An Introduction to Meta-Analysis With R

Calum MacGillivray, PhD Student (Education), University of Dundee

c.y.macgillivray@dundee.ac.uk | University Profile | ORCiD

University of Dundee

Scottish Graduate School of Social Science
Sgoil Cheumnaichean Saidheans Sòisealta na h-Alba

Housekeeping

- Breaks: we will take a short break every hour and an hour for lunch from 12:00-13:00
- In the morning we will be talking about the theory behind meta-analysis and getting started with R
- In the afternoon we will get more hands on with the code
- Questions: we will stop regularly for questions but if you have any as we go then feel free to jump in, raise your hand or put them in the chat.
- Afterwards: feel free to email me with any questions: c.y.macgillivray@dundee.ac.uk
- Resources for this workshop, including the notebook with the annotated code, data files, further reading and this presentation can be found on my website here: https://calummacgillivray.github.io/Training/

A Quick Note on Jupyter Notebooks and Colab

- We will be working from a Jupyter Notebook hosted by Google Colab
- Jupyter Notebooks are great for sharing code as you can write in legible text around the chunks of code. This makes it easier to follow.
  - When aesthetic legibility is less of a concern I tend to work directly in Rstudio. The code works the same regardless.
- Google Colab allows us to run R code without having R installed on the browser which is very helpful and prevents technical issues
  - You just need a Google account
- You can find the interactive notebook with the code here: https://colab.research.google.com/github/calummacgillivray/calummacgillivray.github.io/blob/main/Meta_Analysis_jupyter.ipynb

Introductions

Who am I?

Name: Calum MacGillivray

University: University of Dundee, School of Humanities, Social Science and Law

Discipline: Education (with a developmental psychology background)

PhD Project: Primary-secondary transitions experiences and associated educational outcomes

Now who are you?

What’s your name? | Your discipline/background? | What are you working on?

What do you know already?

- What is meta-analysis?
- Why might you conduct it?
- Have you done any?
- What do you hope to get out of today?

Learning Outcomes

Feel more confident in understanding and conducting meta-analysis by:

- Understanding the basics of how meta-analysis works and when to conduct it
- Understanding how effect sizes are calculated and the data you require for meta-analysis
- Building knowledge of various methods in R to conduct meta-analysis
- Learning how to build a forest plot to visualise meta-analyses with R

A Note on Terminology

- Literature Review
- Systematic Review
- Rapid Review
- Review of Reviews/Umbrella Review
- Mapping review
- Scoping Review
- Meta-synthesis
  - **Meta analysis**
  - Narrative synthesis

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

An Anatomy of a Systematic review

Designing the Review

Research Questions	Methodology	Search Strategy	Scoping	Protocol	The Team

↓

Searching

Databases	Journals	Grey Literature	Citation Searching	Websites	Emails

↓

Screening

Titles	Abstracts	Full texts	(Dis)agreement

↓

Data Extraction

Coding	Critical Appraisal (Risk of Bias)

↓

Analysis/Meta-synthesis

Narrative Synthesis	Meta Analysis

A brief (and incomplete) history

- 1975: Meta-Analysis (psychotherapy)
- Archie Cochrane: evidence based medicine
- Cochrane Systematic Reviews: health care intervention efficacy
- Cambell Collaboration: public policy scope
- EPPI Centre: education and welfare
- QUOROM, then PRISMA: reporting guidelines

https://eppi.ioe.ac.uk/cms/Resources/EvidenceInformedPolicyandPractice/HistoryofSystematicReviews/tabid/68/Default.aspx

Addressing the Replication Crisis

2011 – a bad year for psychology

- Precognition – dubiously published
- Diedrik Stapel – fabricating data
- Using common methods to argue that listening to a song could reduce a participants age – to prove a point
- Name-letter effect – failed to replicate
- Many studies suffer from large amounts of unrecognised bias

The result: “Crisis of Confidence”

But this is not limited to psychology

- Dubious practice exists across scientific disciplines

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology, 39*(4), 202–217. https://doi.org/10.1037/teo0000137

Pashler, H., & Wagenmakers, E. (2012). Editors’ Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, *7*(6), 528-530. https://doi.org/10.1177/1745691612465253

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*, 452–454. https://doi.org/10.1038/533452a

Bias

- Bias from incorrectly applied (or understood) methodology
  - Failure to address confounding factors
  - Faulty metrics
  - Participant selection
  - Missing data
  - Misinterpretation
  - Congruity between philosophical, theoretical, and methodological approach
- Publication bias (the file drawer problem)

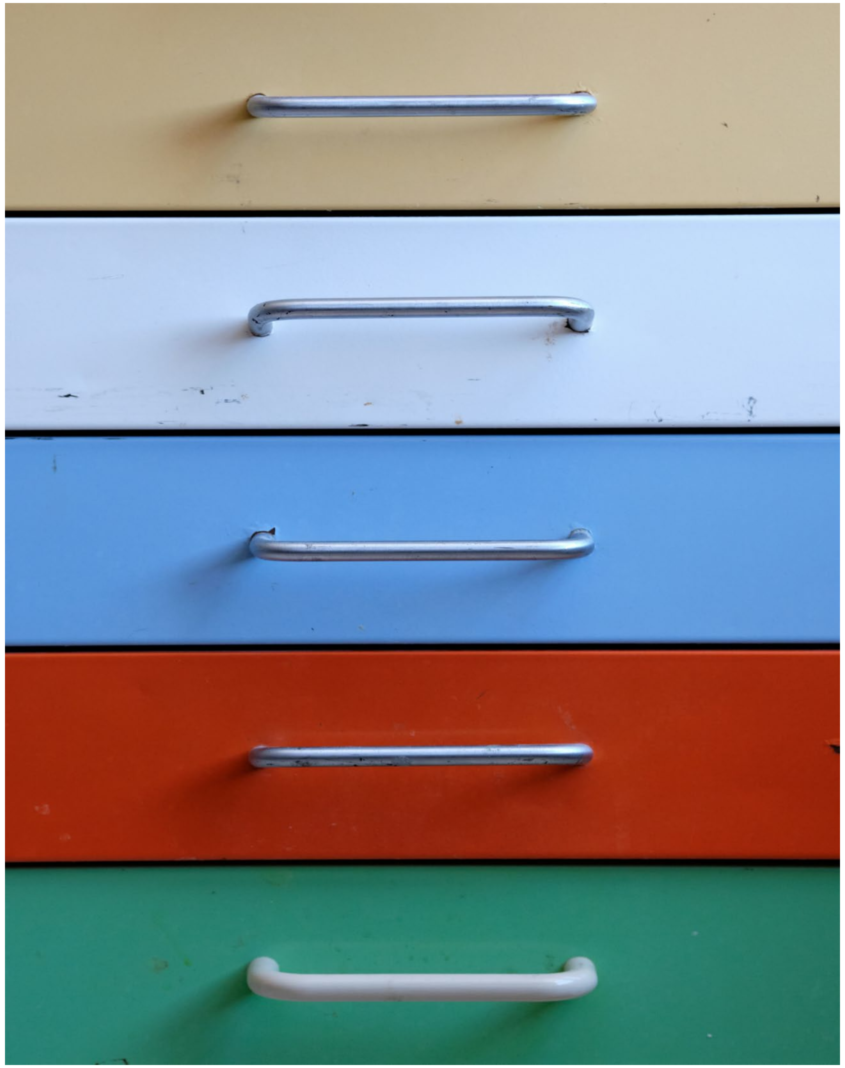


Photo by Drew Beamer on Unsplash

Addressing the Replication Crisis

- Systematic reviews pre-date the replication crisis
  - Rubbish in > Rubbish out
- But can begin to address the file drawer problem
- Can also make informed decisions on what to include
  - Exclusion
  - Weighting
- Can be a tool for systematic, rigorous critique
- Can highlight the methodological shortcomings/strengths in the literature
- Can assess risk of bias

Photo by Evan Demicoli on Unsplash

Synthesis of the literature

- Bringing together findings
- Creating a new perspective
- Can cover any methodology
  - Quantitative
  - Qualitative
  - Mixed Methods
- Narrative Synthesis
- Meta Analysis
- Carefully consider what you are bringing together
- Don't accidentally equate apples with oranges!

Photo by Gowtham AGM on Unsplash

But this is not a systematic review workshop!

Meta-Analysis as Part of a Systematic Review

- The point of meta-analysis is usually to pool effect sizes across different studies with the aim of investigating a common effect
- Through combining the rigorously gathered data, you gain the benefit of combining the work across many studies

- Problem 1: The Completeness of the Data
- Problem 2: The File Drawer Problem
- Problem 3: Comparing Apples and Oranges
- Problem 4: Rubbish In -> Rubbish Out

- Solutions?
  - Systematic approach, rigour and critical appraisal
  - Scoping: sometimes meta-analysis is not the right choice
    - The "I have a hammer so every problem looks like a nail" problem

Photo by Wisnu Amaludin on Unsplash

Effect sizes

- Meta-analysis pools effect sizes
- An effect size is a measure of the magnitude of a relationship
- A p-value can suggest if a relationship is unlikely to be random, supporting evidence to refute a null hypothesis – however it says nothing about the scall of that relationship
    - That is where the effect size comes in
    - A classic example is Cohen's *d*
    - We can approximate from this value the magnitude of the effect
        - A rough reckoner is:
            - 0.2 = small effect
            - 0.5 = medium effect
            - 0.8 = large effect
    - From one study we would call this the observed effect
    - We will look at calculating different effect sizes later, although often we only need the data for calculating the effect sizes rather than the effect sizes themselves

Fixed Effects Meta-Analysis

- For fixed effects meta-analysis we take the stance that the observed effect ($\hat{\theta}_k$) from each study differs from the true effect (θ) because of sampling error (ϵk): $\hat{\theta}_k = \theta + \epsilon k$
  - True effect = the actual underlying effect across samples i.e. the thing the meta-analysis intends to estimate
  - Observed effect = an effect size from a single study
  - Sampling error = the unrepresentativeness of our sample

Fixed Effects Meta-Analysis

- The lower the standard error of an effect size the more precise it is at estimating the effect size, this is directly related to sample size.
  - As such to pool effects across studies we weight them by standard error
  - To calculate the weight for each study (w_k) we divide 1 by the variance (standard error squared; s_k^2): $w_k = \frac{1}{s_k^2}$
- Our final model then takes the sum of each effect size multiplied by its weight: $\sum_{k=1}^{K} \hat{\theta}_k w_k$
- Then we divide that by the sum of all the weights: $\sum_{k=1}^{K} w_k$
- Altogether this looks like: $\frac{\sum_{k=1}^{K} \hat{\theta}_k w_k}{\sum_{k=1}^{K} w_k}$

Fixed Effects Meta-Analysis

- Technically speaking we ran through the inverse variance approach, however when dealing with sparse binary data we may want to weight differently by using either:
  - The Mantel-Haenszel approach
  - The Peto Approach
  - Bakbergenuly's method
- You can find out more here: https://doing-meta.guide/pooling-es#binary-outcomes

Random Effects Meta-Analysis

- But what if there is more causing differences between studies than just sampling error?
- What if outcomes are slighty measured differently?
- Or the intended population varies?
- Or an intervention is delivered in a slighty different way?
- Or any other numbers of potentially biasing factors is in play?
- This means that the true effect that each study is trying to measure will be slightly different.
  - This difference is known as between study heterogeneity
- To account for this we need a more complex model than: $\hat{\theta}_k = \theta + \epsilon k$

Random Effects Meta-Analysis

- In random effects meta-analysis we no longer assume each study differs from one true effect by sampling error.
- Rather we anticipate a distribution of many true effects, the mean of which is denoted as μ.
- Additionally as well as sampling error (ϵ_k) we also anticipate error that deviates from the mean of the distribution of effects (ζ_k).
- This gives us the equation for our true effect of interest: $\hat{\theta}_k = \mu + \zeta_k + \epsilon_k$
- A key assumption here is that the source of error (ζ_k) is entirely random.

Random Effects Meta-Analysis

- Our practical equation is similar to fixed effects however we use adjusted weights (denoted by asterisks): $\frac{\sum_{k=1}^{K} \hat{\theta}_k \, w_k^*}{\sum_{k=1}^{K} w_k^*}$
- We calculate these adjusted weights through the addition of Tau-squared (τ^2) to the standard error squared (s_k^2) which we then divide 1 by: $w_k^* = \frac{1}{s_k^2 + \tau^2}$
- Therefore the key difference between fixed and random effects meta-analysis is the calculation of Tau-squared.

Tau-Squared

- There are many ways to estimate tau-squared and a decision will need to be made, this is worth reading into in depth:
  - The most common is by DerSimonian and Laird but may be more biased when there are few studies with high heterogeneity (Hartung and Knapp, 2001).
  - There is one by Paule and Mandel which performs well for binary and continuous data at low study numbers (Bakbergenuly et al., 2020).
  - The restricted maximum likelihood approach is also best suited to continuous outcomes (Veroniki et al., 2016).
  - Sidik and Jonkman provided an approach that is best suited when tau-squared is high.

Knapp-Hartung Adjustment

- The Knapp-Hartung adjustment can be used with random effects meta-analysis
  - This usually increases the confidence intervals and reduces the risk of a false positive.
  - This is usually done, but not always.

Random or Fixed Effects?

- The assumptions of the fixed effects model are difficult to meet in most situations, as was outlined previously.
- Therefore in many cases researchers, especially in the social sciences, decide to conduct random effects meta-analyses.
- However, there is an argument that random-effects meta-analyses ca, in some cases, overemphasise small sample studies which may be more likely to include more biases (Poole and Greenland, 1999).

Assessing Heterogeneity

- Cochrane's *Q* – a chi-square test
  - to detect heterogeneity – on significant p-value
- Higgins and Thompson's I^2 – to quantify heterogeneity
- Tau-squared – forms the basis of random effects analyses

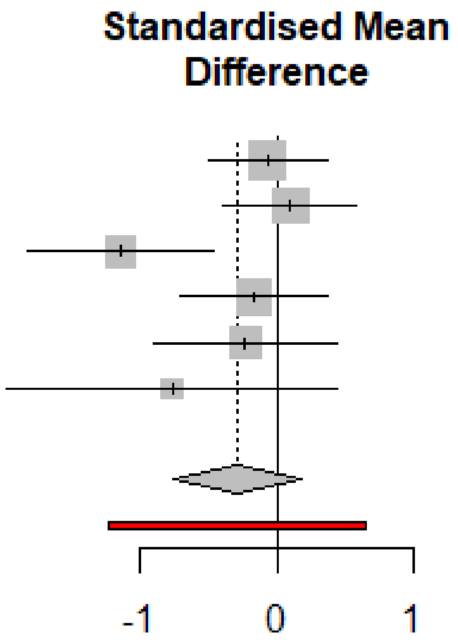
- According to the Cochrane handbook:
- Cochrane's Q is prone to bias at small study numbers, or sample sizes
  - A non-significant finding does not mean heterogeneity can be ruled out
  - At the extreme end – high numbers of studies can lead to unimpactful heterogeneity being detected
  - 0.1 (rather than 0.05) is often used as the threshold for this test
- For I^2
  - 0% - 40%: possibly less important
  - 30% - 60%: possibly moderate
  - 50% - 90%: possibly substantial
  - 75% - 100%: possibly considerable

https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-10

Forest Plots – Example Data

Simulating 6 studies comparing an outcome at T2 after an intervention for an experimental group vs a control group

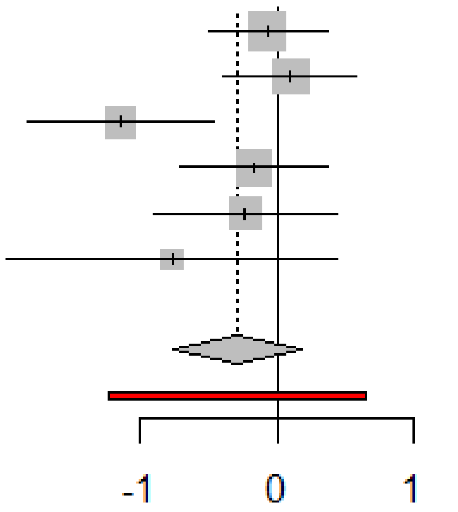
Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	



Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$

Forest Plots – Measures

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	



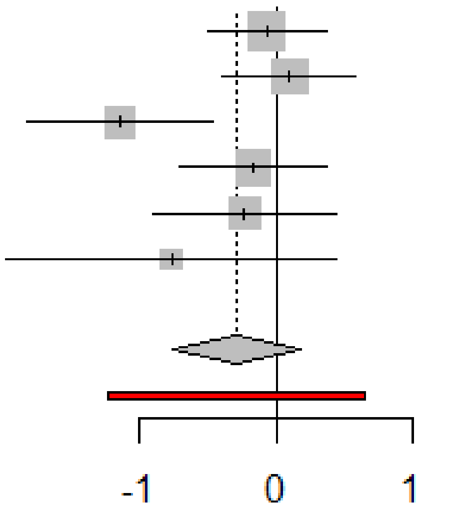
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$

Forest Plots – Totals

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

Heterogeneity: I^2 = 49.8%, p = 0.0764

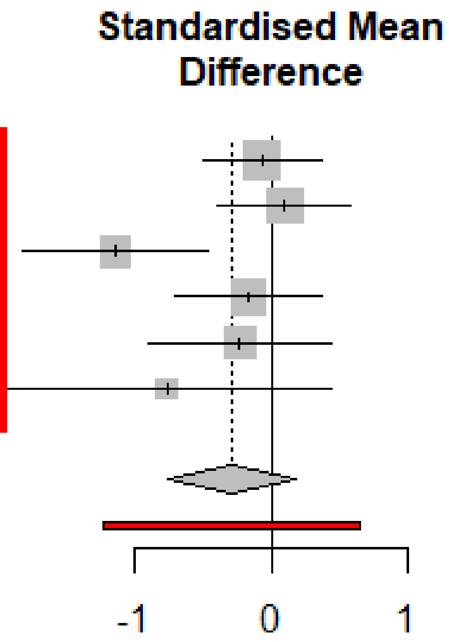
Standardised Mean Difference



Forest Plots – Means and Standard Deviations

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

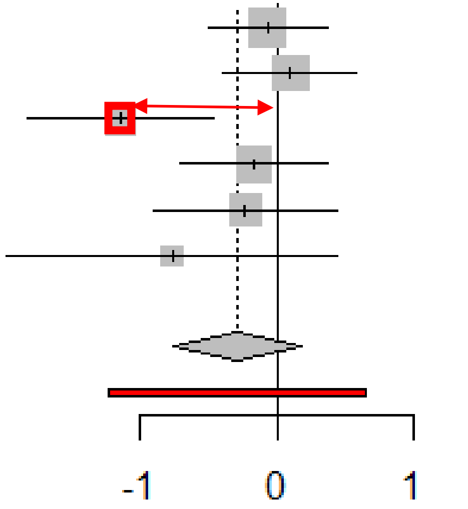
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Forest Plots – Standardised Mean Difference

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	Standardised Mean Difference	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000		-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000		0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400		-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600		-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200		-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000		-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**				**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval									**[-1.23; 0.65]**	

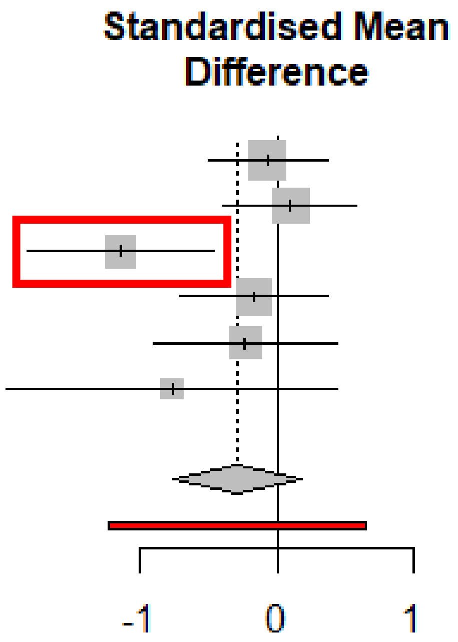
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Forest Plots – Confidence Intervals

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

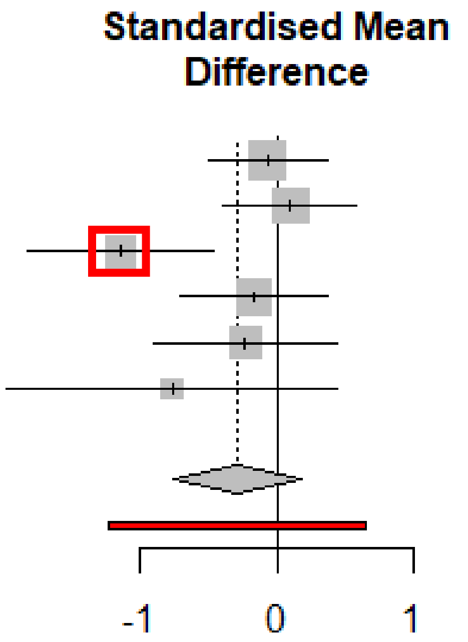
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Forest Plots – Weighting

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

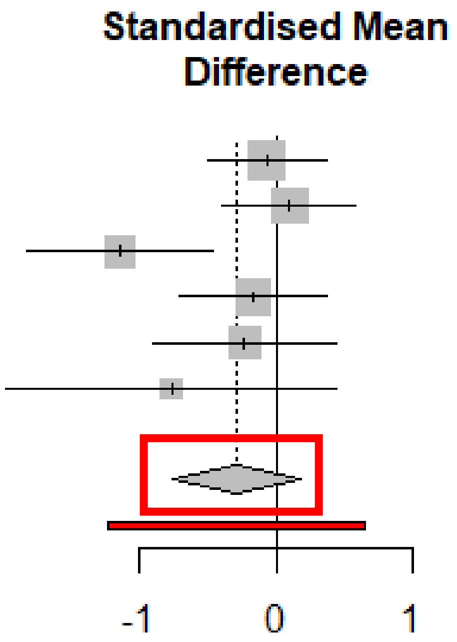
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Forest Plots – Pooled Statistics

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

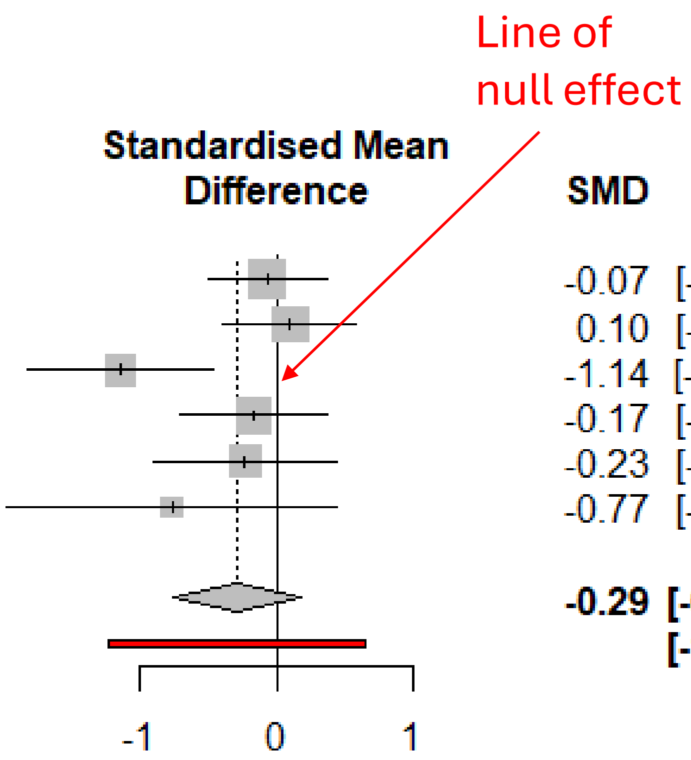
Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Forest Plots – Interpreting the Plot

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	SMD	95%-CI	Weight
Study 1	41	26.60	10.1000	39	27.30	10.6000	-0.07	[-0.51; 0.37]	22.4%
Study 2	34	4.00	1.9000	31	3.80	2.0000	0.10	[-0.39; 0.59]	20.7%
Study 3	20	13.25	2.6300	19	16.26	2.5400	-1.14	[-1.82; -0.46]	15.2%
Study 4	27	1.95	1.5700	26	2.16	0.7600	-0.17	[-0.71; 0.37]	19.1%
Study 5	18	6.30	1.2600	17	6.58	1.1200	-0.23	[-0.89; 0.44]	15.6%
Study 6	7	3.30	3.4000	5	6.30	3.9000	-0.77	[-1.97; 0.44]	7.0%
Random effects model	**147**			**137**			**-0.29**	**[-0.77; 0.19]**	**100.0%**
Prediction interval								**[-1.23; 0.65]**	

Heterogeneity: $I^2 = 49.8\%$, $p = 0.0764$



Publication Bias

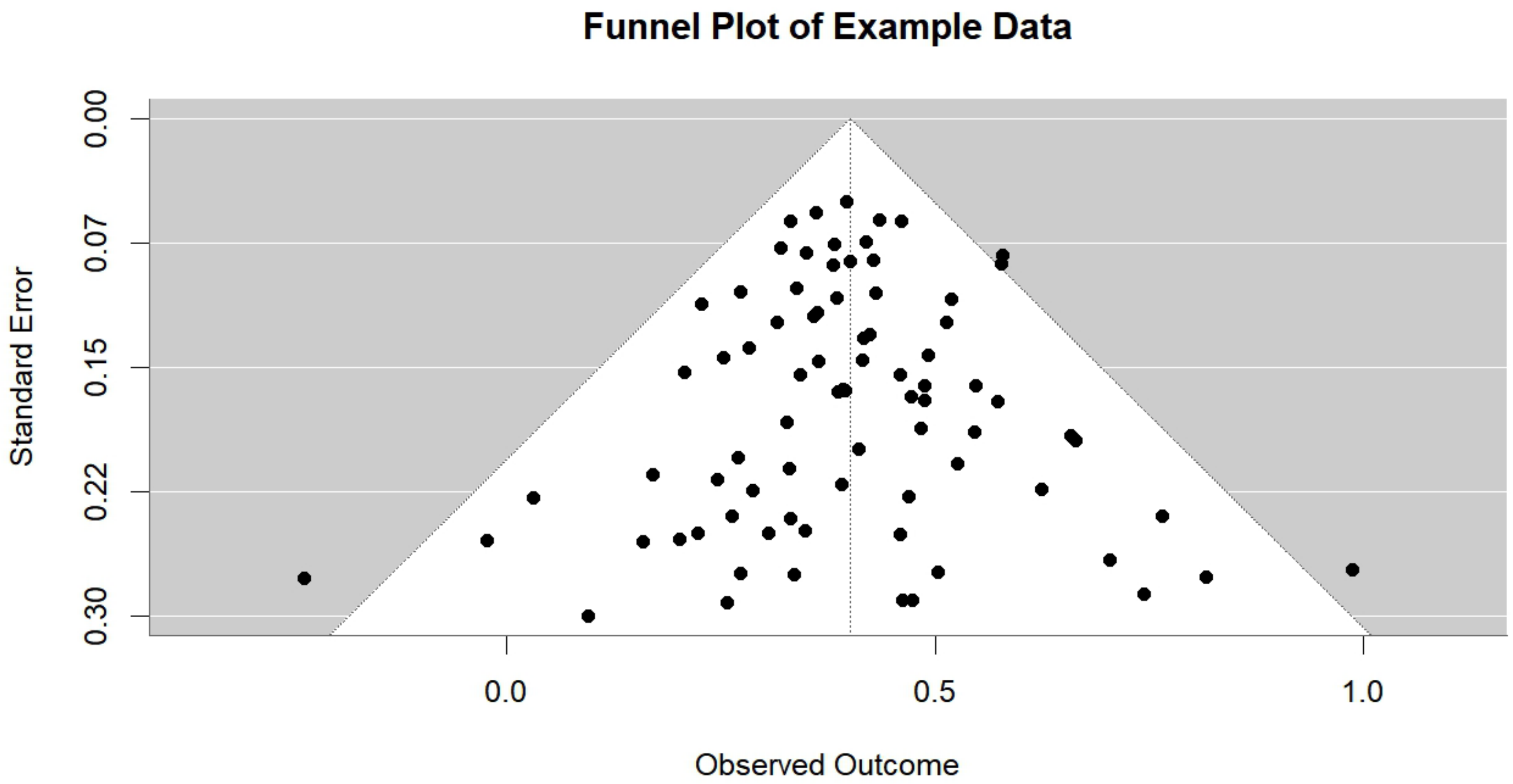
- There are many ways to investigate publication bias
  - We won't have time to cover them all, and you can read more here: https://doing-meta.guide/pub-bias
- Today we will talk about funnel plots and Egger's test which looks at the small study effect

Funnel Plots

- Larger studies are more likely to be published - More resources are used
  - Sample size is related to standard error – higher sample leads to narrower SE and wider confidence intervals
    - Smaller studies less likely to find a significant finding
    - Non-significant studies are less likely to be published
- Funnel Plots – more studies is better (10+ at least)
  - Plotting effect sizes by Standardised mean difference and standard error
  - An exemplar expected funnel-shape in dotted lines
  - A middle line showing the average effect size
  - A symmetrical plot suggests publication bias is less likely
    - Can also look at contours related to significance

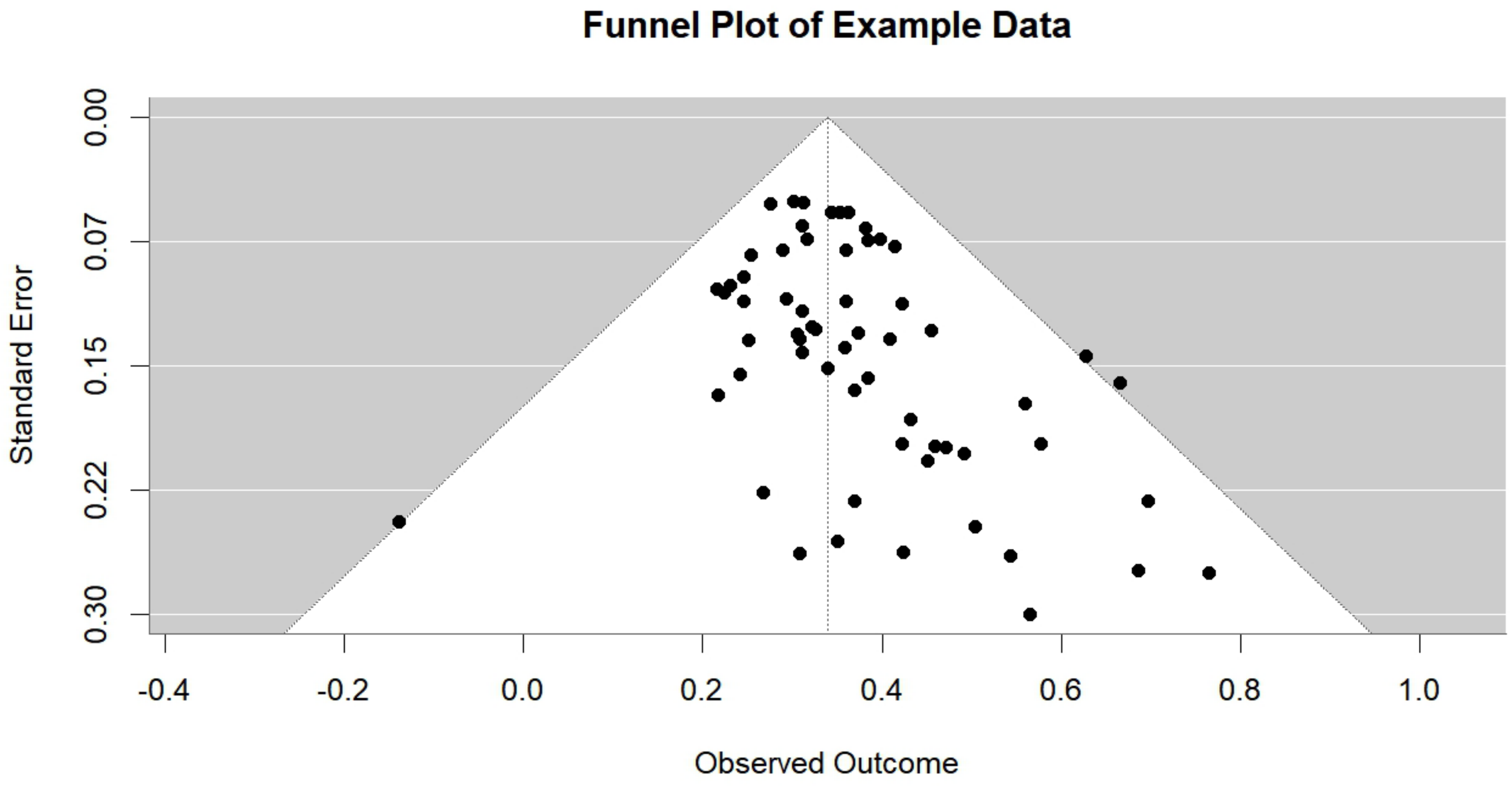
Borenstein, Michael, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2011. *Introduction to Meta-Analysis*. John Wiley & Sons.

Funnel Plots



From example data, synthesised to be relatively symmetrical

Funnel Plots



From example data, synthesised to be less likely to be published if non-significant

Egger’s Test

- Egger’s test uses a regression model to investigate funnel plot asymmetry
- If the plot has a bite out of it, then it will certainly be asymmetrical
  - But the asymmetry is not often easy to notice – especially with fewer studies
  - A significant p-value indicates asymmetry
  - Too few studies makes this unreliable
  - We can never truly know if the pattern is truly caused by publication bias

Further Information

- These are some excellent resources:
  - The Cochrane Handbook is an excellent resource for information: https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-10#section-10-10-2
  - The Doing Meta-Analysis with R Guide is the only R guide you will ever need: https://doing-meta.guide/
  - PRISMA is brilliant for all things related to reporting your research: https://www.prisma-statement.org/