Introduction to Meta-Analysis with R

KITE-Trainees Workshop Series

Mohammadreza Amiri

2021-10-26

Contents

1 Making Everything Ready

Visit this link to find out how to install R, RStudio IDE, and install R packages.

1.1 R

R is a free software environment for statistical computing and graphics. R software can be downloaded here. The most recent version of R is 4.1.1 (Kick Things) and has been released on 2021-08-10. However, some packages that I frequently use do not support the newest version of R. Facilitating this issue, I use version 3.6.3 on my system. Nonetheless, you are able to install and use multiple version of R on your system and using RStudio you can switch back and forth to any version of R based on your needs. To get help using R access their webpage here.

1.2 RStudio

RStudio's mission is to create free and open-source software for data science, scientific research, and technical communication. Inspired by innovators in science, education, government, and industry, RStudio develops free and open tools for R, and enterprise-ready professional products for teams who use both R and Python, to scale and share their work. RStudio Integrated development environment (IDE) can be downloaded here. We are using RStudio for this workshop.

2 What is a meta-analysis (MA)?

2.1 Definition

The following definitions summarize the MA statistical technique:

> "Analysis of analyses" (Glass, 1976)

> "Meta-analysis can be understood as a form of survey research in which research reports, rather than people, are surveyed." (Lipsey & Wilson, 2001)

3 The Best packages to Conduct a MA

In my opinion, the best packages are meta (for ease of use and has a book) and metafor for best documentation (after installation type `vignette("metafor")`).

4 Data Preparation

Loading required packages (don't forget to install the packages first if you have not done so):

Let's view the data for this session:

```
dat <- dat.bcg

knitr::kable(dat)
```

trial	author	year	tpos	tneg	cpos	cneg	ablat	alloc
1	Aronson	1948	4	119	11	128	44	random
2	Ferguson & Simes	1949	6	300	29	274	55	random
3	Rosenthal et al	1960	3	228	11	209	42	random
4	Hart & Sutherland	1977	62	13536	248	12619	52	random
5	Frimodt-Moller et al	1973	33	5036	47	5761	13	alternate
6	Stein & Aronson	1953	180	1361	372	1079	44	alternate
7	Vandiviere et al	1973	8	2537	10	619	19	random
8	TPT Madras	1980	505	87886	499	87892	13	random
9	Coetzee & Berjak	1968	29	7470	45	7232	27	random
10	Rosenthal et al	1961	17	1699	65	1600	42	systematic
11	Comstock et al	1974	186	50448	141	27197	18	systematic
12	Comstock & Webster	1969	5	2493	3	2338	33	systematic
13	Comstock et al	1976	27	16886	29	17825	33	systematic

Now we need to calculate the log risk ratios and corresponding sampling variances (and use the 'slab' argument to store study labels as part of the data frame):

```
ex1data <- escalc(measure = "RR", ai = tpos, bi = tneg, ci = cpos,
    di = cneg, data = dat, slab = paste(author, year, sep = ", "))

# set_flextable_defaults( font.family = 'Arial', font.size
# = 10, font.color = 'black', table.layout = 'autofit',
# digits = 3, theme_fun = '' )

ex1data %>%
    select(trial, author, year, yi, vi) %>%
    head(5) %>%
    flextable() %>%
    set_caption("Effect size and variance") %>%
    set_table_properties(layout = "autofit") %>%
    colformat_double(digits = 3) %>%
    theme_booktabs()
```

Table 1: Effect size and variance

trial	author	year	yi	vi
1	Aronson	1,948	-0.889	0.326
2	Ferguson & Simes	1,949	-1.585	0.195
3	Rosenthal et al	1,960	-1.348	0.415
4	Hart & Sutherland	1,977	-1.442	0.020
5	Frimodt-Moller et al	1,973	-0.218	0.051

Effect size (yi): Log(RR)

Variance of effect size (*vi*): Variance Log(RR)

Note (Viechtbauer, 2010):

In various fields (such as the health and medical sciences), the response variable measured is often dichotomous (binary), so that the data from a study comparing two different groups can be expressed in terms of a table, such as:

Table 2: Example of dichotomous outcomes

	Outcome 1	Outcome 2	Total
Group 1	ai	bi	n1i
Group 2	ci	di	n2i

where ai, bi, ci, and di denote the cell frequencies (i.e., the number of people falling into a particular category) and n1i and n2i are the row totals (i.e., the group sizes).

For example, in a set of randomized clinical trials, group 1 and group 2 may refer to the treatment and placebo/control group, respectively, with outcome 1 denoting some event of interest (e.g., death, complications, failure to improve under the treatment) and outcome 2 its complement. Similarly, in a set of cohort studies, group 1 and group 2 may denote those who engage in and those who do not engage in a potentially harmful behavior (e.g., smoking), with outcome 1 denoting the development of a particular disease (e.g., lung cancer) during the follow-up period. Finally, in a set of case-control studies, group 1 and group 2 may refer to those with the disease (i.e., cases) and those free of the disease (i.e., controls), with outcome 1 denoting, for example, exposure to some environmental risk factor in the past and outcome 2 non-exposure. Note that in all of these examples, the stratified sampling scheme fixes the row totals (i.e., the group sizes) by design.

A meta-analysis of studies reporting results in terms of tables can be based on one of several different outcome measures, including the risk ratio (also called the relative risk), the odds ratio, the risk difference, and the arcsine square root transformed risk difference (Rücker et al., 2009). For any of these outcome measures, one needs to specify the cell frequencies via the ai, bi, ci, and di arguments (or alternatively, one can use the ai, ci, n1i, and n2i arguments).

The options for the measure argument are then:

"RR" for the log risk ratio,

"OR" for the log odds ratio,

"RD" for the risk difference,

"AS" for the arcsine square root transformed risk difference (Rücker et al., 2009),

"PETO" for the log odds ratio estimated with Peto's method (Yusuf et al., 1985).

Note that the log is taken of the risk ratio and the odds ratio, which makes these outcome measures symmetric around 0 and yields corresponding sampling distributions that are closer to normality. Also, when multiplied by 2, the arcsine square root transformed risk difference is actually identical to Cohen's h (Cohen, 1977).

5 Hands-on practice

5.1 Forest Plot Using metafor Package

5.1.1 Unordered Forest Plot

```
# Fitting a random-effects model
res <- rma(yi, vi, data = ex1data)


# Forest plot combined with (necessary) annotations
forest(res, atransf = exp, at = log(c(0.05, 0.25, 1, 4)), xlim = c(-16,
    6), ilab = cbind(dat.bcg$tpos, dat.bcg$tneg, dat.bcg$cpos,
    dat.bcg$cneg), ilab.xpos = c(-9.5, -8, -6, -4.5), cex = 0.75,
    header = "Author(s) and Year", mlab = "")

op <- par(cex = 0.75, font = 2)

text(c(-9.5, -8, -6, -4.5), 15, c("TB+", "TB-", "TB+", "TB-"))

text(c(-8.75, -5.25), 16, c("Vaccinated", "Control"))

par(op)


# add text with Q-value, dfs, p-value, and I^2 statistic
text(-16, -1, pos = 4, cex = 0.75, bquote(paste("RE Model (Q = ",
    .(formatC(res$QE, digits = 2, format = "f")), ", df = ",
    .(res$k - res$p), ", p = ", .(formatC(res$QEp, digits = 2,
        format = "f")), "; ", I^2, " = ", .(formatC(res$I2, digits = 1,
        format = "f")), "%)")))
```

5.1.2 Ordered Forest Plot

```
# Fitting a random-effects model
res <- rma(yi, vi, data = ex1data)


# Forest plot combined with (necessary) annotations
forest(res, atransf = exp, at = log(c(0.05, 0.25, 1, 4)), xlim = c(-16,
    6), ilab = cbind(dat.bcg$tpos, dat.bcg$tneg, dat.bcg$cpos,
    dat.bcg$cneg), ilab.xpos = c(-9.5, -8, -6, -4.5), cex = 0.75,
    header = "Author(s) and Year", mlab = "", order = "obs")
```

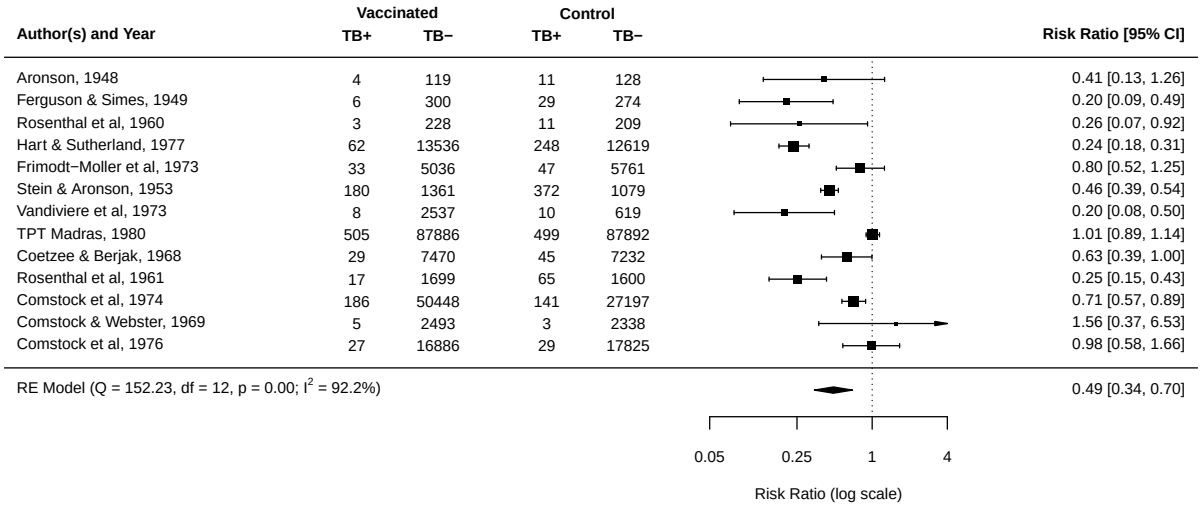

Figure 1: Unordered Forest Plot

```
op <- par(cex = 0.75, font = 2)

text(c(-9.5, -8, -6, -4.5), 15, c("TB+", "TB-", "TB+", "TB-"))

text(c(-8.75, -5.25), 16, c("Vaccinated", "Control"))

par(op)


# add text with Q-value, dfs, p-value, and I^2 statistic
text(-16, -1, pos = 4, cex = 0.75, bquote(paste("RE Model (Q = ",
    .(formatC(res$QE, digits = 2, format = "f")), ", df = ",
    .(res$k - res$p), ", p = ", .(formatC(res$QEp, digits = 2,
        format = "f")), "; ", I^2, " = ", .(formatC(res$I2, digits = 1,
        format = "f")), "%)")))
```

5.2 Forest Plot Using meta Package

```
nfr <- read.xlsx(xlsxFile = "NFR_v_Control_final.xlsx")

nfrModel <- metacont(data = nfr, n.e = ss1, mean.e = m1, sd.e = sd1,
    n.c = ss2, mean.c = m2, sd.c = sd2, studlab = Study.Label,
    sm = "SMD")

forest(x = nfrModel, comb.random = F, digits.mean = 1, digits.sd = 1)
```

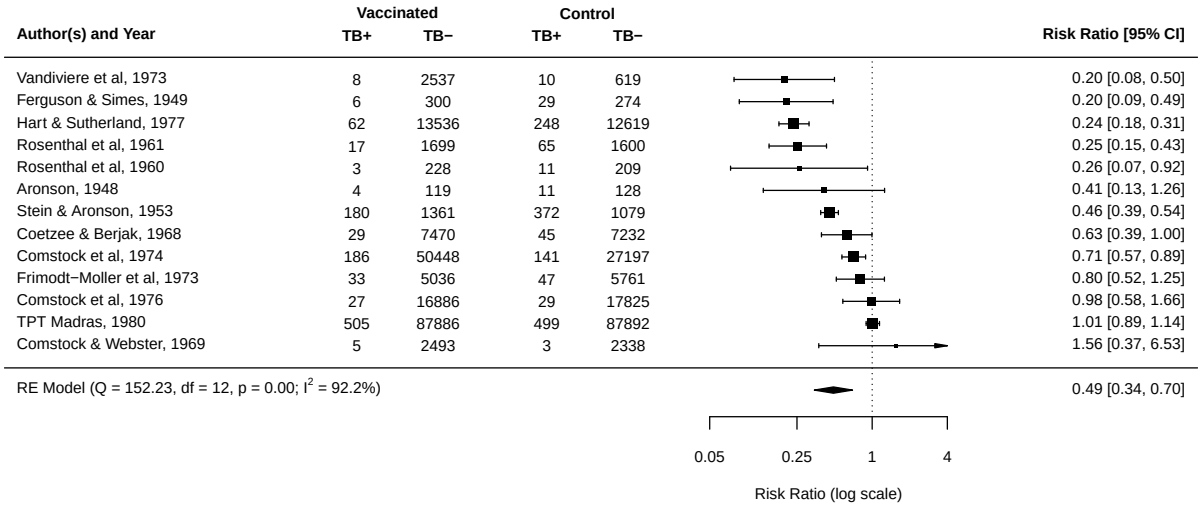

Figure 2: Ordered Forest Plot

5.3 Funnel Plot Using metafor Package

A funnel plot shows the observed effect sizes or outcomes on the x-axis against a precision measure of the observed effect sizes on the y-axis. In metafor package, the recommended choice for the y-axis is the standard error (in decreasing order) which is in line with Sterne & Egger (2001). If there is no publication bias or heterogeneity among studies, the expectation is to see all points to fall within the pseudo-confidence shape funnel, i.e., the triangle shape funnel for the case of standard errors on the y-axis.

```
funnel(res, main = "Standard Error")
```

```
funnel(res, yaxis = "seinv", main = "Inverse Standard Error")
```

5.4 Funnel Plot Using meta Package

```
nfr <- read.xlsx(xlsxFile = "NFR_v_Control_final.xlsx")

nfrModel <- metacont(data = nfr, n.e = ss1, mean.e = m1, sd.e = sd1,
    n.c = ss2, mean.c = m2, sd.c = sd2, studlab = Study.Label,
    sm = "SMD")
```

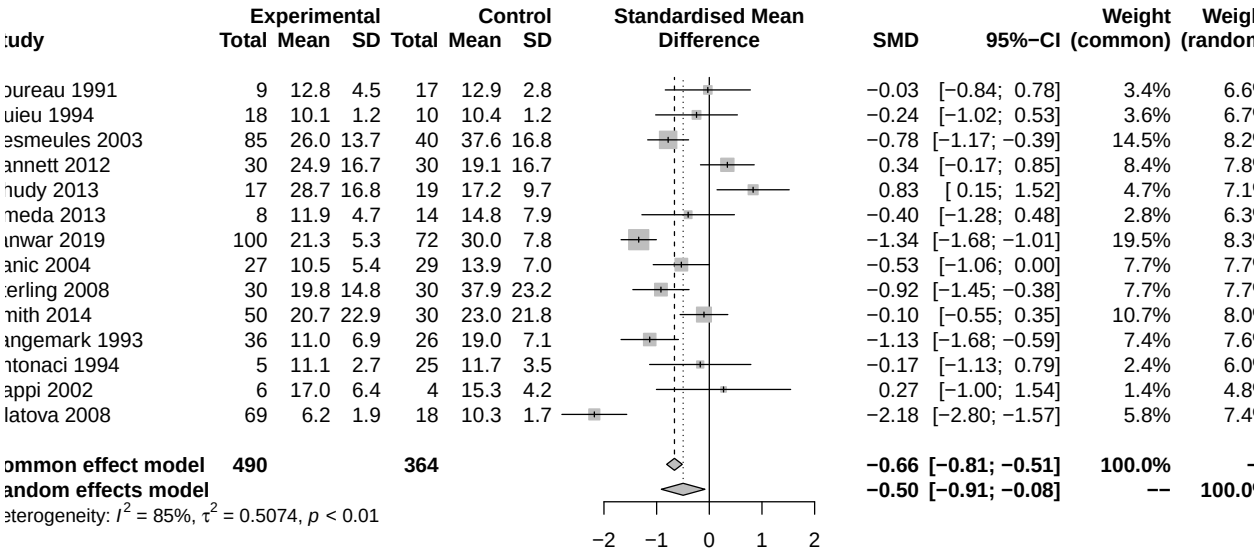

Figure 3: Forest Plot Using meta Package

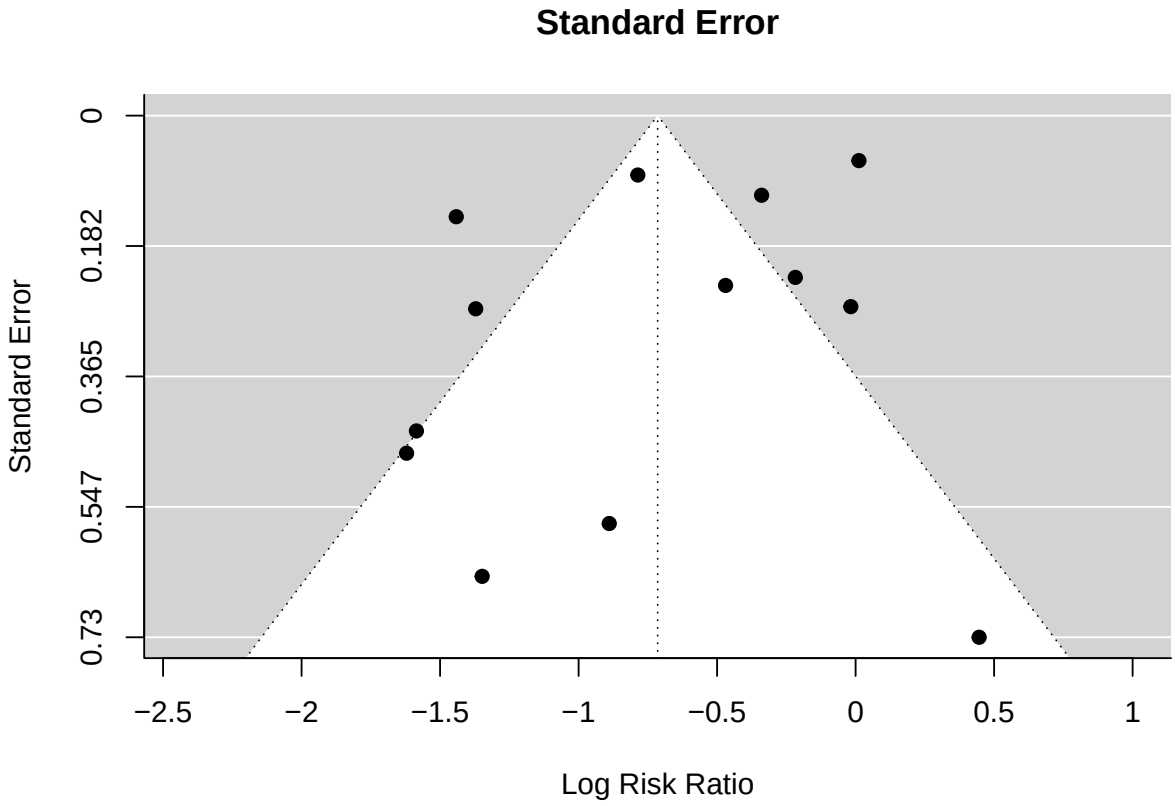

Figure 4: Common Funnel Plots

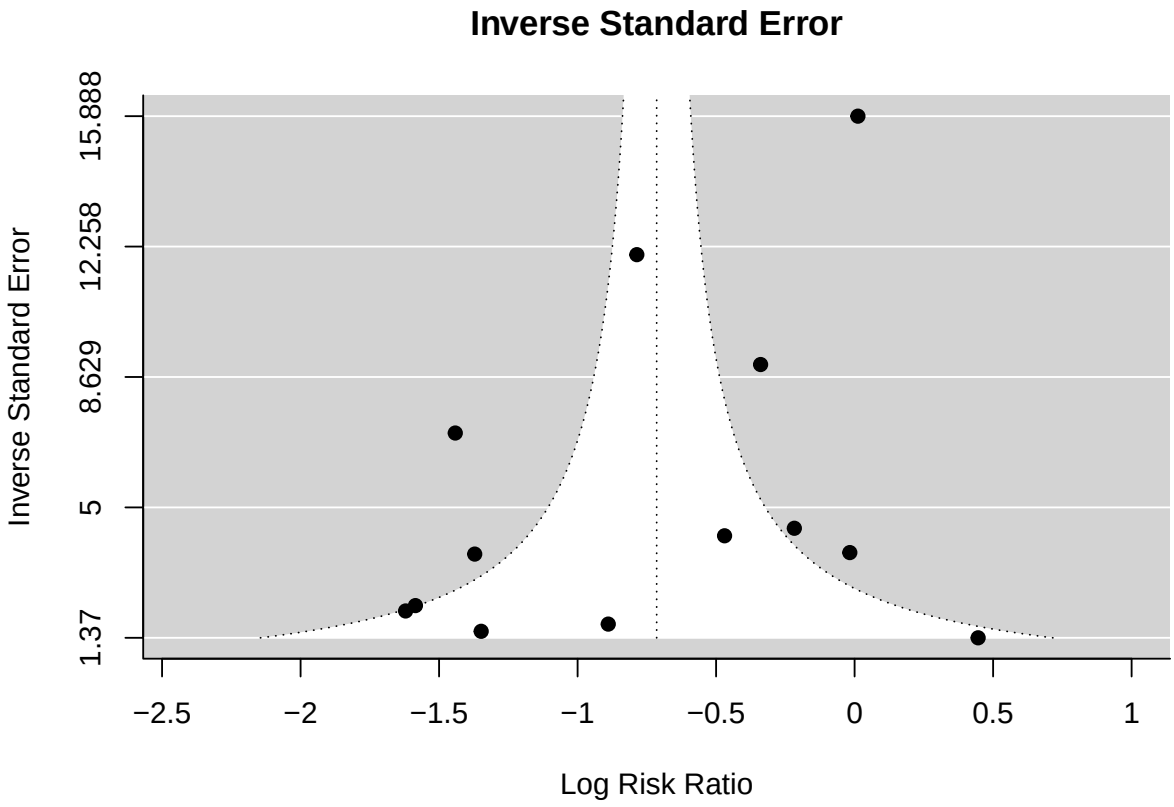

Figure 5: Common Funnel Plots

```
meta::funnel.meta(nfrModel, contour.levels = c(0.9, 0.95, 0.99),
    yaxis = "invvar")
```

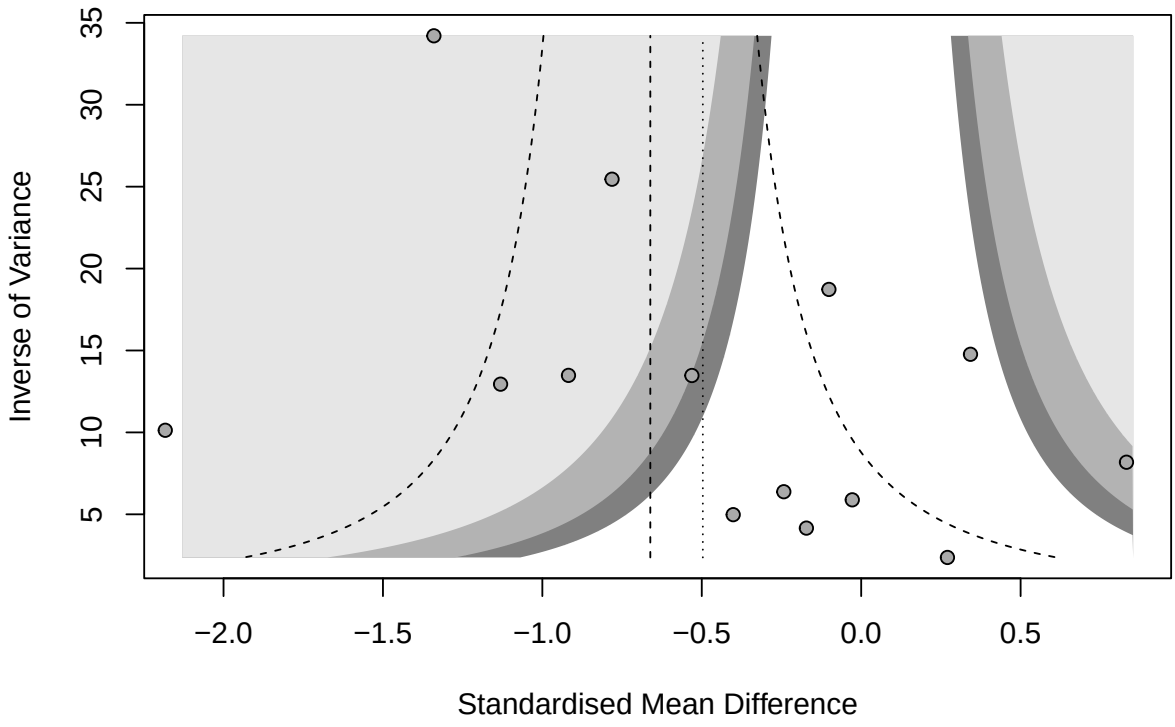

Figure 6: Common Funnel Plots Using meta Package

References

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Elsevier. https://doi.org/10.1016/c2013-0-10517-x

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.3102/0013189x005010003

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Rücker, G., Schwarzer, G., Carpenter, J., & Olkin, I. (2009). Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Statistics in Medicine*, *28*(5), 721–738. https://doi.org/10.1002/sim.3511

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055. https://doi.org/10.1016/s0895-4356(01)00377-8

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Yusuf, S., Peto, R., Lewis, J., Collins, R., & Sleight, P. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases*, *27*(5), 335–371. https://doi.org/10.1016/s0033-0620(85)80003-7