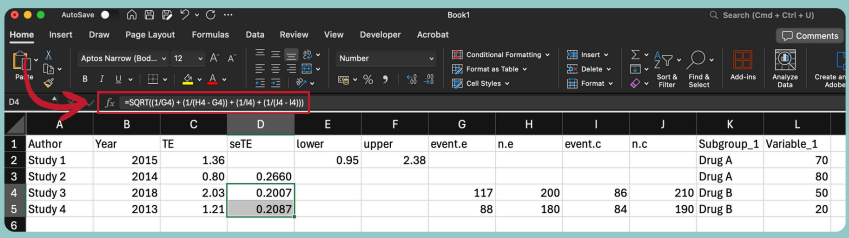
CALCULATING THE STANDARD ERROR

$$SE(\log(OR)) = \sqrt{\frac{1}{\text{Experimental Events}} + \frac{1}{\text{Experimental Non-Events}} + \frac{1}{\text{Control Events}} + \frac{1}{\text{Control Non-Events}}}$$

In the column labeled **seTE**, insert the following formula to calculate the **SE**:

```
=SQRT((1/event.e) + (1/(n.e - event.e)) + (1/event.c) + (1/(n.c - event.c)))
```

Replace **event.e**, **n.e**, **event.c**, and **n.c** with their respective cell references (e.g., G4, H4, I4, J4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	1.36		0.95	2.38					Drug A	70
3	Study 2	2014	0.80	0.2660							Drug A	80
4	Study 3	2018	2.03	0.2007			117	200	86	210	Drug B	50
5	Study 4	2013	1.21	0.2087			88	180	84	190	Drug B	20
6												

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 3 COMPLETE DATA REPORT USING RR

ALL studies reporting **RR**

along with their **CI** or **STANDARD ERROR (SE)**

Structure your dataset following the example below:

Calculated effect sizes = RR

Standard Error

Sample size in experimental ('n.e') and control (n.c) groups

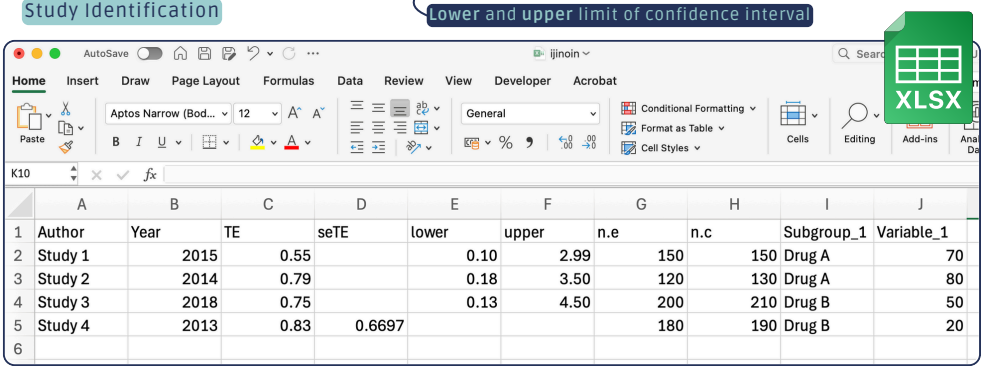
Subgroups

Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
Study 1	2015	0.55		0.10	2.99	150	150	Drug A	70
Study 2	2014	0.79		0.18	3.50	120	130	Drug A	80
Study 3	2018	0.75		0.13	4.50	200	210	Drug B	50
Study 4	2013	0.83	0.6697			180	190	Drug B	20

Meta-regression

Study Identification

Lower and upper limit of confidence interval



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	0.55		0.10	2.99	150	150	Drug A	70
3	Study 2	2014	0.79		0.18	3.50	120	130	Drug A	80
4	Study 3	2018	0.75		0.13	4.50	200	210	Drug B	50
5	Study 4	2013	0.83	0.6697			180	190	Drug B	20
6										

If the study reports the **SE**, input the **SE** directly into the spreadsheet. If the study provides a **CI**, enter the **upper** and **lower** limits of the **CI** into the respective columns of the spreadsheet.

If the study reports both the **SE** and the **CI** (rare!), choose one to report. **SE** is preferred in such cases, as the **CI** would typically be used to estimate and calculate the SE.

This ensures that the data are standardized and ready for meta-analysis without additional processing.

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 4 MIXED DATA REPORT USING RR

SOME studies reporting **RR + CI** or **SE**

WHILE OTHERS provide only **RAW DATA**

1st Step Structure your dataset following the example below:

N. events experimental group

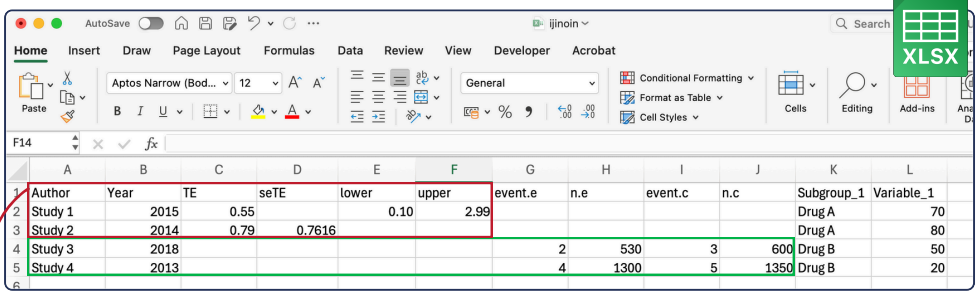
Sample size in experimental group

Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
Study 1	2015	0.55		0.10	2.99					Drug A	70
Study 2	2014	0.79	0.7616							Drug A	80
Study 3	2018					2	530	3	600	Drug B	50
Study 4	2013					4	1300	5	1350	Drug B	20

Sample size in experimental group

N. events experimental group

XLSX



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	0.55		0.10	2.99					Drug A	70
3	Study 2	2014	0.79	0.7616							Drug A	80
4	Study 3	2018					2	530	3	600	Drug B	50
5	Study 4	2013					4	1300	5	1350	Drug B	20
6												

These studies require **no additional preparation**. The provided values for **RR** and either **CI** or **SE** can be directly used in the **metagen()** function imported into R.

Identify the studies that reported raw data **2nd Step**

And then calculate the **RR** and **SE(log[RR])** using the appropriate formulas!

MORE ON THE OTHER SIDE

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 4 MIXED DATA REPORT USING RR

SOME studies reporting **RR + CI** or **SE**

WHILE OTHERS provide only **RAW DATA**

3rd Step Calculate the **OR** and **SE** from raw data:

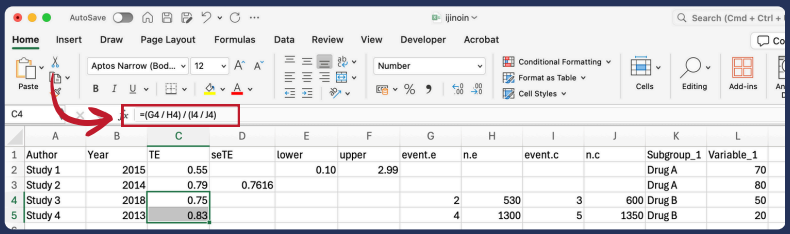
CALCULATING THE RELATIVE RISK (OR RISK RATIO)

$$RR = \frac{\text{event.e}/\text{n.e}}{\text{event.c}/\text{n.c}}$$

In the column labeled **TE**, insert the following formula to calculate the **RR**:

```
=(event.e / n.e) / (event.c / n.c)
```

Replace **event.e**, **n.e**, **event.c**, and **n.c** with their respective cell references (e.g., G4, H4, I4, J4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	0.55		0.10	2.99					Drug A	70
3	Study 2	2014	0.79	0.7616							Drug A	80
4	Study 3	2018	0.75				2	530	3	600	Drug B	50
5	Study 4	2013	0.83				4	1300	5	1350	Drug B	20

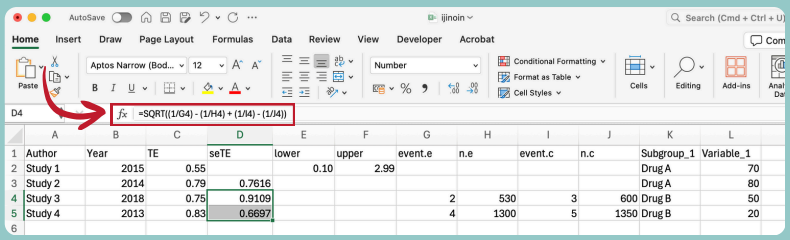
CALCULATING THE STANDARD ERROR

$$SE(\log(RR)) = \sqrt{\frac{1}{\text{event.e}} - \frac{1}{\text{n.e}} + \frac{1}{\text{event.c}} - \frac{1}{\text{n.c}}}$$

In the column labeled **seTE**, insert the following formula to calculate the **SE**:

```
=SQRT((1/event.e) - (1/n.e) + (1/event.c) - (1/n.c))
```

Replace **event.e**, **n.e**, **event.c**, and **n.c** with their respective cell references (e.g., G4, H4, I4, J4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Author	Year	TE	seTE	lower	upper	event.e	n.e	event.c	n.c	Subgroup_1	Variable_1
2	Study 1	2015	0.55		0.10	2.99					Drug A	70
3	Study 2	2014	0.79	0.7616							Drug A	80
4	Study 3	2018	0.75	0.9109			2	530	3	600	Drug B	50
5	Study 4	2013	0.83	0.6697			4	1300	5	1350	Drug B	20
6												

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

SCENARIO 1

Available data: **ALL** studies reporting **OR** along with their **CI** or **SE**.
Data Handling: **NO** manipulation needed. If both **CI** and **SE** are reported, prefer **SE**.

SCENARIO 2

Available data: **SOME** studies reporting **OR** along with their **CI** or **SE**, **WHILE** others provide only **RAW** data.
Data Handling: calculate the **OR** and **SE** using **event counts** and **totals**. Then, integrate the calculated values with the pre-reported OR and SE into the dataset.

SCENARIO 3

Available data: **ALL** studies reporting **RR** along with their **CI** or **SE**.
Data Handling: **NO** manipulation needed. If both **CI** and **SE** are reported, prefer **SE**.

SCENARIO 4

Available data: **SOME** studies reporting **RR** along with their **CI** or **SE**, **WHILE** others provide only **RAW** data.
Data Handling: calculate the **RR** and **SE** using **event counts** and **totals**. Then, integrate the calculated values with the pre-reported RR and SE into the dataset.

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

HELPFUL TIPS

FOR RATIOS EFFECTS MEASURES

1. **Hazard Ratios** CANNOT be combined directly with **Odds Ratios** OR or **Risk Ratios** due to differences in their statistical meaning and interpretation.

2. Always prioritize **adjusted ratio effect measures** derived from **multivariate analysis** or **propensity score matching** (PSM). These **account for confounding** factors and provide more reliable estimates compared to unadjusted ratios

   But keep in mind that it could be a source of heterogeneity, especially when you're pooling adjusted with unadjusted effect size data.

3. **Standardize** all **effect measures** to a **single metric** (e.g., OR) when possible, using appropriate conversion formulas, to maintain consistency across the meta-analysis.

4. **Logistic regression** models report **adjusted ratio effect measures** as **OR**, not **RRs**. As a result, many meta-analyses using adjusted ratio effect measures are conducted with **ORs** rather than **RRs**.

5. Clearly state any transformations, conversions, or assumptions made during the meta-analysis. Transparency ensures reproducibility and strengthens the study's credibility.

IV.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR TIME-TO-EVENT ANALYSIS

SCENARIO 1 — COX PROPORTIONAL HAZARDS MODEL

ALL studies reporting **HR** or **LN(HR)** along with their **CI** or **STANDARD ERROR (SE)**

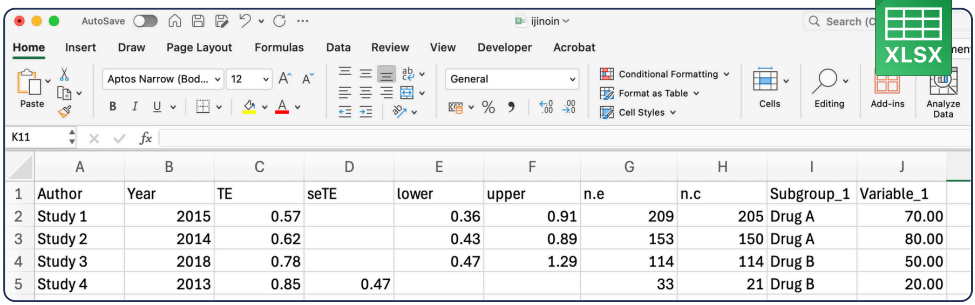
Structure your dataset following the example below:

Calculated effect sizes = HR | Standard Error | Sample size in experimental ('n.e') and control (n.c) groups | Subgroups | Meta-regression

Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
Study 1	2015	0.57		0.36	0.91	209	205	Drug A	70.00
Study 2	2014	0.62		0.43	0.89	153	150	Drug A	80.00
Study 3	2018	0.78		0.47	1.29	114	114	Drug B	50.00
Study 4	2013	0.85	0.47			33	21	Drug B	20.00

Study Identification | Lower and upper limit of confidence interval

XLSX



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	0.57		0.36	0.91	209	205	Drug A	70.00
3	Study 2	2014	0.62		0.43	0.89	153	150	Drug A	80.00
4	Study 3	2018	0.78		0.47	1.29	114	114	Drug B	50.00
5	Study 4	2013	0.85	0.47			33	21	Drug B	20.00

If the study reports the **SE**, input the **SE** directly into the spreadsheet. If the study provides a **CI**, enter the **upper** and **lower** limits of the **CI** into the respective columns of the spreadsheet.

If the study reports both the **SE** and the **CI** (rare!), choose one to report. **SE** is preferred in such cases, as the **CI** would typically be used to estimate and calculate the SE.

This ensures that the data are standardized and ready for meta-analysis without additional processing.

POSSIBLE SCENARIOS FOR TIME-TO-EVENT ANALYSIS

SCENARIO 2 USING LOG-RANK ANALYSIS

The second approach is to estimate the hazard ratio and its CI or SE approximately using statistics computed during a log-rank analysis.

Estimating the ln(HR):

$$\ln(HR) = \frac{O - E}{V}$$

Where:

- **O:** Observed number of events in the experimental group.
- **E:** Expected number of events in the experimental group (based on the log-rank test).
- **O - E:** Log-rank statistic.
- **V:** Variance of the log-rank statistic.

Estimating the SE:

$$SE = \frac{1}{\sqrt{V}}$$

This explanation is a simplification of the many possible scenarios for estimating **HR** and their **SE** from log-rank analysis data.

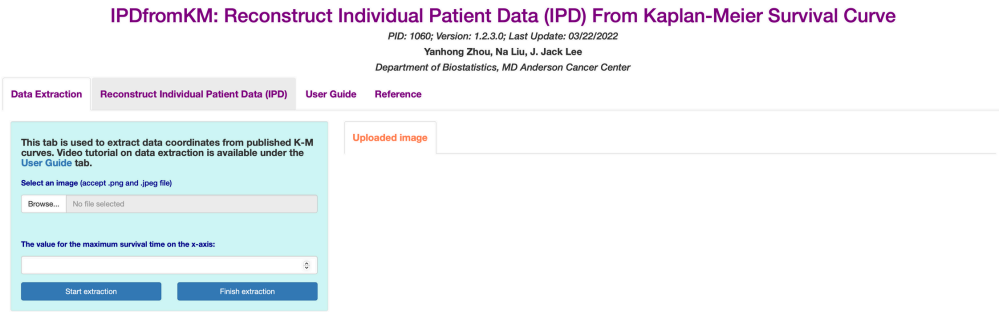
For a comprehensive understanding of all scenarios and detailed practical guidance, review authors are encouraged to consult Tierney and colleagues (Tierney et al., 2007). Scan the QR code below to access the full article.

Tierney and colleagues (Tierney et al., 2007) provide a calculator to simplify the estimation of hazard ratios (HR) and their standard errors (SE) from log-rank analysis data. To access this helpful tool, scan the QR code below.

POSSIBLE SCENARIOS FOR TIME-TO-EVENT ANALYSIS

SCENARIO 3 RECONSTRUCT APPROXIMATE INDIVIDUAL PARTICIPANT DATA FROM PUBLISHED KAPLAN-MEIER CURVES

This allows reanalysis of the data to estimate the hazard ratio, and also allows alternative approaches to analysis of the time-to-event data.

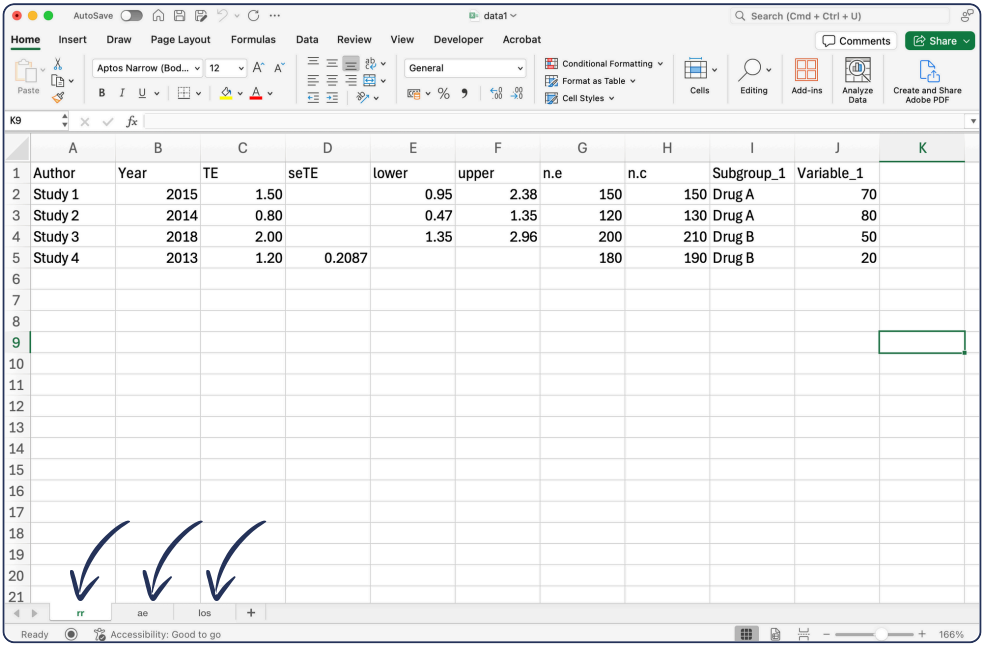


IV.I – Preparing your Dataset

Create a spreadsheet for each outcome of your meta-analysis in separate worksheets in your Excel workbook!

This approach saves time by allowing you to import **ALL** your endpoints at once into RStudio.

Here is an example of a dataset with 3 outcomes, each one in its own worksheet.



	A	B	C	D	E	F	G	H	I	J	K
1	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1	
2	Study 1	2015	1.50		0.95	2.38	150	150	Drug A	70	
3	Study 2	2014	0.80		0.47	1.35	120	130	Drug A	80	
4	Study 3	2018	2.00		1.35	2.96	200	210	Drug B	50	
5	Study 4	2013	1.20	0.2087			180	190	Drug B	20	
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											

IV.II – Visualizing your Data

Now, we can have a look at each of the lists (our endpoints) of our data frame object.

The object **ma,** we've just imported, has **3** lists of data:

1. "**rr**" for Recurrence Rate data
2. "**ae**" for Adverse Event data
3. "**los**" for Length of Stay data

To print one specific endpoint, you can do so by adding the dollar sign operator "**$**" after "**ma**" inside de **View()** function:

```
View(ma$rr) #to visualize Recurrence Rate data
View(ma$ae) #to visualize Adverse Event data
View(ma$los) #to visualize Length of Stay data
```

This is how your endpoint dataset will appear in RStudio. See the example below:

EXAMPLE

ma$rr | Untitled1* | Filter

	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
1	Study 1	2015	1.5	*NA*	0.95	2.38	150	150	Drug A	70
2	Study 2	2014	0.8	*NA*	0.47	1.35	120	130	Drug A	80
3	Study 3	2018	2.0	*NA*	1.35	2.96	200	210	Drug B	50
4	Study 4	2013	1.2	0.2087	*NA*	*NA*	180	190	Drug B	20

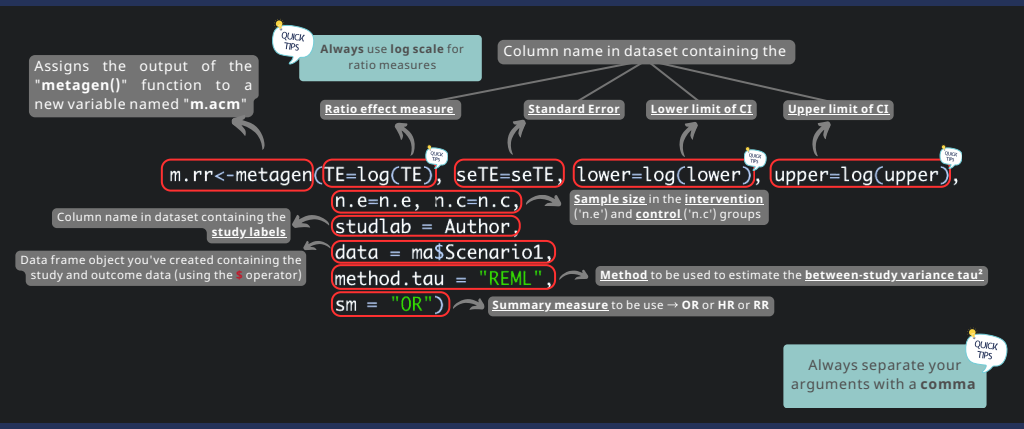
Showing 1 to 4 of 4 entries, 10 total columns

IV.III – Performing the Meta-Analysis

FOR RATIO MEASURES

To perform the meta-analysis of pre-calculated effect size data, using **ratio effect measures**, we will use the **metagen()** function.

We need to provide some **instructions** for the **metagen()** function. These instructions are known as **arguments**, which can be one, two, three, or more inputs that the function uses to perform its task. The main arguments of the **metagen()** function are:



The provided code using the **metagen()** function is applicable to **all five aforementioned scenarios** of available data **for binary outcomes** because:

1. If the **Standard Error ('seTE')** is not provided, the **'upper'** (upper limit of CI) and **'lower'** (lower limit of CI) arguments will be automatically used to estimate the **'seTE'**.

2. If the **'seTE'** is included in the dataset, the function will directly use it, bypassing the need to estimate from the confidence intervals.

To perform the meta-analysis, the ratio effect measure (e.g., OR, RR, or HR) **must be provided in its log scale**, along with its confidence interval. Ensure to include **'TE=log(TE)'**, **'lower=log(lower)'**, and **'upper=log(upper)'** in the dataset before running the analysis.

Over the next pages, we will detail the following aspects of binary outcome meta-analysis:
(1) methods for **pooling the studies**;
(2) methods to estimate the **between-study variance tau²**; and
(3) the **summary measures**.

MORE ON THE OTHER SIDE

IV.III – Performing the Meta-Analysis

FOR RATIO MEASURES

To visualize the results of a meta-analysis conducted with the **metagen()** function, we will use the **summary()** function. It will generate a summary of the analysis

```
summary(m.rr)
```

Call here the **metagen()** object you have just created

The **summary()** function will display the estimated treatment effect (i.e., the summary measure), the corresponding confidence interval, the between-study variance, and other relevant statistics.

```
Console | Terminal | Background Jobs
R 4.3.0 · ~/Desktop/WD RStudio/
> summary(m.rr)
                OR            95%-CI %W(common) %W(random)
Study 1 1.5000 [0.9500; 2.3800]       22.8       24.4
Study 2 0.8000 [0.4700; 1.3500]       17.3       21.8
Study 3 2.0000 [1.3500; 2.9600]       31.2       27.3
Study 4 1.2000 [0.7971; 1.8065]       28.7       26.5

Number of studies: k = 4
Number of observations: o = 1330 (o.e = 650, o.c = 680)

                          OR           95%-CI    z p-value
Common effect model   1.3806 [1.1088; 1.7191] 2.88  0.0039
Random effects model  1.3335 [0.9242; 1.9241] 1.54  0.1240

Quantifying heterogeneity:
 tau^2 = 0.0883 [0.0000; 2.0189]; tau = 0.2972 [0.0000; 1.4209]
 I^2 = 63.0% [0.0%; 87.5%]; H = 1.64 [1.00; 2.83]

Test of heterogeneity:
    Q d.f. p-value
 8.11    3  0.0438

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
>
```

This section contains:
(1) the individual studies;
(2) their effect sizes and confidence intervals;
(3) their weights (%) in common (i.e. fixed-) and random-effects

The **total number of studies** in our meta-analysis

The **total number of patients** in our meta-analysis

The POOLED effect size

Results concerning the **between-study heterogeneity**

Method to be used **for pooling the studies**

Method to be used to estimate the **between-study variance tau^2**

IV.III – Performing the Meta-Analysis

FOR RATIO MEASURES

METHODS FOR POOLING THE PRE-CALCULATED EFFECT SIZES MEASURES

Which method should I use?

The **metagen()** function uses the **generic inverse-variance** method for pooling the studies. This approach weights each study by the inverse of its variance, ensuring more precise studies contribute more to the overall estimate.

This method is recommended by the **Cochrane Collaboration** and is the **default** in the **metagen()** function.

You **do not need** to **specify** this method explicitly when using **metagen()**, as it is automatically applied.

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

1. Binary outcome?
   - Yes
2. Is the **heterogeneity** of effects in your sample **very large**? Is **avoiding** false-positives a **priority**?
   - Yes → Sidik-Jonkman (**"SJ"**)
   - No
3. Is there **extreme variation** (small and large studies) in the **sample sizes** among included studies?
   - Yes → DerSimonian-Laird (**"DL"**) or Restricted Maximum Likelihood (**"REML"**)
   - No → BEST CHOICE **Paule-Mandel** (**"PM"**)

DerSimonian-Laird (**"DL"**) estimator can be **biased**, particularly when the number of studies is **small** and **heterogeneity** is **high.**

LEARNING R SYNTAX

Methods to estimate the between-study variance

```
method.tau = "REML" # Restricted maximum-likelihood estimator
method.tau = "ML" # Maximum-likelihood estimator
method.tau = "PM" # Paule-Mandel estimator
method.tau = "DL" # DerSimonian-Laird estimator
method.tau = "SJ" # Sidik-Jonkman estimator
method.tau = "HE" # Hedges estimator
method.tau = "EB" # Empirical Bayes estimator
```

Always use " " when referring to these methods

QUICK TIPS

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

There are no strict guidelines determining the choice of estimator. In many cases, there are only slight variations in results among different estimators. So, you **shouldn't** worry **TOO** much about it!

If doubt arise, you can always rerun your analyses using different estimators, and check if it changes the interpretation of your results.

IV.IV – Creating the Forest Plot

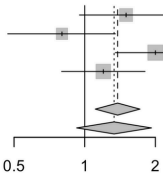
FOR RATIO MEASURES

You can produce a forest plot for **metagen()** object using the **forest()** function from the '**meta**' package

Usually, these forest plots do not have enough quality for publication by default! In the example below, we plot the '**m.acm**' object that we also used in the previous example:

```
#Creating a forest plot for 'm.acm' metagen object
forest(m.acm)
```

Study	logOR	SE(logOR)	Experimental Total	Control Total	OR	95%-CI	Weight (common)	Weight (random)
Study 1	0.4055	0.2343	150	150	1.50	[0.95; 2.38]	22.8%	24.4%
Study 2	-0.2231	0.2692	120	130	0.80	[0.47; 1.35]	17.3%	21.8%
Study 3	0.6931	0.2003	200	210	2.00	[1.35; 2.96]	31.2%	27.3%
Study 4	0.1823	0.2087	180	190	1.20	[0.80; 1.81]	28.7%	26.5%
Common effect model			**650**	**680**	**1.38**	**[1.11; 1.72]**	**100.0%**	**.**
Random effects model					**1.33**	**[0.92; 1.92]**	**.**	**100.0%**

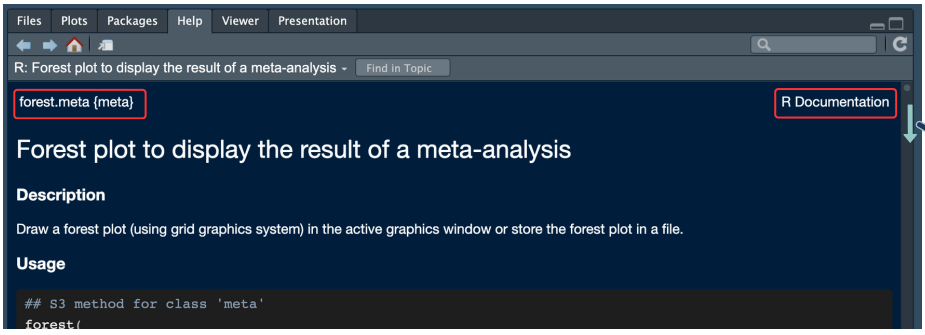


Heterogeneity: I^2 = 63%, τ^2 = 0.0883, p = 0.04

However, the **forest()** function also has countless additional arguments to further customize the appearance of the forest plot.

All these arguments are detailed and described in the function documentation. You can accesses this documentation by running the following code:

```
help(forest)
```



IV.IV – Creating the Forest Plot

FOR RATIO MEASURES

But don't worry. I've put together a code that brings together the main arguments for generating **elegant** and **high-quality** forest plots for publication.

Now, let's understand the most important arguments to customize your forest plot!

This is how your code looks in the end:

```
forest(m.rr,
       smlab = "Arrhythmia Recurrence",
       layout ="Revman5",
       sortvar = TE,
       lab.e = "Experimental", label.left = "Favors Experimental",
       lab.c = "Control", label.right = "Favors Control",
       ff.lr = "bold",
       leftcols = c("studlab", "Year", "n.e", "n.c", "w.random", "effect", "ci"),
       leftlabs = c("Studies", "Year", "Experimental\nTotal", "Control\nTotal", NA, NA, NA),
       text.random = "Random effects model",
       random = TRUE,
       common = FALSE,
       test.overall.random = TRUE,
       rightcols = FALSE,
       colgap = "3mm",
       fs.heading = 12,
       fs.study = 12,
       fs.hetstat = 10,
       digits = 2,
       digits.pval = 2,
       pooled.totals = TRUE,
       col.square="darkcyan", col.square.lines="black",
       prediction = TRUE, col.predict = "#CEF2EE",col.predict.lines = "black", ff.predict = 1)
```

The first argument always will be the **metagen()** object

1 Label printed at top of the plot. We usually use to print the endpoint analyzed!

If you omit this argument, the summary measure will be printed

To specify how to order the studies:
(1) 'TE' = effect size
(2) 'studlab' = alphabetical
(3) 'w.common' or 'w.random' = weight of the studies
(4) 'Year'
... and so forth

Write the name of your **experimental group** 2

To specify the **weight** to be used in the forest plot, use either **'w.random'** → for random-effects) or **'w.common'** → (for fixed-effects)

3 Write the name of your **control group**

To specify whether to use a **random-effects** (random = TRUE, common = FALSE) or **fixed-effects model** (random = FALSE, common = TRUE)

Write the name of your **experimental group** 2

Write the name of your **control group** 3

To specify whether to write **"Random effects model"** or **"Fixed effects model"** 4

To specify whether to print results of test for overall effect:
test.overall.random = **TRUE** → For **random effects model**
est.overall.common = **TRUE** → For **fixed effects model** 5

To specify the gap between the columns
This can be particularly useful when you have long names for the control and/or intervention groups, as these names can overlap and become difficult to read if the gap is not set appropriately.

To specify the minimal number of significant **digits** to print for the **effect estimates** (OR/RR/HR and CI)

To specify the minimal number of significant **digits** to print for p-value of overall treatment effect

To plot the **total number of participants** 6

To specify the **colour** for the **outer lines** of **squares** reflecting study's weight

To specify the **colour** for the **squares** reflecting study's weight

7 To print the **prediction interval**

To specify the **background colour** for the **prediction interval**

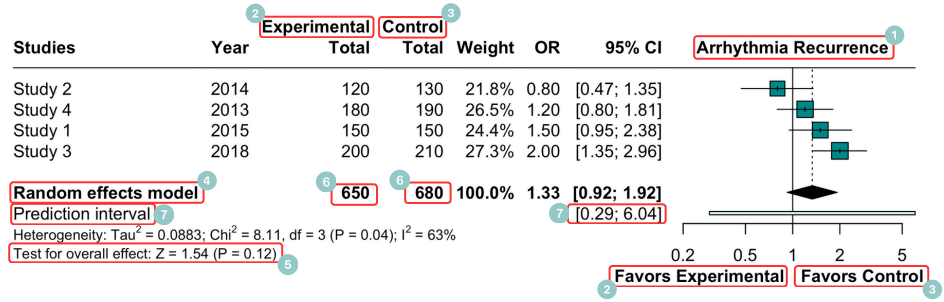
To specify the **colour** of the outer lines of **prediction interval**

MORE ON THE OTHER SIDE

IV.IV – Creating the Forest Plot

FOR RATIO MEASURES

This is how our forest plot looks in the end:



Studies	Year	Experimental Total	Control Total	Weight	OR	95% CI
Study 2	2014	120	130	21.8%	0.80	[0.47; 1.35]
Study 4	2013	180	190	26.5%	1.20	[0.80; 1.81]
Study 1	2015	150	150	24.4%	1.50	[0.95; 2.38]
Study 3	2018	200	210	27.3%	2.00	[1.35; 2.96]
Random effects model		**650**	**680**	**100.0%**	**1.33**	**[0.92; 1.92]**
Prediction interval						[0.29; 6.04]

Heterogeneity: Tau² = 0.0883; Chi² = 8.11, df = 3 (P = 0.04); I² = 63%
Test for overall effect: Z = 1.54 (P = 0.12)

Arrhythmia Recurrence
0.2 0.5 1 2 5
Favors Experimental Favors Control

V – CONTINUOUS OUTCOMES

DIFFERENCE MEASURES META-ANALYSIS

MEAN DIFFERENCE

STANDARDIZED MEAN DIFFERENCE

V.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 1 COMPLETE DATA REPORT USING MD

ALL studies reporting **MD**

along with their **CI** or **STANDARD ERROR (SE)**

Structure your dataset following the example below:

Calculated effect sizes = MD

Standard Error

Sample size in experimental ('n.e') and control (n.c) groups

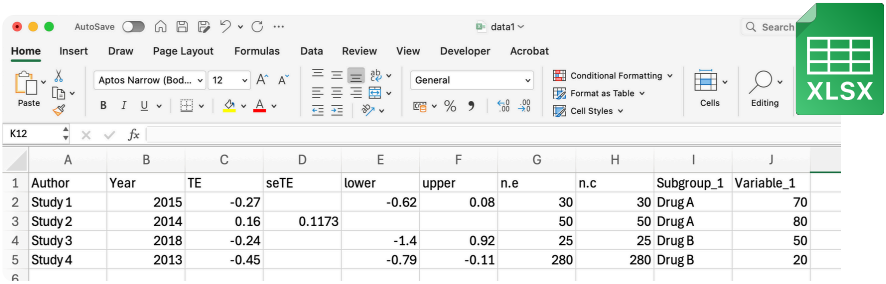
Subgroups

Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
Study 1	2015	-0.27		-0.62	0.08	30	30	Drug A	70
Study 2	2014	0.16	0.1173			50	50	Drug A	80
Study 3	2018	-0.24		-1.40	0.92	25	25	Drug B	50
Study 4	2013	-0.45		-0.79	-0.11	280	280	Drug B	20

Meta-regression

Study Identification

Lower and upper limit of confidence interval



	A	B	C	D	E	F	G	H	I	J
1	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	-0.27		-0.62	0.08	30	30	Drug A	70
3	Study 2	2014	0.16	0.1173			50	50	Drug A	80
4	Study 3	2018	-0.24		-1.4	0.92	25	25	Drug B	50
5	Study 4	2013	-0.45		-0.79	-0.11	280	280	Drug B	20
6										

If the study reports the **SE**, input the **SE** directly into the spreadsheet. If the study provides a **CI**, enter the **upper** and **lower** limits of the **CI** into the respective columns of the spreadsheet.

If the study reports both the **SE** and the **CI** (rare!), choose one to report. **SE** is preferred in such cases, as the **CI** would typically be used to estimate and calculate the SE.

This ensures that the data are standardized and ready for meta-analysis without additional processing.

V.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 2 MIXED DATA REPORT

SOME studies reporting **MD + CI** or **SE**

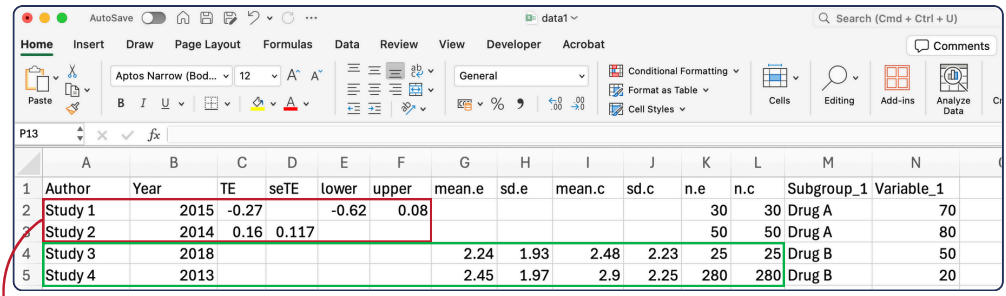
WHILE OTHERS provide only **MEANS AND SD**

1st Step Structure your dataset following the example below:

Author	Year	TE	seTE	lower	upper	mean.e	sd.e	mean.c	sd.c	n.e	n.c	Subgroup_1	Variable_1
Study 1	2015	-0.27		-0.62	0.08					30	30	Drug A	70
Study 2	2014	0.16	0.1173							50	50	Drug A	80
Study 3	2018					2.24	1.93	2.48	2.23	25	25	Drug B	50
Study 4	2013					2.45	1.97	2.90	2.25	280	280	Drug B	20

- Calculated effect sizes = MD
- Standard Error
- Mean - experimental group
- Standard Deviation - experimental group
- Study Identification
- Mean - control group
- Standard Deviation - control group
- Lower and upper limit of confidence interval
- Sample size in experimental ('n.e') and control (n.c) groups

Identify the studies that reported raw data **2nd Step**



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Author	Year	TE	seTE	lower	upper	mean.e	sd.e	mean.c	sd.c	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	-0.27		-0.62	0.08					30	30	Drug A	70
3	Study 2	2014	0.16	0.117							50	50	Drug A	80
4	Study 3	2018					2.24	1.93	2.48	2.23	25	25	Drug B	50
5	Study 4	2013					2.45	1.97	2.9	2.25	280	280	Drug B	20

These studies require **no additional preparation**. The provided values for **MD** and either **CI** or **SE** can be directly used in the **metagen()** function imported into R.

And then calculate the **MD** and **SE** using the appropriate formulas!

MORE ON THE OTHER SIDE

V.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 2 MIXED DATA REPORT

SOME studies reporting **MD + CI** or **SE**

WHILE OTHERS provide only **MEANS AND SD**

3rd Step Calculate the **MD** and **SE** from raw data:

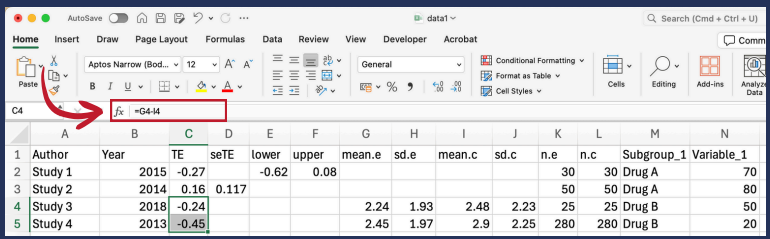
CALCULATING THE MEAN DIFFERENCE

$$MD = \text{Mean}_e - \text{Mean}_c$$

In the column labeled **TE**, insert the following formula to calculate the **MD**:

```
=mean.e - mean.c
```

Replace **mean.e** and **mean.c** with their respective cell references (e.g., G4, I4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Author	Year	TE	seTE	lower	upper	mean.e	sd.e	mean.c	sd.c	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	-0.27		-0.62	0.08					30	30	Drug A	70
3	Study 2	2014	0.16	0.117							50	50	Drug A	80
4	Study 3	2018	-0.24				2.24	1.93	2.48	2.23	25	25	Drug B	50
5	Study 4	2013	-0.45				2.45	1.97	2.9	2.25	280	280	Drug B	20

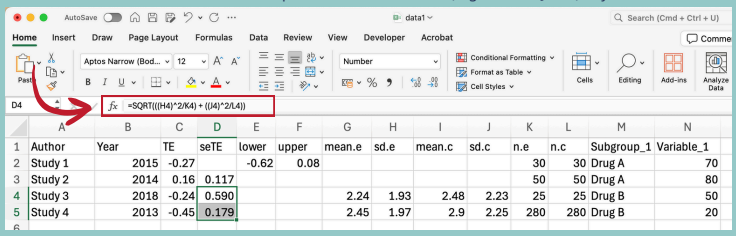
CALCULATING THE STANDARD ERROR

$$SE_{MD} = \sqrt{\frac{SD_e^2}{n_e} + \frac{SD_c^2}{n_c}}$$

In the column labeled **seTE**, insert the following formula to calculate the **SE**:

```
=SQRT(((sd.e)^2/n.e) + ((sd.c)^2/n.c))
```

Replace **sd.e**, **n.e**, **sd.c**, and **n.c** with their respective cell references (e.g., G4, H4, J4, L4) in your dataset.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Author	Year	TE	seTE	lower	upper	mean.e	sd.e	mean.c	sd.c	n.e	n.c	Subgroup_1	Variable_1
2	Study 1	2015	-0.27		-0.62	0.08					30	30	Drug A	70
3	Study 2	2014	0.16	0.117							50	50	Drug A	80
4	Study 3	2018	-0.24	0.590			2.24	1.93	2.48	2.23	25	25	Drug B	50
5	Study 4	2013	-0.45	0.179			2.45	1.97	2.9	2.25	280	280	Drug B	20

V.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 3 MIXED DATA REPORT

SOME studies reporting **MD + CI** or **SE**

WHILE OTHERS provide **MEANS AND SD** or **MEDIAN, IQR, MIN-MAX**

1st Step Structure your dataset following the example below:

Median - experimental group
Quartile 1 and 3 - experimental group
Minimum and Maximum - experimental group
Sample Size - experimental group

Author	Year	TE	seTE	lower	upper	mean.e	sd.e	median.e	q1.e	q3.e	min.e	max.e	n.e	mean.c	sd.c	median.c	q1.c	q3.c	min.c	max.c	n.c
Study 1	2015	-0.27		-0.62	0.08								30								30
Study 2	2014	0.16	0.1173										50								50
Study 3	2018					2.24	1.93						25	2.24	1.93						25
Study 4	2013							2.45	1.12	3.78	0.4	6.3	280			2.90	1.38	4.42	0.5	7.3	280

Median - control group
Quartile 1 and 3 - control group
Minimum and Maximum - control group
sample Size - control group

2nd Step For studies reporting **Median**, **Interquartile Range**, and/or **Minimum** and **Maximum** values, **first** estimate the **Mean** and **SD** using methods recommended by the Cochrane Collaboration (Wan & Luo method).

Scan the QR code below to visit a user-friendly website that helps you apply this method to find means ± SDs as needed.

V.I – Preparing your Dataset

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 3 MIXED DATA REPORT

SOME studies reporting **MD + CI** or **SE**

WHILE OTHERS provide **MEANS AND SD** or **MEDIAN, IQR, MIN-MAX**

2nd Step Estimating the Mean and SD:

Scenario 1

RARE

280	Size of the sample
0.4	Minimum of the sample
2.45	Median of the sample
6.3	Maximum of the sample

Reset | Detect the skewness | Calculate

2.4997	Estimated mean of the sample from Luo et al. (2018)
1.0374	Estimated standard deviation of the sample from Wan et al. (2014)

Median

Min

Max

Important Note: Ensure the study explicitly states that the range represents the minimum and maximum values of the sample. Do not confuse this with the Interquartile Range (IQR), which represents the range between the 25th percentile (Q1) and the 75th percentile (Q3). Properly distinguishing these measures is essential before performing any calculations.

MOST POPULAR

Scenario 2

280	Size of the sample
1.12	First quartile of the sample
2.45	Median of the sample
3.78	Third quartile of the sample

Reset | Detect the skewness | Calculate

2.45	Estimated mean of the sample from Luo et al. (2018)
1.9822	Estimated standard deviation of the sample from Wan et al. (2014)

Median

Q1

Q3

Important Note: When a study reports the median (IQR), the value before the comma inside the parentheses represents the first quartile (Q1), and the value after the comma represents the third quartile (Q3). Ensure these are correctly interpreted before performing any calculations.

POSSIBLE SCENARIOS FOR CONTINUOUS OUTCOMES

SCENARIO 3 MIXED DATA REPORT

SOME studies reporting **MD + CI** or **SE**

WHILE OTHERS provide **MEANS AND SD** or **MEDIAN, IQR, MIN-MAX**

2nd Step Estimating the Mean and SD:

Scenario 3

280	Size of the sample
0.4	Minimum of the sample
1.12	First quartile of the sample
2.45	Median of the sample
3.78	Third quartile of the sample
6.3	Maximum of the sample

Reset | Detect the skewness | Calculate

2.478	Estimated mean of the sample from Luo et al. (2018)
1.6732	Estimated standard deviation of the sample from Shi et al. (2020)

Median

Min

Q1

Q3

Max

By organizing your dataset to include columns for **Median**, **Q1**, **Q3**, **Min**, and **Max**, you can easily estimate missing values when needed. This ensures flexibility without losing valuable studies.

You need at least **one pair** (Q1-Q3 or Min-Max) of dispersion measures. **The more** measures you have, **the more accurate** the estimation will be.

3rd Step Calculate the **MD** and **SE** from raw data, as explained previously in **Scenario 2**.

POSSIBLE SCENARIOS FOR BINARY OUTCOMES

FOR DIFFERENCE MEASURES

1. If the data are **significantly skewed**, take caution when applying normal-based methods for estimating the Mean and Standard Deviation. **Carefully** evaluate whether these methods (Wan & Luo) are appropriate, as they **assume normality** and may lead to biased results in skewed distributions.

2. **Least-square means** (also known as Marginal Means) are derived from a linear model and are **adjusted for covariates** or confounding variables, whereas the **arithmetic mean** is **unadjusted**. While both represent continuous effect measures of the same outcome, **combining them using the estimate and SE** (calculated directly or derived from the CI) is possible.

   But keep in mind that it could be a source of heterogeneity, especially when you're pooling adjusted with unadjusted effect size data.

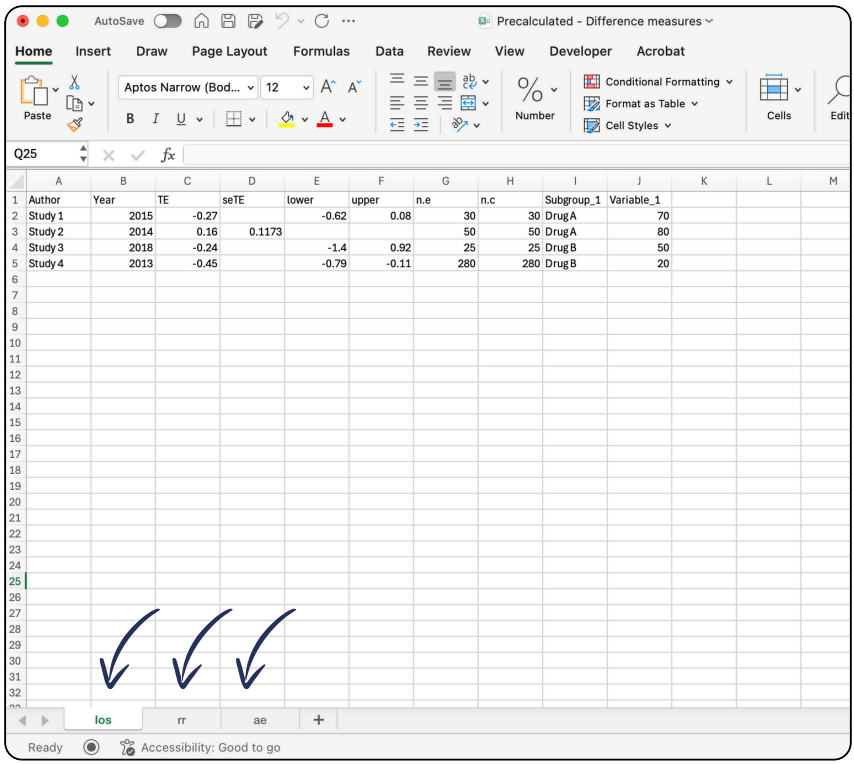
3. Always ensure all studies follow the **same direction of comparison** (e.g., Intervention 1 vs Intervention 2). This step is crucial for accurate pooling and interpretation in your meta-analysis.

V.I – Preparing your Dataset

Create a spreadsheet for each outcome of your meta-analysis in separate worksheets in your Excel workbook!

This approach saves time by allowing you to import **ALL** your endpoints at once into RStudio.

Here is an example of a dataset with 3 outcomes, each one in its own worksheet.



IV.II – Visualizing your Data

Now, we can have a look at each of the lists (our endpoints) of our data frame object.

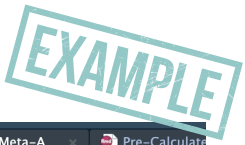
The object **ma,** we've just imported, has **3** lists of data:

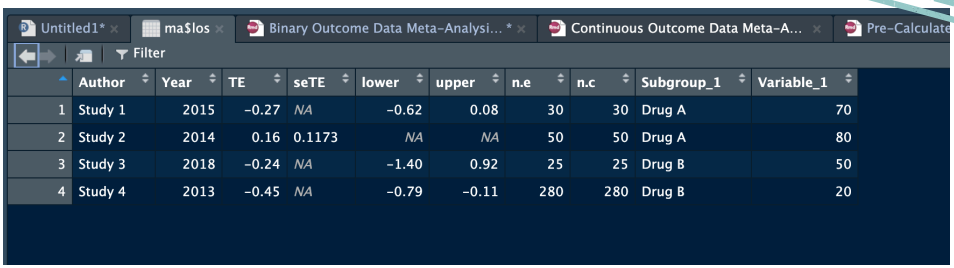
1. "**rr**" for Recurrence Rate data
2. "**ae**" for Adverse Event data
3. "**los**" for Length of Stay data

To print one specific endpoint, you can do so by adding the dollar sign operator "**$**" after "**ma**" inside de **View()** function:

```
View(ma$rr) #to visualize Recurrence Rate data
View(ma$ae) #to visualize Adverse Event data
View(ma$los) #to visualize Length of Stay data
```

This is how your endpoint dataset will appear in RStudio. See the example below:





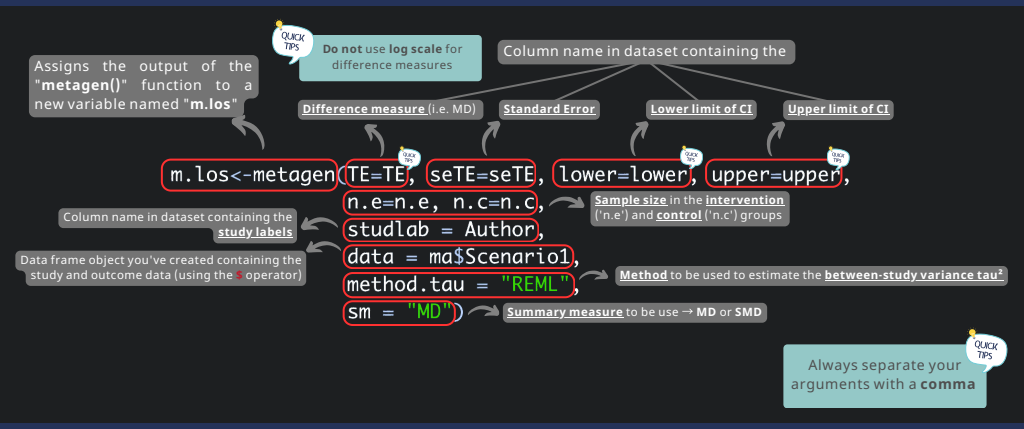
	Author	Year	TE	seTE	lower	upper	n.e	n.c	Subgroup_1	Variable_1
1	Study 1	2015	-0.27	NA	-0.62	0.08	30	30	Drug A	70
2	Study 2	2014	0.16	0.1173	NA	NA	50	50	Drug A	80
3	Study 3	2018	-0.24	NA	-1.40	0.92	25	25	Drug B	50
4	Study 4	2013	-0.45	NA	-0.79	-0.11	280	280	Drug B	20

V.III – Performing the Meta-Analysis

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

To perform the meta-analysis of pre-calculated effect size data, using **difference measures**, we will use the **metagen()** function.

We need to provide some **instructions** for the **metagen()** function. These instructions are known as **arguments**, which can be one, two, three, or more inputs that the function uses to perform its task. The main arguments of the **metagen()** function are:



The provided code using the **metagen()** function is applicable to **all three aforementioned scenarios** of available data **for continuous outcomes** because:

1. If the **Standard Error ('seTE')** is not provided, the **'upper'** (upper limit of CI) and **'lower'** (lower limit of CI) arguments will be automatically used to estimate the **'seTE'**.

2. If the **'seTE'** is included in the dataset, the function will directly use it, bypassing the need to estimate from the confidence intervals.

To perform the meta-analysis, continuous effect measures (e.g., MD, SMD) **do not require transformation into the log scale**, unlike ratio measures, which must be log-transformed.
Ensure to include **'TE=TE'**, **'lower=lower'**, and **'upper=upper'** in the dataset before running the analysis.

Over the next pages, we will detail the following aspects of binary outcome meta-analysis:
(1) methods for **pooling the studies**;
(2) methods to estimate the **between-study variance tau²**; and
(3) the **summary measures**.

MORE ON THE OTHER SIDE

V.III – Performing the Meta-Analysis

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

To visualize the results of a meta-analysis conducted with the **metagen()** function, we will use the **summary()** function. It will generate a summary of the analysis

```
summary(m.los)
```

Call here the **metagen()** object you have just created

The **summary()** function will display the estimated treatment effect (i.e., the summary measure), the corresponding confidence interval, the between-study variance, and other relevant statistics.

```
> summary(m.los)
                    MD             95%-CI %W(common) %W(random)
Study 1 -0.2700 [-0.6200;  0.0800]        22.4        28.8
Study 2  0.1600 [-0.0699;  0.3899]        51.9        34.9
Study 3 -0.2400 [-1.4000;  0.9200]         2.0         7.0
Study 4 -0.4500 [-0.7900; -0.1100]        23.7        29.3

Number of studies: k = 4
Number of observations: o = 770 (o.e = 385, o.c = 385)

                          MD            95%-CI     z p-value
Common effect model  -0.0890 [-0.2546; 0.0765] -1.05  0.2919
Random effects model -0.1704 [-0.5077; 0.1668] -0.99  0.3219

Quantifying heterogeneity:
 tau^2 = 0.0710 [0.0012; 0.8987]; tau = 0.2665 [0.0350; 0.9480]
 I^2 = 69.8% [13.1%; 89.5%]; H = 1.82 [1.07; 3.09]

Test of heterogeneity:
    Q d.f. p-value
 9.93    3  0.0192

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
```

This section contains:
(1) the individual studies;
(2) their effect sizes and confidence intervals;
(3) their weights (%) in common (i.e. fixed-) and random-effects

The **total number of studies** in our meta-analysis

The **total number of patients** in our meta-analysis

The POOLED effect size

Results concerning the **between-study heterogeneity**

Method to be used **for pooling the studies**

Method to be used to estimate the **between-study variance tau^2**

V.III – Performing the Meta-Analysis

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

METHODS FOR POOLING THE PRE-CALCULATED EFFECT SIZES MEASURES

Which method should I use?

The **metagen()** function uses the **generic inverse-variance** method for pooling the studies. This approach weights each study by the inverse of its variance, ensuring more precise studies contribute more to the overall estimate.

This method is recommended by the **Cochrane Collaboration** and is the **default** in the **metagen()** function.

You **do not need** to **specify** this method explicitly when using **metagen()**, as it is automatically applied.

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

BEST CHOICE

Restricted Maximum Likelihood ("REML")

VS

DerSimonian-Laird ("DL") estimator can be **biased**, particularly when the number of studies is **small** and **heterogeneity** is **high.**

This can lead to **overconfident** results with **narrower** confidence intervals than warranted.

False-positive results
Type 1 error

LEARNING R SYNTAX

Methods to estimate the between-study variance

```
method.tau = "REML" # Restricted maximum-likelihood estimator
method.tau = "ML" # Maximum-likelihood estimator
method.tau = "PM" # Paule-Mandel estimator
method.tau = "DL" # DerSimonian-Laird estimator
method.tau = "SJ" # Sidik-Jonkman estimator
method.tau = "HE" # Hedges estimator
method.tau = "EB" # Empirical Bayes estimator
```

Always use “ ” when referring to these methods

QUICK TIPS

METHODS TO ESTIMATE THE BETWEEN-STUDY VARIANCE

Which estimator should I use?

There are no strict guidelines determining the choice of estimator. In many cases, there are only slight variations in results among different estimators. So, you **shouldn't** worry **TOO** much about it!

If doubt arise, you can always rerun your analyses using different estimators, and check if it changes the interpretation of your results.

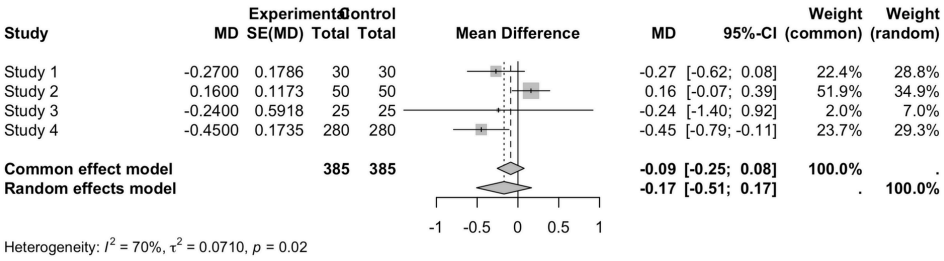
V.IV – Creating the Forest Plot

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

You can produce a forest plot for **metagen()** object using the **forest()** function from the '**meta**' package

Usually, these forest plots do not have enough quality for publication by default! In the example below, we plot the '**m.los**' object that we also used in the previous example:

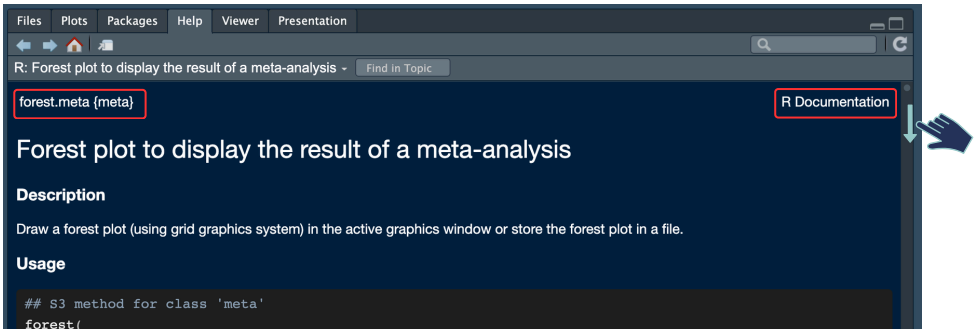
```
#Creating a forest plot for 'm.los' metagen object
forest(m.los)
```



However, the **forest()** function also has countless additional arguments to further customize the appearance of the forest plot.

All these arguments are detailed and described in the function documentation. You can accesses this documentation by running the following code:

```
help(forest)
```



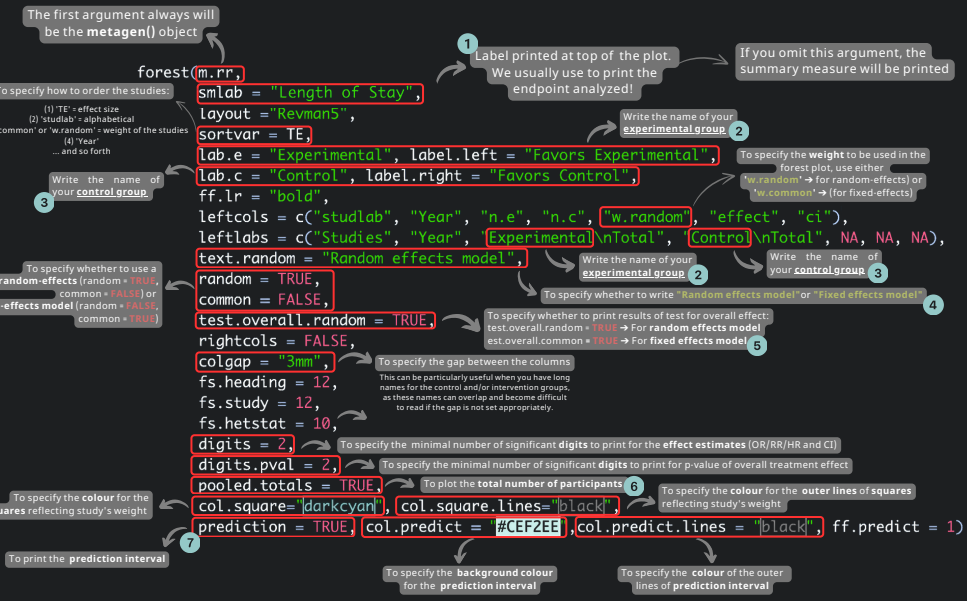
V.IV – Creating the Forest Plot

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

But don't worry. I've put together a code that brings together the main arguments for generating **elegant** and **high-quality** forest plots for publication.

Now, let's understand the most important arguments to customize your forest plot!

This is how your code looks in the end:



MORE ON THE OTHER SIDE

V.IV – Creating the Forest Plot

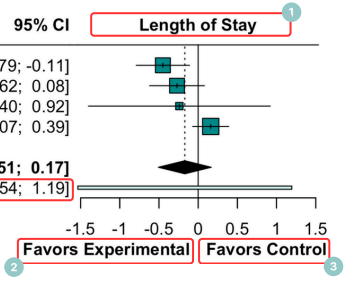
FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

This is how our forest plot looks in the end:

Studies	Year	Experimental Total	Control Total	Weight	MD	95% CI
Study 4	2013	280	280	29.3%	-0.45	[-0.79; -0.11]
Study 1	2015	30	30	28.8%	-0.27	[-0.62; 0.08]
Study 3	2018	25	25	7.0%	-0.24	[-1.40; 0.92]
Study 2	2014	50	50	34.9%	0.16	[-0.07; 0.39]
Random effects model		**385**	**385**	**100.0%**	**-0.17**	**[-0.51; 0.17]**
Prediction interval						[-1.54; 1.19]

Heterogeneity: $Tau^2 = 0.0710$; $Chi^2 = 9.93$, df = 3 (P = 0.02); $I^2 = 70\%$

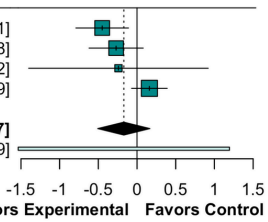
Test for overall effect: Z = -0.99 (P = 0.32)



Studies	Year	Experimental Total	Control Total	Weight	MD	95% CI
Study 4	2013	280	280	29.3%	-0.45	[-0.79; -0.11]
Study 1	2015	30	30	28.8%	-0.27	[-0.62; 0.08]
Study 3	2018	25	25	7.0%	-0.24	[-1.40; 0.92]
Study 2	2014	50	50	34.9%	0.16	[-0.07; 0.39]
Random effects model		**385**	**385**	**100.0%**	**-0.17**	**[-0.51; 0.17]**
Prediction interval						[-1.54; 1.19]

Heterogeneity: $Tau^2 = 0.0710$; $Chi^2 = 9.93$, df = 3 (P = 0.02); $I^2 = 70\%$

Test for overall effect: Z = -0.99 (P = 0.32)



V.IV – Creating the Forest Plot

FOR CONTINUOUS OUTCOMES – DIFFERENCE MEASURES

DON'T FORGET

TO **ADJUST** THE **LABELS** IN YOUR FOREST PLOT

It's critical to correctly adjust the "**Favors Experimental**" and "**Favors Control**" labels in forest plots. The interpretation of these labels depends on whether an **increase** or **decrease** in the (standardized) **mean difference** favors the experimental group. **Mislabeling** can mislead readers and **compromise the clarity** of your analysis.

SCENARIO 1 HIGHER SCORE OR VALUE = BETTER OUTCOME

When a **higher score or value** represents a **better** outcome, an **increase** in the **MD** favors the **experimental group**.

EXAMPLE

- **Quality of Life:** Higher scores indicate improved well-being.
- **Cognitive Function Tests:** Higher scores represent better performance.
- **Physical Activity Levels:** Higher values mean better results.

Always **check** the **clinical meaning** of an **increase** or **decrease** in the outcome.

SCENARIO 2 LOWER SCORE OR VALUE = BETTER OUTCOME

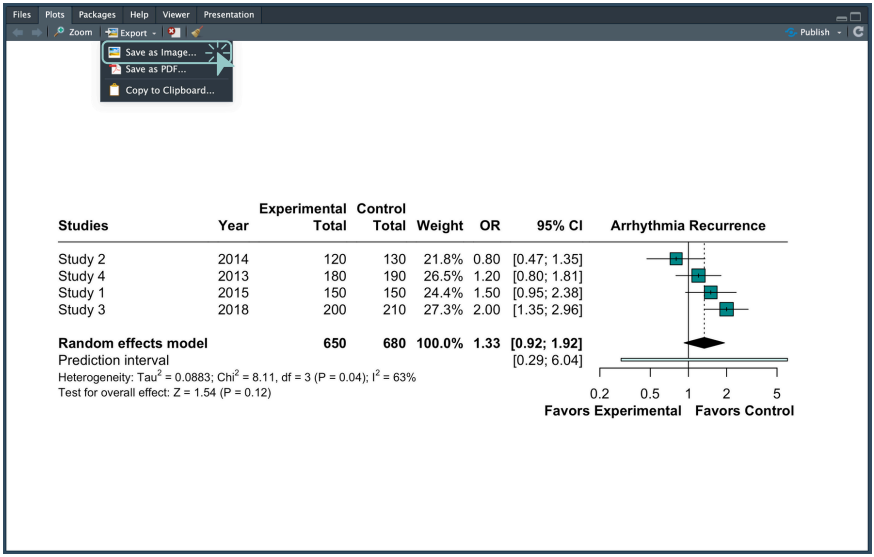
When a **lower score or value** represents a **better** outcome, a **decrease** in the **MD** favors the **experimental group**.

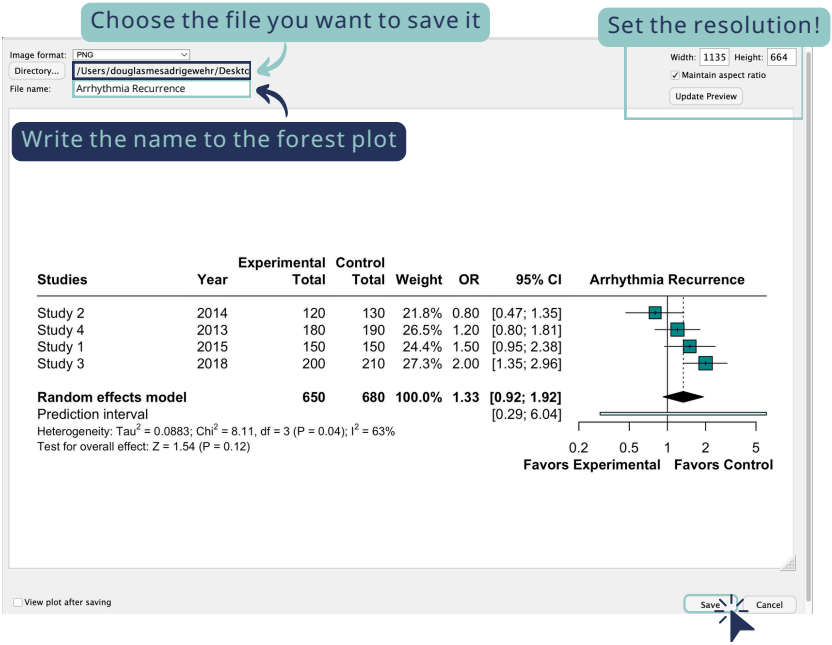
EXAMPLE

- **Blood Pressure:** Lower values reduce cardiovascular risk.
- **Cholesterol Levels:** Lower levels are beneficial.
- **Body Mass Index:** Lower BMI values indicate successful weight-loss interventions.

Always **check** the **clinical meaning** of an **increase** or **decrease** in the outcome.

THE EASIER METHOD





ADDRESSING THE DIRECTION OF COMPARISON

When combining studies, it's essential to ensure all reported treatment effects are **consistent** in their **direction of comparison**. Some studies may report effects as **Intervention 1** vs **Intervention 2**, while others use the reverse, **Intervention 2** vs **Intervention 1**. This inconsistency must be corrected before pooling data.

THE DIRECTION OF COMPARISON MATTERS !

RATIOS MEASURES = OR | RR | HR

Ratios are multiplicative measures, and inverting the comparison changes the magnitude of the effect.

EXAMPLE

RR (95% CI) for Intervention 1 vs Intervention 2 = **1.66 (0.94-2.93)**

RR (95% CI) for Intervention 2 vs Intervention 1 = **0.60 (0.34-1.06)**

DIFFERENCE MEASURES = MD | SMD

Differences are additive measures, and reversing the comparison flips the sign.

EXAMPLE

MD (95% CI) for Intervention **1** vs Intervention **2** = **1.5 (1.23; 1.77)**

MD (95% CI) for Intervention **2** vs Intervention **1** = **-1.5 (-1.23; -1.77)**

ADDRESSING THE DIRECTION OF COMPARISON

To standardize the direction of comparison in a meta-analysis, all treatment effects must represent the same comparison (e.g., Intervention 1 vs Intervention 2 or vice versa).

Here's how to adjust the direction for both ratios and differences:

RATIOS MEASURES = OR | RR | HR

$$\text{New Treatment Effect (TE)} = \frac{1}{\text{Reported Treatment Effect (TE)}}$$

$$\text{New Lower CI} = \frac{1}{\text{Reported Upper CI}}$$

$$\text{New Upper CI} = \frac{1}{\text{Reported Lower CI}}$$

Standard Error remains **unchanged**. It's independent of the direction and does not change.

DIFFERENCE MEASURES = MD | SMD

$$\text{New TE} = -\text{Reported TE}$$

$$\text{New Lower CI} = -\text{Reported Upper CI}$$

$$\text{New Upper CI} = -\text{Reported Lower CI}$$

Standard Error remains **unchanged**. It's independent of the direction and does not change.

WORKING WITH DIFFERENT CONFIDENCE INTERVALS

In meta-analyses, studies may report confidence intervals (CIs) with **different levels of confidence** (e.g., **90%**, **95%**, or **99%**). To standardize the data for analysis, you must **first calculate the Standard Error** for all studies using their reported CIs, regardless of the confidence level.
Once the **SE** is calculated, the treatment effect and its **SE** can be used uniformly in the **metagen()** function.

FROM **CONFIDENCE INTERVAL** TO **STANDARD ERROR**:

ONLY FOR STUDIES WITH >60 PATIENTS IN EACH GROUP

Ratio measures OR | RR | HR	**Difference measures** MD | SMD
From 90% CI to SE	
$SE = \frac{\log(\text{Upper CI}) - \log(\text{Lower CI})}{3.29}$	$SE = \frac{\text{Upper CI} - \text{Lower CI}}{3.29}$
`=(log(Upper_CI) - log(Lower_CI))/3.29`	`=(Upper_CI - Lower_CI)/3.29`
From 95% CI to SE	
$SE = \frac{\log(\text{Upper CI}) - \log(\text{Lower CI})}{3.92}$	$SE = \frac{\text{Upper CI} - \text{Lower CI}}{3.92}$
`=(log(Upper_CI) - log(Lower_CI))/3.92`	`=(Upper_CI - Lower_CI)/3.92`
From 99% CI to SE	
$SE = \frac{\log(\text{Upper CI}) - \log(\text{Lower CI})}{5.15}$	$SE = \frac{\text{Upper CI} - \text{Lower CI}}{5.15}$
`=(log(Upper_CI) - log(Lower_CI))/5.15`	`=(Upper_CI - Lower_CI)/5.15`

WORKING WITH DIFFERENT CONFIDENCE INTERVALS

FOR STUDIES WITH LESS THAN 60 PATIENTS IN EACH GROUP

When the sample size in each group is small (e.g., fewer than 60 participants per group), confidence intervals should be calculated using the **t-distribution instead** of the **normal distribution**. The constants 3.92 , 3.29 , and 5.15 (derived from the normal distribution) must be replaced with larger values specific to the **t distribution**, accounting for the sample size and degrees of freedom (df).

1st Step Before calculating the t statistic and subsequent SE for confidence intervals, it is essential to clearly define:

Confidence Interval

90% | 95% | 99%

+

n.e **n.c**

Sample Size in Experimental and Control Groups

2nd Step In an Excel cell, type the following formula using the variables:

`=tinv(1-CI, n.e+n.c-2)` = *t-statistics*

Where:

- **CI** is the confidence level of the interval (e.g., 0.95 for 95% CI).
- **n.e** is the sample size of the experimental group.
- **n.c** is the sample size of the control group.

EXAMPLE

MORE ON THE OTHER SIDE

3rd Step Calculate the **SE** Using the **t-Statistic**

Ratio measures

$$SE = \frac{\log(\text{Upper CI}) - \log(\text{Lower CI})}{2 \times t}$$

Difference measures

$$SE = \frac{\text{Upper CI} - \text{Lower CI}}{2 \times t}$$

WORKING WITH DIFFERENT CONFIDENCE INTERVALS

FOR STUDIES WITH LESS THAN 60 PATIENTS IN EACH GROUP

EXAMPLE 1 Ratio Measure

90% CI (1.20 - 2.50) | **n.e** = 25 | **n.c** = 22

```
=tinv(1-0.90, 25+22-2)
```

t-statistics = 1.6794

```
=(LN(2.5)-LN(1.2))/2*1.6794
```

SE = 0.6163

EXAMPLE 2 Difference Measure

99% CI (2.00 - 8.00) | **n.e** = 20 | **n.c** = 18

```
=tinv(1-0.99, 20+18-2)
```

t-statistics = 2.719

```
=(8-2)/2*2.719
```

SE = 8.157

FINALLY ...

WE NEED TO DECIDE IF WE NEED TO APPLY THE SO-CALLED

KNAPP-HARTUNG ADJUSTMENTS

What effect does it have on meta-analysis?

These adjustments affect the way the **standard error** (and thus the **confidence intervals**) of our pooled effect size is calculated.

Usually leading to **WIDER confidence intervals** of the pooled effect

How does this method do it?

Knapp-Hartung method uses quantiles of the **t-distribution** rather than the standard normal distribution in the more conventional method when computing a confidence interval (CI) for the average effect.

Why to apply this method?

The Knapp-Hartung adjustments try to **control for the uncertainty** in between-study heterogeneity → **widening** the **confidence interval** → **reducing** the chance of **false-positives**

In which situations is it most appropriate to use this method?

1) When working with random-effects model **AND**

2) We include a small number of studies our meta-analysis **AND**

3) The included studies have similar sample sizes

WARNING Extra caution is needed when there are ⩽ 5 studies of very unequal sizes

NEXT

LERN HOW TO

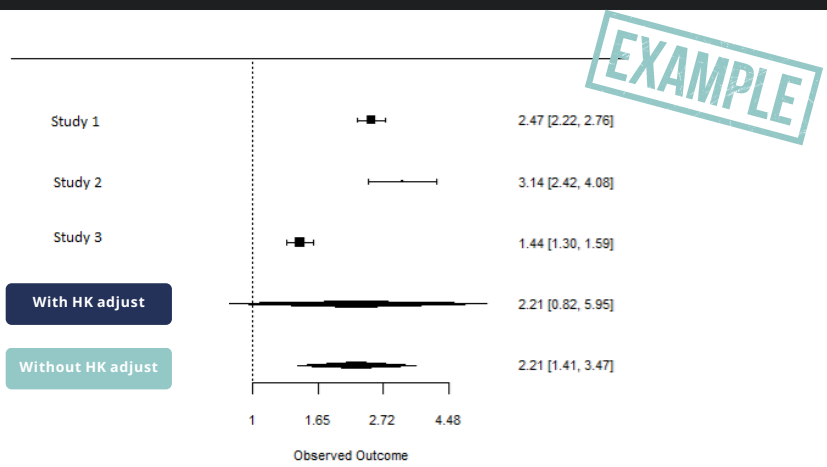
TO APPLY THE

KNAPP-HARTUNG

ADJUSTMENTS

```
m.rr<-metagen(TE=TE, seTE=seTE, lower=lower, upper=upper,
              n.e=n.e, n.c=n.c,
              studlab = Author,
              data = ma$rr,
              method.tau = "REML",
              hakn = TRUE,
              sm = "OR")
```

To apply the **Knapp-Hartung** adjustments when working with random effects model, you just need to add the argument highlighted in **red** in the **metagen()** function



VIII– Reporting the Methods

It is highly advised to specify as the aspects of the model you used in the methods section of your meta-analysis report, including:

(1) the summary measure;
(2) the method to pool the studies;
(3) the between-study variance estimator;
(4) and other details you applied in your analyses.

EXAMPLE

Statistical Analysis

We summarized pre-calculated odds ratios (OR), and mean differences (MD) using the inverse-variance random-effects model, with 95% confidence interval (CI). The restricted-maximum likelihood estimator was used to calculate the between-study variance t^2.[1] Missing means and standard deviations (SD) were estimated using the formula proposed by Wan and Luo, which derives these values from medians, interquartile ranges, and ranges, as recommended by the Cochrane Collaboration.[2-4] Pre-calculated hazard ratios (HR) and their standard errors (SE), derived from Cox proportional hazard models, were also extracted and pooled using the inverse-variance random-effects model with 95% CI. Heterogeneity was assessed using Cochrane's Q statistic and Higgins and Thompson's I^2 statistic, with $p \leq 0.10$ indicating statistical significance. I^2 values were categorized as 0%, ≤ 25%, ≤ 50%, and > 50%, reflecting no observed, low, moderate, and substantial heterogeneity, respectively.[5] All statistical tests were two-tailed, and a p-value < 0.05 was considered statistically significant. All calculations and graphics were performed using R version 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) and the "meta" extension package.[6,7]

References

1. Viechtbauer, W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics. 2005;30(3):261-293.
2. Luo D, Wan X, Liu J, Tong T (2018): Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Statistical Methods in Medical Research, 27, 1785–805.
3. Wan X, Wang W, Liu J, Tong T (2014): Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology, 14, 135.
4. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane, 2024. Available from www.training.cochrane.org/handbook.
5. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine.* 2002;21(11):1539-1558.
6. Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
7. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health.* 22nd ed.; 2019:153-160.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Between-study variance estimators

Langan D, Higgins JP, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Research Synthesis Methods. 2005;10(1):83–98.

Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Research Synthesis Methods.2016;7(1):55–79.

Paule-Mandel estimator:
Paule RC, Mandel J. Consensus Values and Weighting Factors. Journal of Research of the National Bureau of Standards.1982;87(5):377-385.

Restricted maximum likelihood estimator:
Viechtbauer, W. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. Journal of Educational and Behavioral Statistics. 2005;30(3):261-293.

Sidik-Jonkman estimator:
Sidik K, Jonkman JN. Simple Heterogeneity Variance Estimation for Meta-Analysis. Journal of the Royal Statistical Society Series C: Applied Statistics. 2005;54(2):367–384.

DerSimonian-Laird estimator:
DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials. 1986;7(3):177–188.

Heterogeneity Assessment

Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 2002;21(11):1539-1558.

Methods for Estimating Means and SD

Luo D, Wan X, Liu J, Tong T (2018): Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Statistical Methods in Medical Research, 27, 1785–805

Wan X, Wang W, Liu J, Tong T (2014): Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Medical Research Methodology, 14, 135

Time-to-Event Analysis

Tierney, J.F., Stewart, L.A., Ghersi, D. et al. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials 8, 16 (2007). https://doi.org/10.1186/1745-6215-8-16

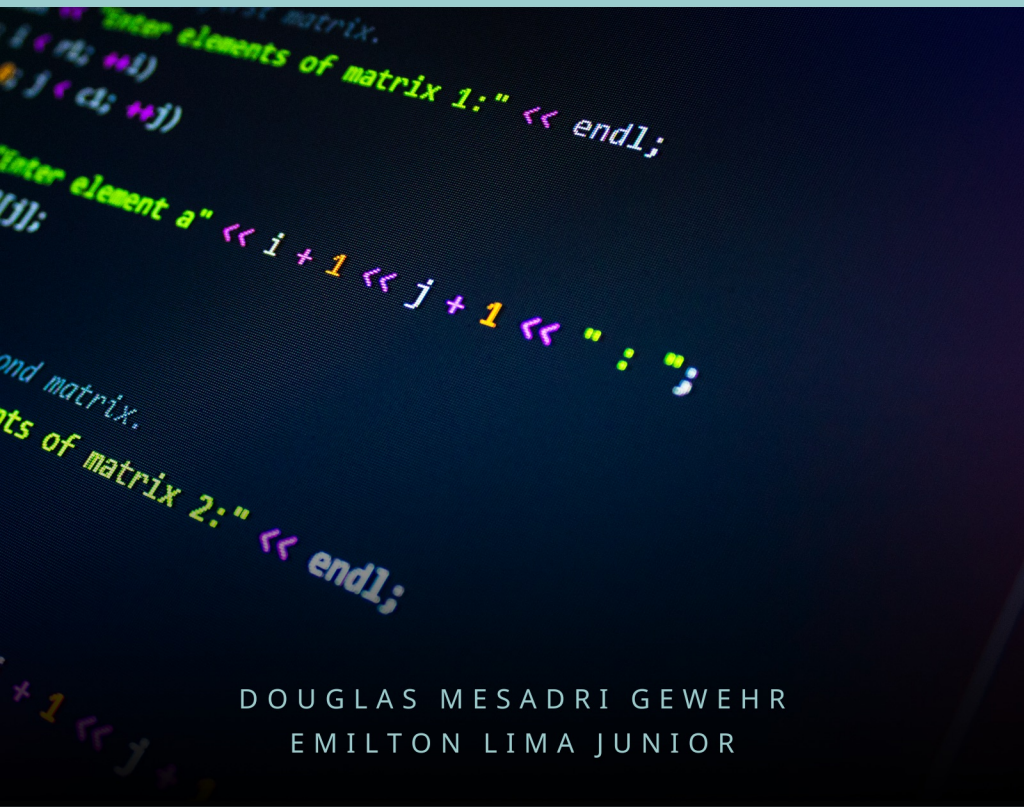
R, RStudio, and R packages

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

meta package:
Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. In: *Evidence-Based Mental Health*. 22nd ed.; 2019:153-160.

readxl package:
Wickham H, Bryan J (2023). readxl: Read Excel Files. https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.



DOUGLAS MESADRI GEWEHR
EMILTON LIMA JUNIOR