Intermediate course: Meta-analysis of data from animal studies in R

Daniel Schulze

Charité - Institute for Biometry and Clinical Epidemiology

Alexandra Bannach-Brown & Maria Economou

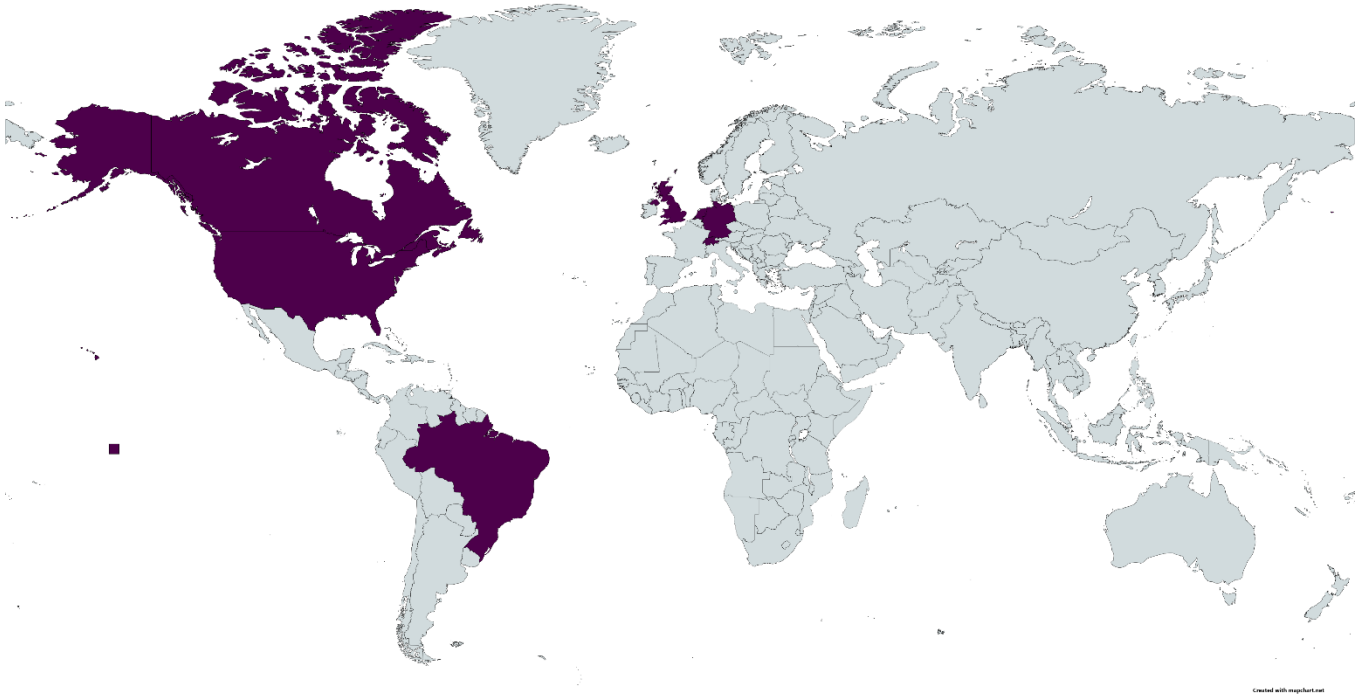
BIH Quest Center

CoReS
Communities for
Open Research Synthesis

Welcome!

CAMARADES

Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies



1. Preclinical meta-research

2. Supporting framework for researchers involved in systematic review and meta-analysis

Collaborative Approach to Meta Analysis and

·C·A·M·A·R·A·D·E·S·

Review of Animal Data from Experimental Studies

CoReS

CAMARADES Berlin - Resources

One-to-one methodological advice

Support for your SR project from research question to meta-analysis
Weekly **drop-in session**: Monday 12-13pm CET

Contact the Helpdesk: **CAMARADES.berlin@charite.de**

Preclinical Systematic Review Software

Free to use online, unlimited projects and reviewers.
http://SyRF.org.uk/

Preclinical Systematic Review wiki website

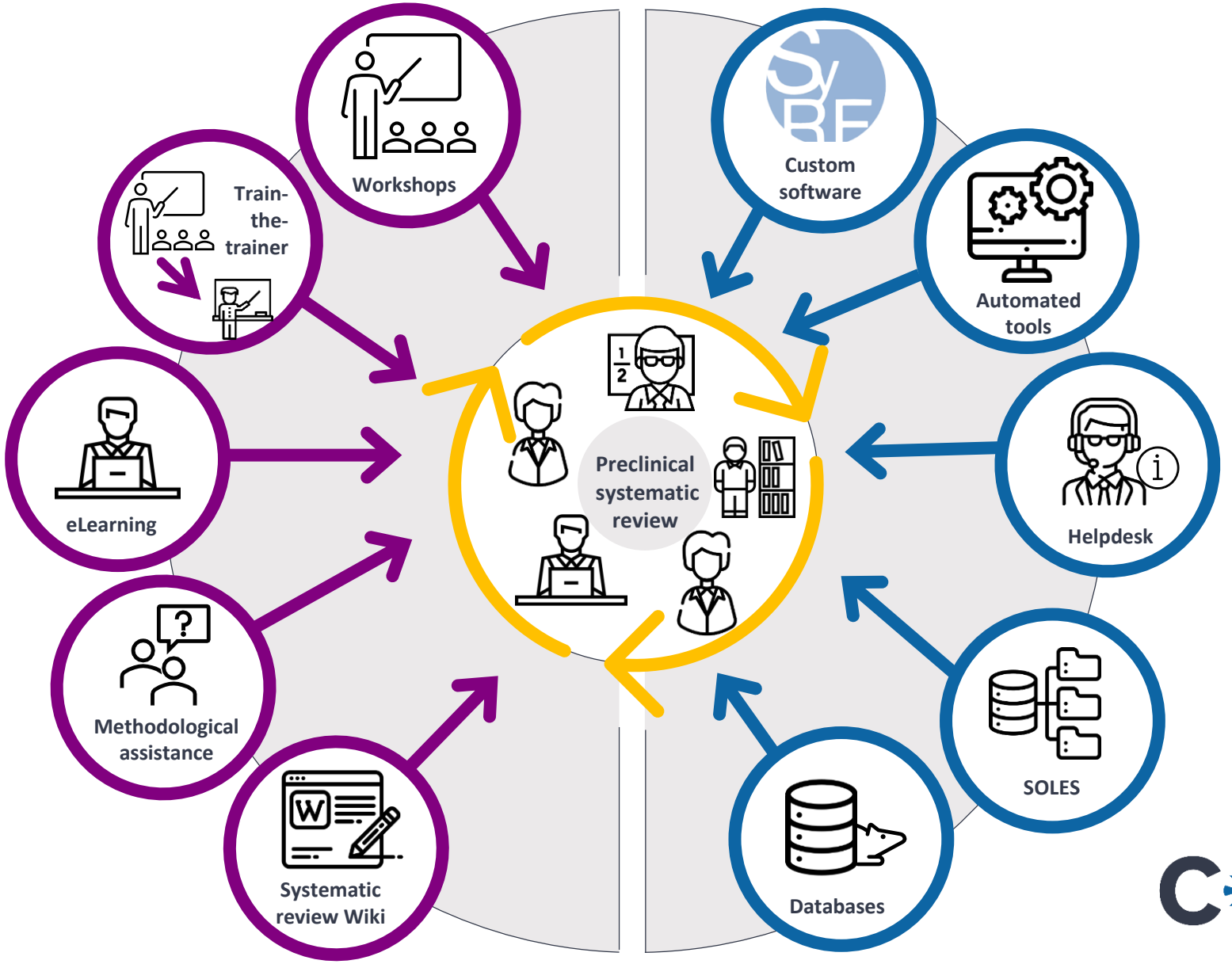
Guide to conducting all steps of a preclinical systematic review.
Information and instructions, links, resources, and further reading.
https://www.CAMARADES.de

Communities for Open Research Synthesis

Community

Education & Support

Infrastructure .



Stay up-to-date with COReS!

Digital Hub

Access resources and sign up for training or advice!

www.cores-hub.io

LinkedIn

Regular updates on education formats, talks, and project developments

https://www.linkedin.com/company/communities-for-open-research-synthesis/

Newsletter

Sign-up to receive updates on the community news, events, workshops, and resources!

http://eepurl.com/h4hsMv

mailchimp

Course structure

1st Day
• *Introduction*
• *Which kind of data? Clarifying data types, outcomes and effect sizes*
• *Data extraction from primary animal studies*
BREAK
• *Calculating effect sizes manually and in* `metafor`
• *Run meta-analysis in R - understanding default settings*
BREAK
• *Plotting: Default forest plots & basic customizations*
• *Day 1 wrap-up & Q&A*

Course structure

2nd Day
• *Heterogeneity in meta-analysis of animal studies*
• *Meta regression:*
• *Regression model & Explaining heterogeneity*
BREAK
• *Running models in R + Plotting*
• *Obstacles*
BREAK
• *Exploring the impacts of Risks of Bias in R*
• *Funnel plotting*
• *Workshop Conclusion & Wrap-up*

Prior knowledge?

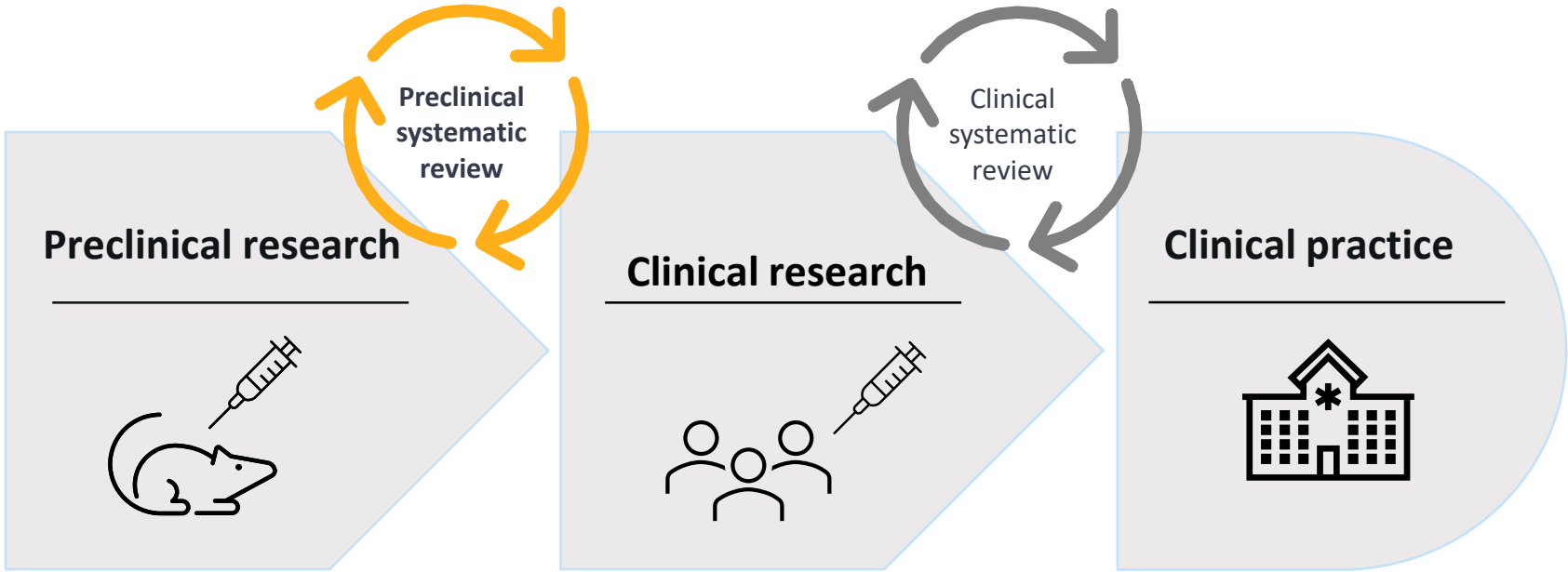
- R
- Rstudio
- Markdown

- Plenty of info on metafor's capabilities:

https://www.metafor-project.org/doku.php/

Systematic review recap

- Systematic review is not routinely used to guide decisions in preclinical research
- Evidence is not synthesized efficiently



Systematic review recap

Systematic review recap

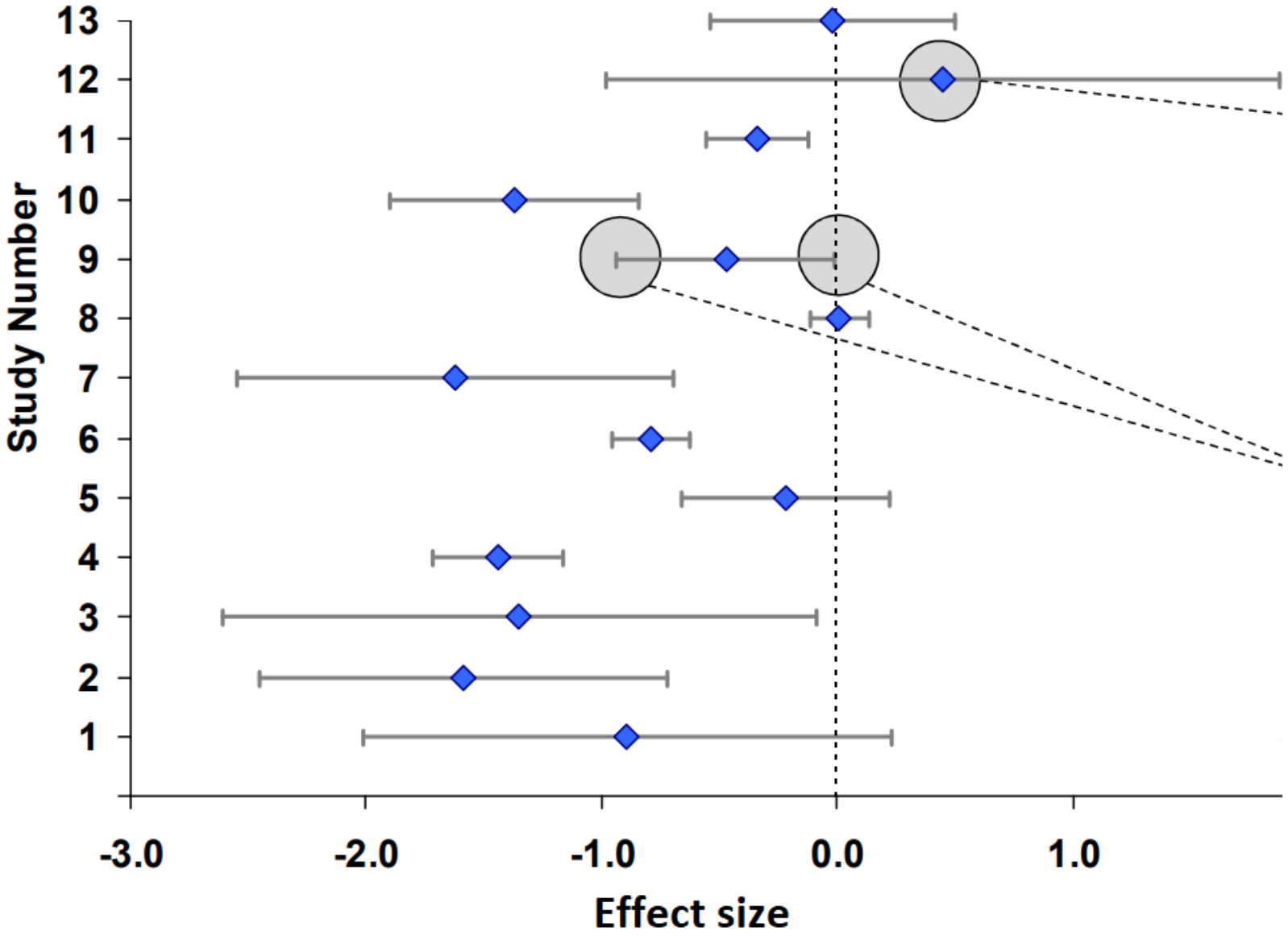
- Clarifying research question
- Search terms
- Quantitative analysis?
- Inclusion / exclusion criteria | PRISMA
- Protocol --> Prospero

- Stack of papers – what next?

1

Data types and effect measures

Meta-analysis



Effect size in study 12

Confidence interval (95%)

Meta-analysis

For quantitatively aggregating across many studies, we need:

- The same measure of effect for every study
- Uncertainty in the effect measure (standard error; 95%-CI; variance)

Data type

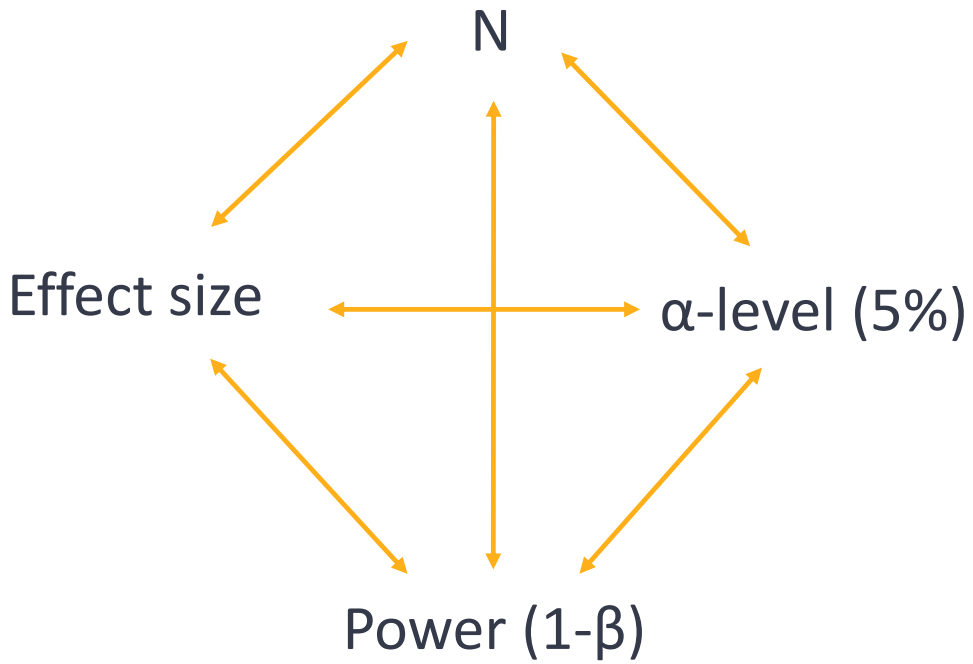
Scale:

- Continuous measures (time, size, ….)
- Binary / categorical outcomes (death, drop-out, therapy response, …)
- Time-to-event (survival, onset of disease, …)

Measure of:

- Difference (groups)
- Association

Effect size



The higher each of these, the likelier a „significant“ result

Raw mean difference

- The outcome is reported on a meaningful scale and all studies in the analysis use the same scale

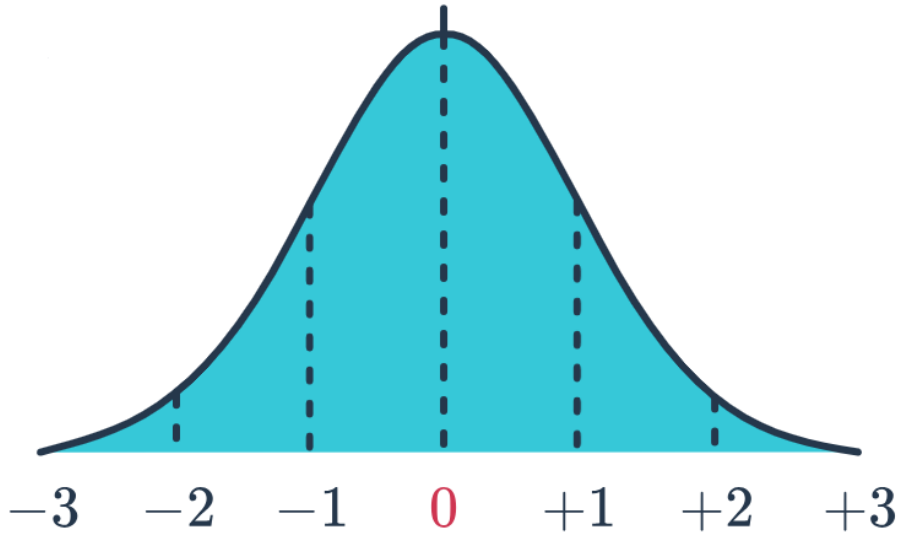
- Meta-analysis is performed directly on the raw difference in means
- E.g. blood pressure – measured in (or converted to) mmHg
- difference = 0 -> no effect

Standardized mean difference (SMD)

- Aka Hedge's g (or Cohen's d, without correction for small samples)

- Logic: Normalizing mean differences with a standard deviation
- SMD = 0 -> no effect

$$\frac{\bar{x}_1 - \bar{x}_2}{s}$$



Standardized mean difference (SMD)

Advantages:

- Applicable to all continuous variables
- Same metric/size for any outcome (cave!)

Disadvantages:

- Abstract metric
- Clinical / biological relevance of effect not self-evident

value	Interpretation
≤ .2	"small" difference
≈ .5	"medium" difference
≥ .8	"large" difference

Standardized mean difference (SMD)

$$ES_i = \frac{\bar{x}_c - \bar{x}_{\text{rx}}}{S_{\text{pooled}}} \times \left(1 - \frac{3}{4N - 9}\right)$$

$$S_{\text{pooled}} = \sqrt{\frac{(n'_c - 1)SD_c^2 + (n_{\text{rx}} - 1)SD_{\text{rx}}^2}{N - 2}}$$

$$VAR = \frac{N}{n_{\text{rx}} \times n'_c} + \frac{ES_i^2}{2\,N}$$

Normalized mean difference

- Can be used when outcomes are on a ratio scale, **where the score on a 'control' or 'sham' animal is known**.

- Advantages: Original unit of measurement is retained (eg tumor or infarct volume mm^3)

Normalized mean difference

$$ES_i = 100 \times \frac{(\bar{x}_c - \bar{x}_{\text{sham}}) - (\bar{x}_{\text{rx}} - \bar{x}_{\text{sham}})}{\bar{x}_c - \bar{x}_{\text{sham}}}$$

$$SD_{\text{c}^*} = 100 \times \frac{SD_c}{\bar{x}_c - \bar{x}_{\text{sham}}}$$

$$SD_{\text{rx}^*} = 100 \times \frac{SD_{\text{rx}}}{\bar{x}_{\text{c}} - \bar{x}_{\text{sham}}}$$

$$VAR = \frac{SD^2_{\text{c}^*}}{n'_c} + \frac{SD^2_{rx*}}{n_{rx*}}$$

Odds ratio

- Difference / association between to binary variables
- How do the odds for O differ between groups G?

	O1	O2	
G1	a	b	n_1
G2	c	d	n_2

Advantages:

- Invariant when reorganizing cross tab
- (directly related to logistic regression)

Disadvantages:

- Abstract metric
- Odds ≠ risk

Odds ratio

$$OR_i = \frac{a_i \times d_i}{b_i \times c_i}$$

- OR = 1 -> no effect. OR = 0.5 corresponds to OR = 2
- For aggregating, ln(OR) is used (0 -> no effect, ln(OR=0.5) = -0.69, ln(OR=2) = 0.69

$$VAR = (1/a_i) + (1/b_i) + (1/c_i) + (1/d_i)$$

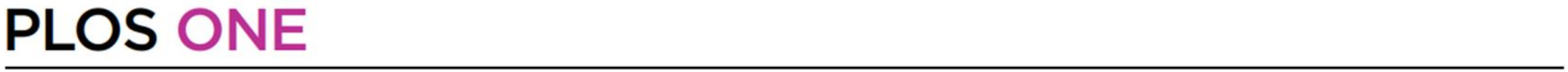
Further effect measures

mean difference	$\bar{x}_1 - \bar{x}_2$	$\mu_1 - \mu_2$	$s_p^2(1/n_1 + 1/n_2)$
response ratio	$\ln[\bar{x}_1 / \bar{x}_2]$	$\ln[\mu_1 / \mu_2]$	$\frac{s_1^2}{n_1\bar{x}_1^2} + \frac{s_2^2}{n_2\bar{x}_2^2}$
risk difference	$p_1 - p_2$	$\pi_1 - \pi_2$	$p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$
risk ratio	$\ln[p_1 / p_2]$	$\ln[\pi_1 / \pi_2]$	$\frac{1}{p_1 n_1} - \frac{1}{n_1} + \frac{1}{p_2 n_2} - \frac{1}{n_2}$
correlation coefficient	r	ρ	$\frac{(1-r^2)^2}{n-1}$
transformed correlation	$\frac{1}{2}\ln\left[\frac{1+r}{1-r}\right]$	$\frac{1}{2}\ln\left[\frac{1+\rho}{1-\rho}\right]$	$1/(n-3)$
proportion	p	π	$p(1-p)/n$
log odds	$\ln[p/(1-p)]$	$\ln[\pi/(1-\pi)]$	$1/[np(1-p)]$
mean	$\bar{x}$	μ	s^2/n

2

Introduction to our working example: Centrofobic rats (in pain)

Centrofobic rats in pain



RESEARCH ARTICLE

A systematic review and meta-analysis of thigmotactic behaviour in the open field test in rodent models associated with persistent pain

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0290382

Zhang XY, Diaz-delCastillo M, Kong L, Daniels N, MacIntosh-Smith W, Abdallah A, Domanski D, Sofrenovic D, Yeung TP, Valiente D, Vollert J. A systematic review and meta-analysis of thigmotactic behaviour in the open field test in rodent models associated with persistent pain. Plos one. 2023 Sep 8;18(9):e0290382.

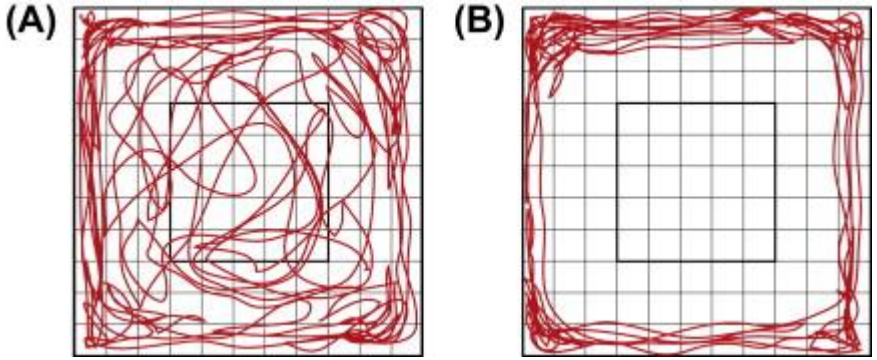
Centrofobic rats in pain

Failure of translating preclinical findings on new analgesic drugs to the clinic[1]

What is a good model for pain? What is a good outcome measure?

- Limb withdrawal
- **Open field test** (-> thigmotaxis)

-> Stress? Anxiety? Pain?



https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/open-field-test

[1] Borsook D, Hargreaves R, Bountra C, Porreca F. Lost but making progress—Where will new analgesic drugs come from? Science Translational Medicine. 2014; 6(249):249sr3–sr3. https://doi.org/10.1126/ scitranslmed.3008320

Is thigmotaxis a valid measure of pain?

"This systematic review aimed to 1) assess whether thigmotaxis can be affected by injury and disease models associated with **persistent pain** and analgesic drug treatments in rodents; …."

- (Spinal) Nerve injury
- Head injury
- Chemotherapy
- Migraine models
- …..

Comparing to sham

Our example

Does the model work? Is pain reflected in the open field test?

We will focus on:

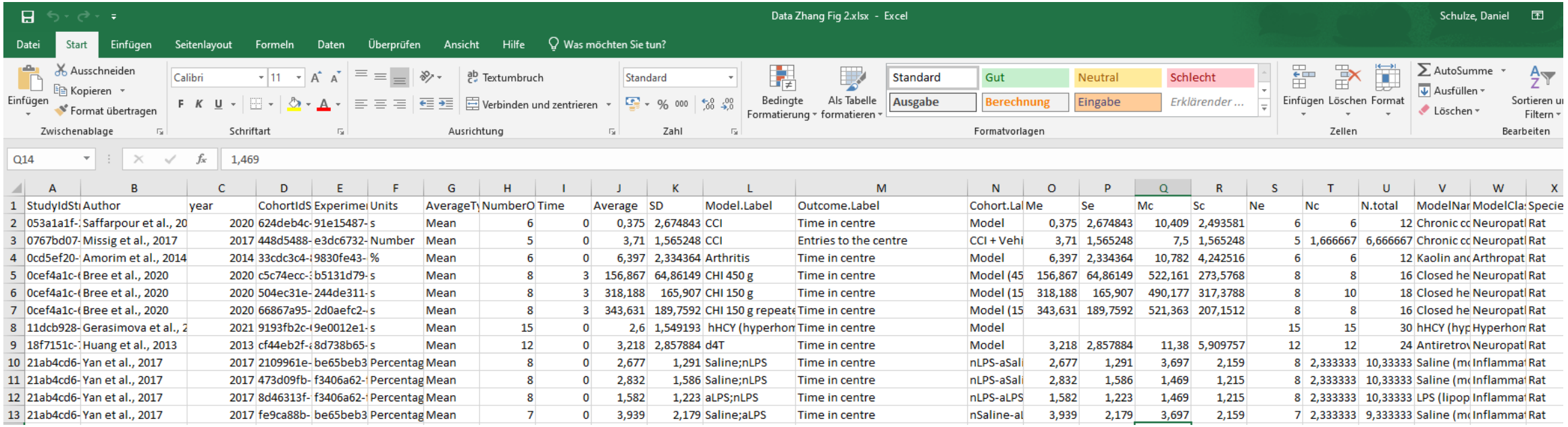
- Rats
- Exp. vs sham control
- Thigmotaxis as measured by time in center / forays to center

- -> Fig 2 in paper by Zhang et al.

Data

Raw data: https://osf.io/rmt97/

Prepared data set:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	StudyIdSt	Author	year	CohortIdS	Experime	Units	AverageT	NumberO	Time	Average	SD	Model.Label	Outcome.Label	Cohort.La	Me	Se	Mc	Sc	Ne	Nc	N.total	ModelNar	ModelCla	Specie
2	053a1a1f-	Saffarpour et al., 20	2020	624deb4c-	91e15487-	s	Mean	6	0	0,375	2,674843	CCI	Time in centre	Model	0,375	2,674843	10,409	2,493581	6	6	12	Chronic cc	Neuropatl	Rat
3	0767bd07-	Missig et al., 2017	2017	448d5488-	e3dc6732-	Number	Mean	5	0	3,71	1,565248	CCI	Entries to the centre	CCI + Vehi	3,71	1,565248	7,5	1,565248	5	1,666667	6,666667	Chronic cc	Neuropatl	Rat
4	0cd5ef20-	Amorim et al., 2014	2014	33cdc3c4-	9830fe43-	%	Mean	6	0	6,397	2,334364	Arthritis	Time in centre	Model	6,397	2,334364	10,782	4,242516	6	6	12	Kaolin anc	Arthropat	Rat
5	0cef4a1c-	Bree et al., 2020	2020	c5c74ecc-	b5131d79-	s	Mean	8	3	156,867	64,86149	CHI 450 g	Time in centre	Model (45	156,867	64,86149	522,161	273,5768	8	8	16	Closed he	Neuropatl	Rat
6	0cef4a1c-	Bree et al., 2020	2020	504ec31e-	244de311-	s	Mean	8	3	318,188	165,907	CHI 150 g	Time in centre	Model (15	318,188	165,907	490,177	317,3788	8	10	18	Closed he	Neuropatl	Rat
7	0cef4a1c-	Bree et al., 2020	2020	66867a95-	2d0aefc2-	s	Mean	8	3	343,631	189,7592	CHI 150 g repeate	Time in centre	Model (15	343,631	189,7592	521,363	207,1512	8	8	16	Closed he	Neuropatl	Rat
8	11dcb928-	Gerasimova et al., 2	2021	9193fb2c-	9e0012e1-	s	Mean	15	0	2,6	1,549193	hHCY (hyperhon	Time in centre	Model					15	15	30	hHCY (hyp	Hyperhon	Rat
9	18f7151c-	Huang et al., 2013	2013	cf44eb2f-	8d738b65-	s	Mean	12	0	3,218	2,857884	d4T	Time in centre	Model	3,218	2,857884	11,38	5,909757	12	12	24	Antiretrov	Neuropatl	Rat
10	21ab4cd6-	Yan et al., 2017	2017	2109961e-	be65beb3	Percentag	Mean	8	0	2,677	1,291	Saline;nLPS	Time in centre	nLPS-aSali	2,677	1,291	3,697	2,159	8	2,333333	10,33333	Saline (m	Inflamma	Rat
11	21ab4cd6-	Yan et al., 2017	2017	473d09fb-	f3406a62-	Percentag	Mean	8	0	2,832	1,586	Saline;nLPS	Time in centre	nLPS-aSali	2,832	1,586	1,469	1,215	8	2,333333	10,33333	Saline (m	Inflamma	Rat
12	21ab4cd6-	Yan et al., 2017	2017	8d46313f-	f3406a62-	Percentag	Mean	8	0	1,582	1,223	aLPS;nLPS	Time in centre	nLPS-aLPS	1,582	1,223	1,469	1,215	8	2,333333	10,33333	LPS (lipop	Inflamma	Rat
13	21ab4cd6-	Yan et al., 2017	2017	fe9ca88b-	be65beb3	Percentag	Mean	7	0	3,939	2,179	Saline;aLPS	Time in centre	nSaline-al	3,939	2,179	3,697	2,159	7	2,333333	9,333333	Saline (m	Inflamma	Rat

Download the dataset

- Github: https://github.com/camaradesberlin/Intermediate-MA-workshop
- Download the “R exercise" folder
- Unzip the folder
- Open the Rproject by clicking the "Intermediate MA.Rproj" file

3

Extracting data from primary studies

What data to extract?

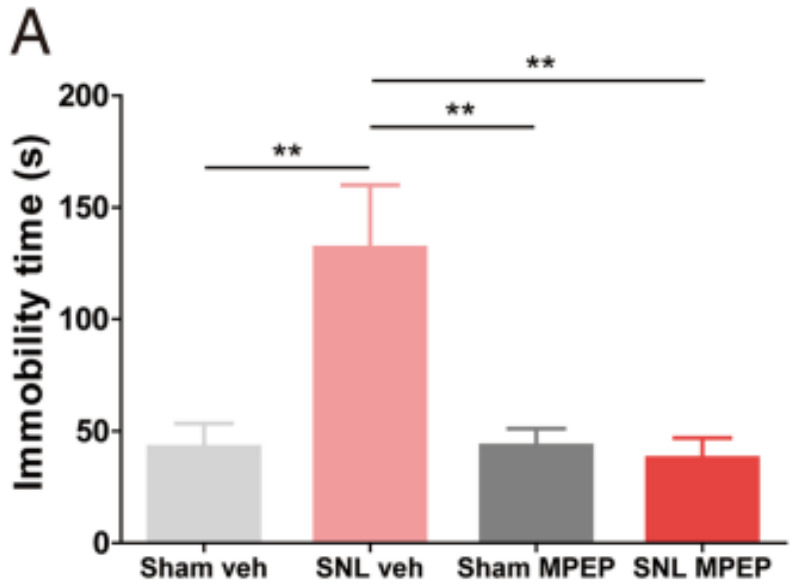
$$ES_i = \frac{\bar{x}_c - \bar{x}_{\text{rx}}}{S_{\text{pooled}}} \times \left(1 - \frac{3}{4N - 9}\right)$$

$$S_{\text{pooled}} = \sqrt{\frac{(n'_c - 1)SD_c^2 + (n_{\text{rx}} - 1)SD_{\text{rx}}^2}{N - 2}}$$

$$VAR = \frac{N}{n_{\text{rx}} \times n'_c} + \frac{ES_i^2}{2\,N}$$

Sources of data

- Figures
- Tables
- Text



Variable	Intervention group mean (s.e.)	Control group mean (s.e.)
Activities of daily living (ADCS ADL total score)	T0: 60.55 (0.91)	T0: 60.53 (0.91)
	T1: 62.35 (0.91)	T1: 57.47 (0.91)
	T2: 61.26 (1.00)	T2: 53.50 (0.94)
Neuropsychiatric symptom profiles (NPI total score)	T0: 11.25 (1.26)	T0: 11.77 (1.26)
	T1: 10.05 (1.26)	T1: 15.71 (1.26)
	T2: 10.40 (1.38)	T2: 16.09 (1.29)
Executive function and language ability (semantic word fluency, number of words)	T0: 13.60 (0.65)	T0: 13.92 (0.65)
	T1: 15.27 (0.65)	T1: 12.46 (0.65)
	T2: 14.15 (0.69)	T2: 12.05 (0.67)

Sources of data

- When SEM is given instead of SD:

$$SD_c = SEM_c \times \sqrt{n_c}$$

- When median & range (or IQR) are given instead of SD:
- Transformation to mean and SD via R package metamedian (Hozo et al.)

Variable	Intervention group mean (s.e.)	Control group mean (s.e.)
Activities of daily living (ADCS ADL total score)	T0: 60.55 (0.91)	T0: 60.53 (0.91)
	T1: 62.35 (0.91)	T1: 57.47 (0.91)
	T2: 61.26 (1.00)	T2: 53.50 (0.94)
Neuropsychiatric symptom profiles (NPI total score)	T0: 11.25 (1.26)	T0: 11.77 (1.26)
	T1: 10.05 (1.26)	T1: 15.71 (1.26)
	T2: 10.40 (1.38)	T2: 16.09 (1.29)
Executive function and language ability (semantic word fluency, number of words)	T0: 13.60 (0.65)	T0: 13.92 (0.65)
	T1: 15.27 (0.65)	T1: 12.46 (0.65)
	T2: 14.15 (0.69)	T2: 12.05 (0.67)

Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC medical research methodology. 2005 Dec;5:1-0.

3a Exercise

Extract data from Gerasimova et al.

- First identify the relevant experiment / comparison
- Find means, SDs
- Hints:
  - N = 15 in both groups
  - The numbers are in the text

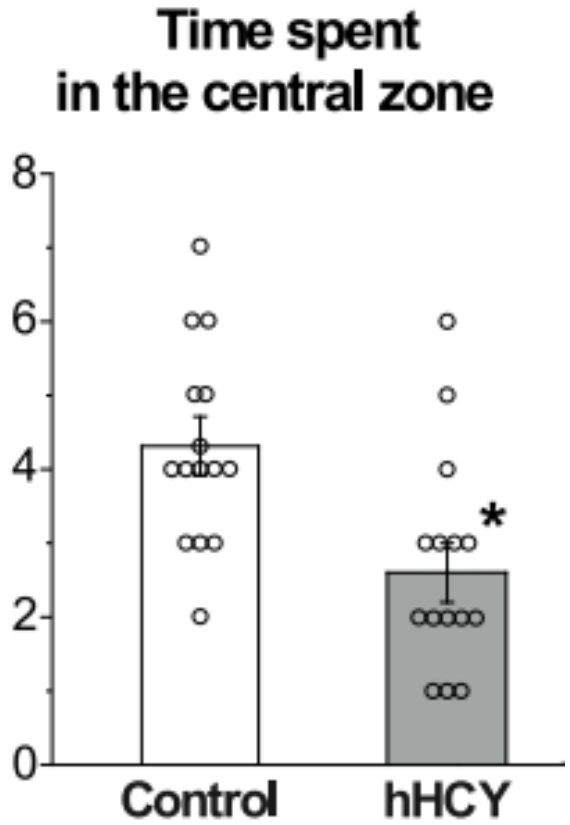
Gerasimova E, Burkhanova G, Chernova K, Zakharov A, Enikeev D, Khaertdinov N, Giniatullin R, Sitdikova G. Hyperhomocysteinemia increases susceptibility to cortical spreading depression associated with photophobia, mechanical allodynia, and anxiety in rats. Behavioural brain research. 2021 Jul 9;409:113324.

Supporting software

- SyRF

- Measuring stuff:
  - Universal Desktop Ruler
  - Adobe Desktop Ruler
  - Webplot Digitizer

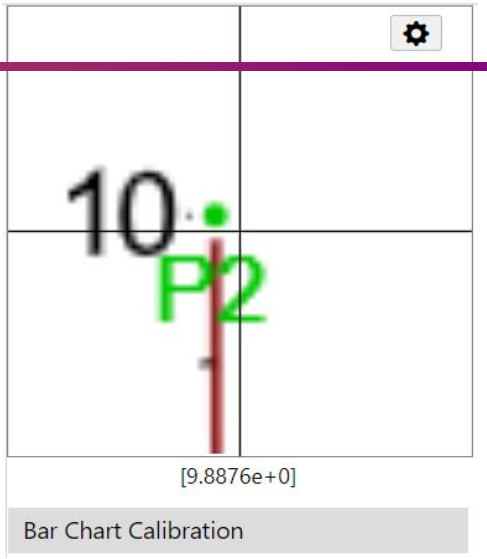
- Further reading: https://osf.io/np3az

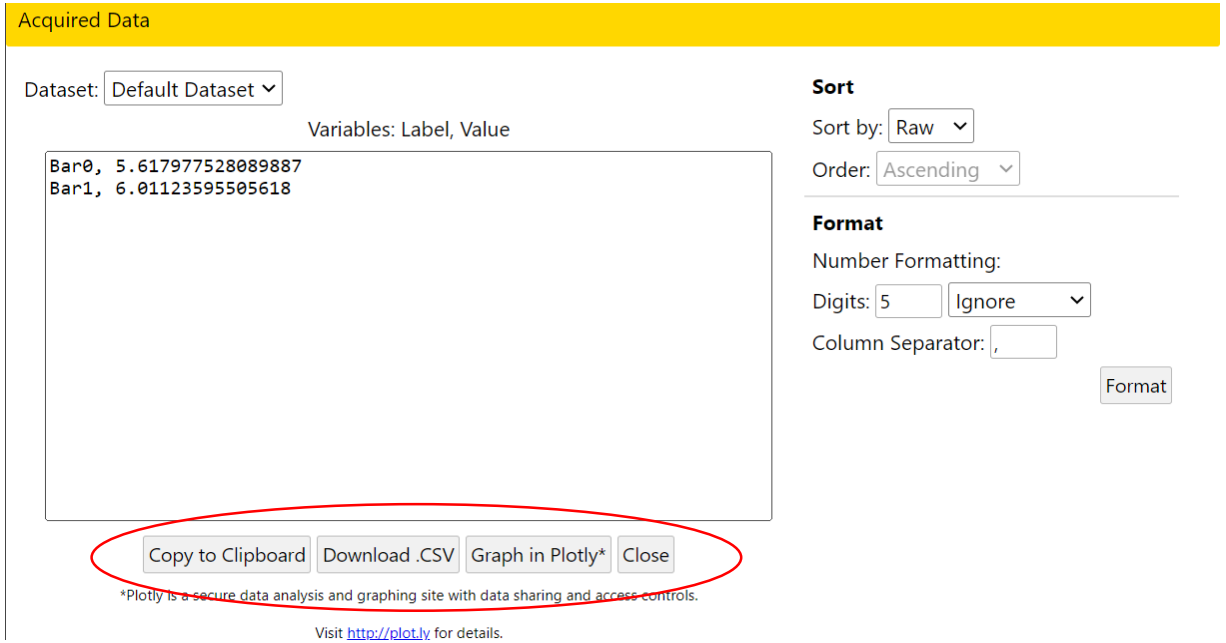


Gerasimova et al.

WebPlotDigitizer

1. Load the PDF + navigate to the pdf page
2. Calibrate Axes
3. Assign points on the graph
4. View Data and export data





4

Calculating effect sizes

Standardized mean difference (SMD)

$$ES_i = \frac{\bar{x}_c - \bar{x}_{\text{rx}}}{S_{\text{pooled}}} \times \left(1 - \frac{3}{4N - 9}\right)$$

$$S_{\text{pooled}} = \sqrt{\frac{(n'_c - 1)SD_c^2 + (n_{\text{rx}} - 1)SD_{\text{rx}}^2}{N - 2}}$$

$$VAR = \frac{N}{n_{\text{rx}} \times n'_c} + \frac{ES_i^2}{2\,N}$$

N control

	A	B	C	D	E	F	G	H	I	J
1	StudyIdSt	Author	Me	Se	Mc	Sc	Ne	Nc	N.total	ModelN
2	053a1a1f-	Saffarpour et al., 20	0,375	2,674843	10,409	2,493581	6	6	12	Chronic
3	0767bd07-	Missig et al., 2017	3,71	1,565248	7,5	1,565248	5	1,666667	6,666667	Chronic
4	0cd5ef20-	Amorim et al., 2014	6,397	2,334364	10,782	4,242516	6	6	12	Kaolin a
5	0cef4a1c-	Bree et al., 2020	156,867	64,86149	522,161	273,5768	8	8	16	Closed h
6	0cef4a1c-	Bree et al., 2020	318,188	165,907	490,177	317,3788	8	10	18	Closed h
7	0cef4a1c-	Bree et al., 2020	343,631	189,7592	521,363	207,1512	8	8	16	Closed h
8	11dcb928-	Gerasimova et al., 2					15	15	30	hHCY (h
9	18f7151c-	Huang et al., 2013	3,218	2,857884	11,38	5,909757	12	12	24	Antiretr
10	21ab4cd6-	Yan et al., 2017	2,677	1,291	3,697	2,159	8	2,333333	10,33333	Saline (r

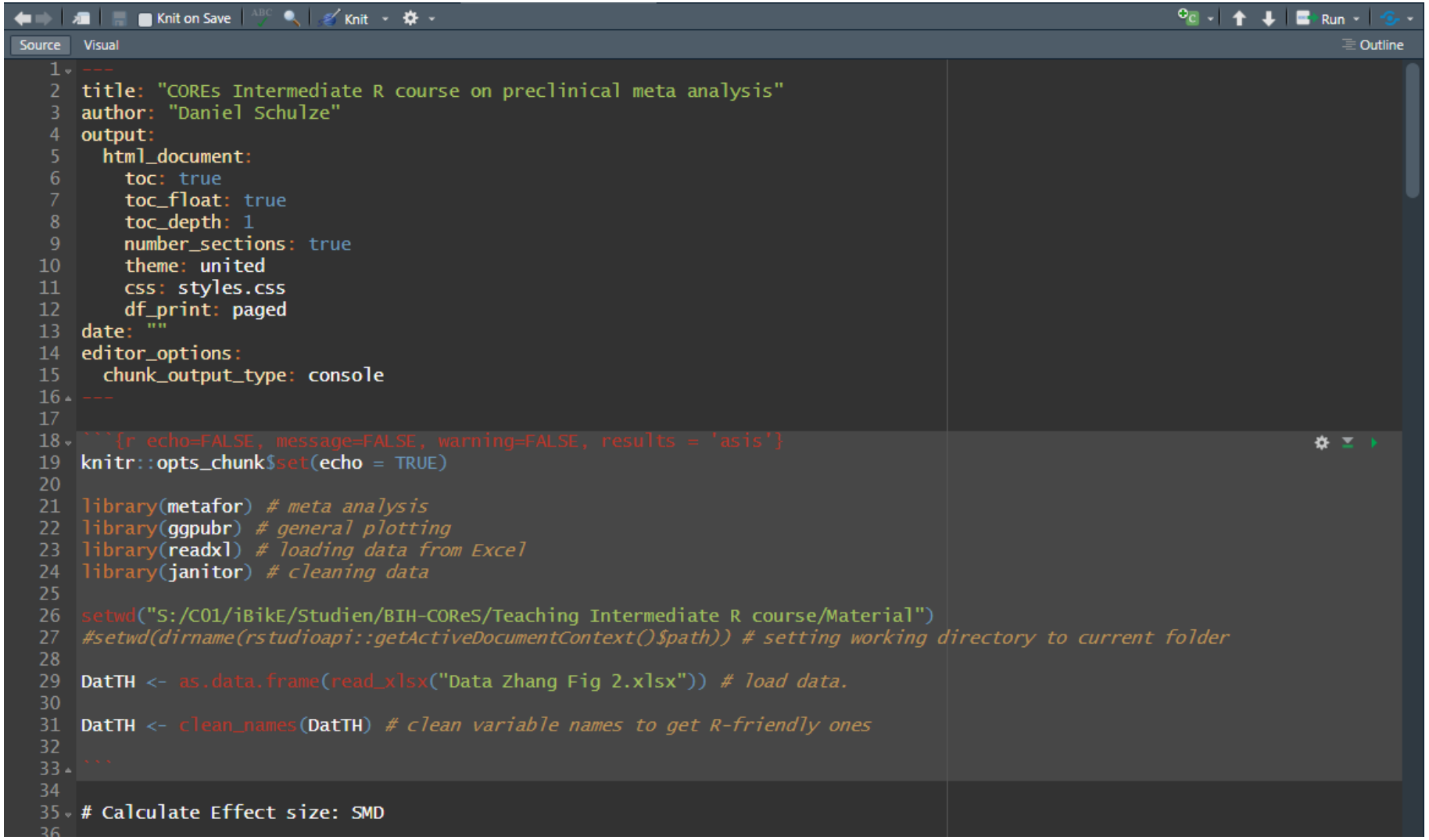
N control

A single experiment can contain a number of comparisons. If the control cohort is serving more than one treatment group, we correct the number of animals reported in the control cohort by the number of treatment groups.

$$n_c' = \frac{n_c}{\text{num. treatmentgroups}}$$

$$N = n_{\text{rx}} + n_c'$$

Intro to R & R Markdown



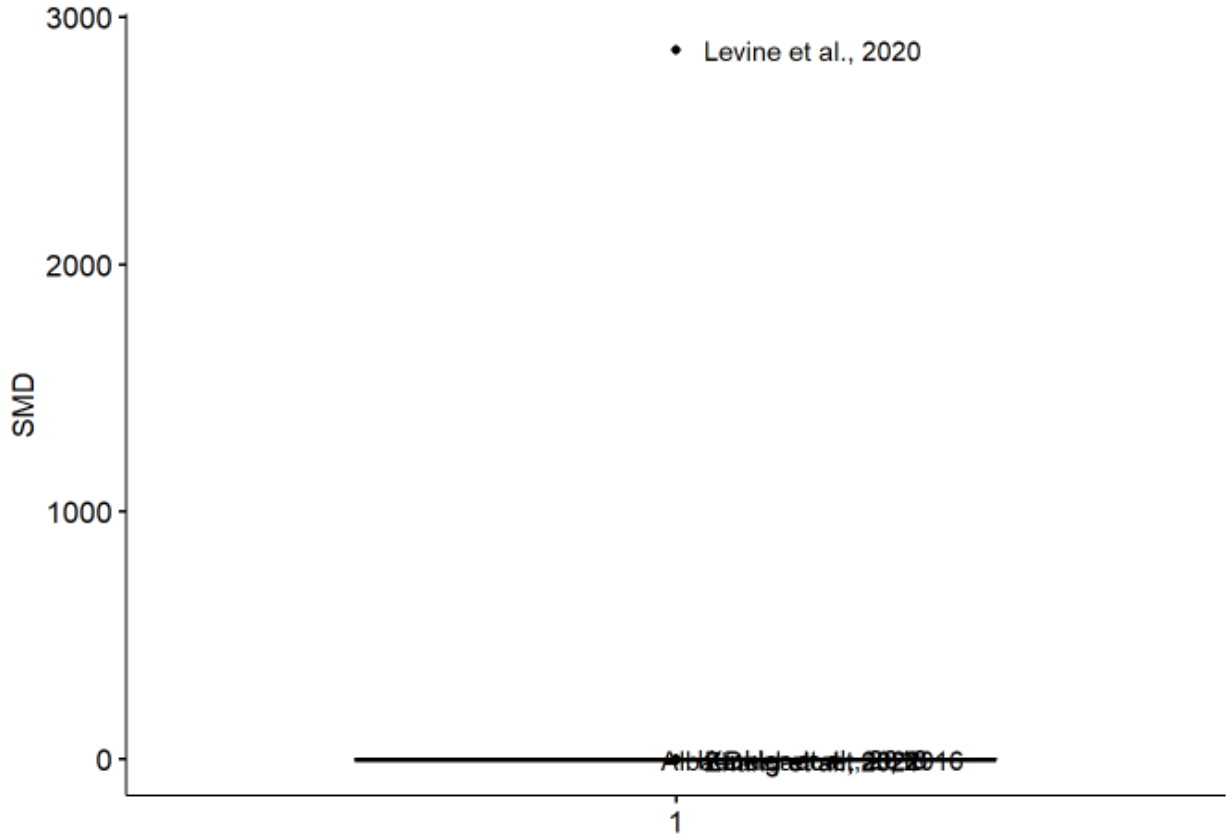
4a Exercise

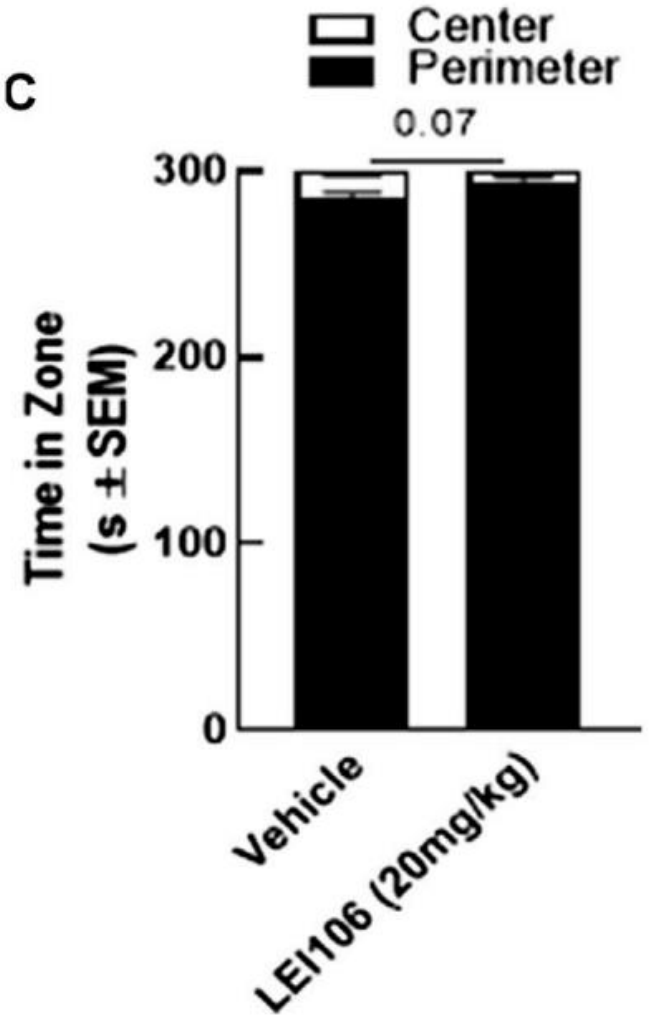
R Markdown Script

Calculate effect sizes manually

- Pooled SD
- SMD
- Variance

Outlier in the data





Levine A, Liktor-Busa E, Karlage KL, Giancotti L, Salvemini D, Vanderah TW, Largent-Milnes TM. DAGLα inhibition as a non-invasive and translational model of episodic headache. Frontiers in pharmacology. 2021 Jan 12;11:615028.

4b Exercise

R Markdown Script

Calculate effect sizes in metafor

- **escalc()**

Specifics for NMD

- Not part of escalc()
- Has to be done manually
- Pay attention to the direction of effect!

Specifics for Odds Ratio

- How to deal with 0 events?

$$OR_i = \frac{a_i \times d_i}{b_i \times c_i}$$

Two approaches:

- Adding 0.5 to all zero cells in the cross table
- Arcsin transformation: "AS" in escalc() (cave: abstract measure of effect!)

$$arcsin\sqrt{(a/a+b)} - arcsin\sqrt{(c/c+d)}$$

5

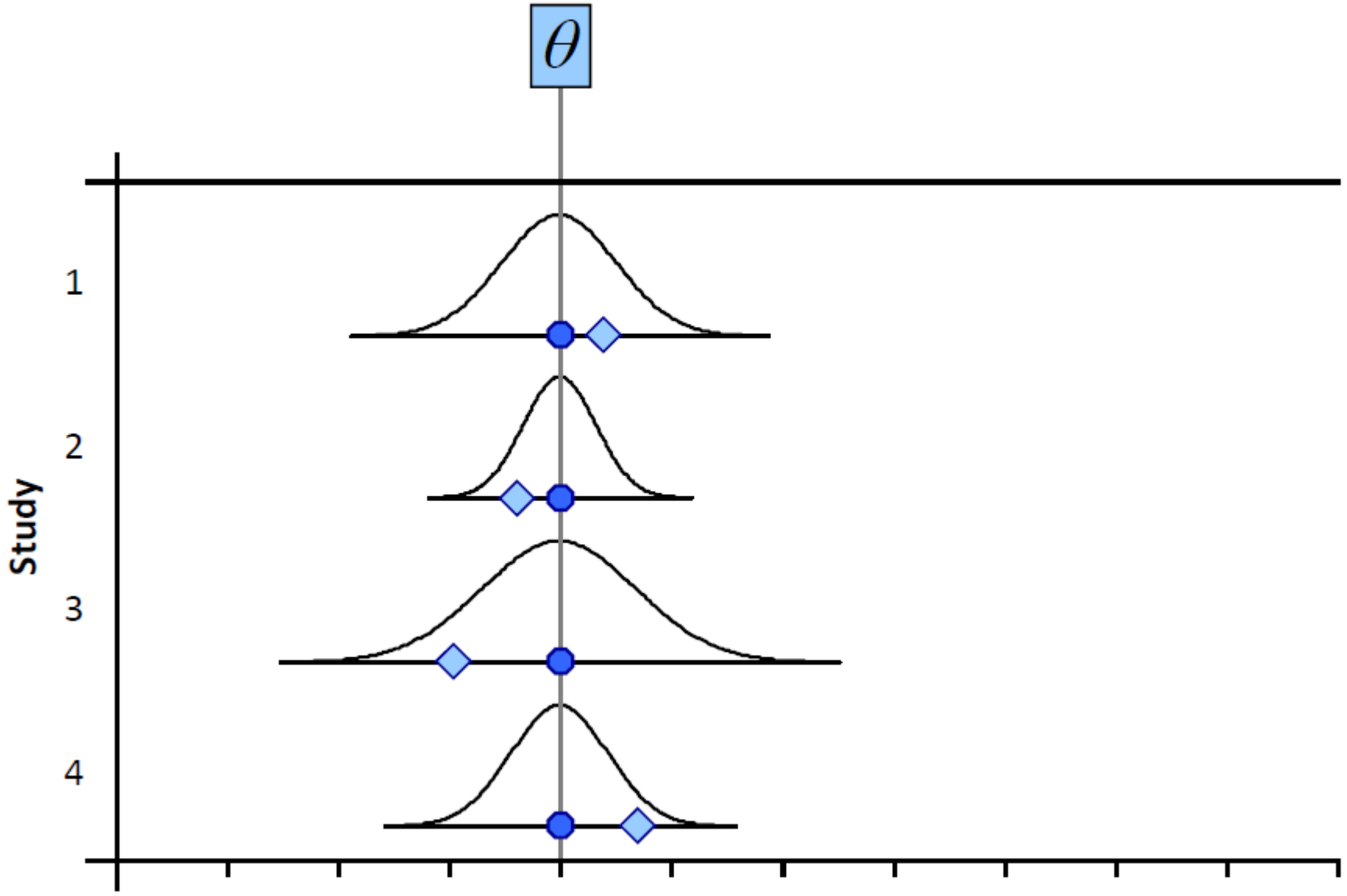
Meta-analytical models

MA = Weighted average of effects

- Logic: Weight study results by study specific uncertainty (size, SD, effect)
- Studies get a larger weight, when:
  - Larger N
  - Smaller SD
  - Larger group difference

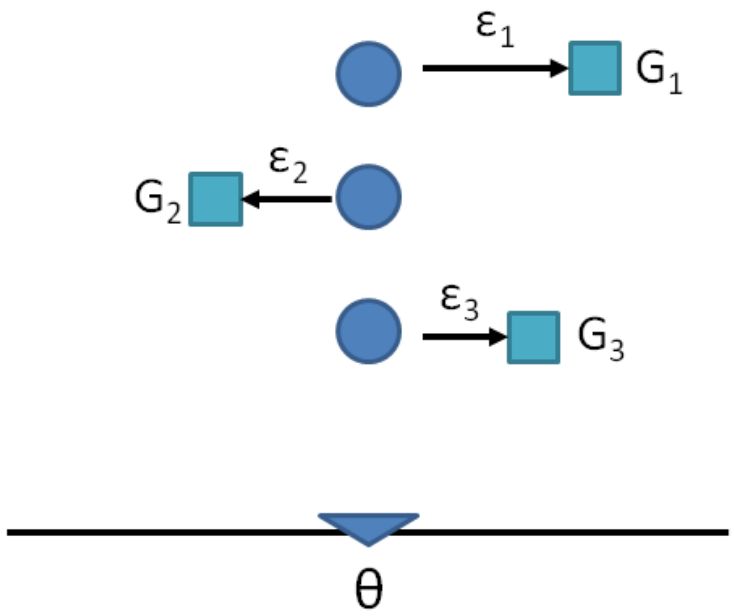
$$\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i} \qquad w_i = \frac{1}{VAR}$$

Equal/fixed effects model



Equal/fixed effects model

The Fixed-Effect Model

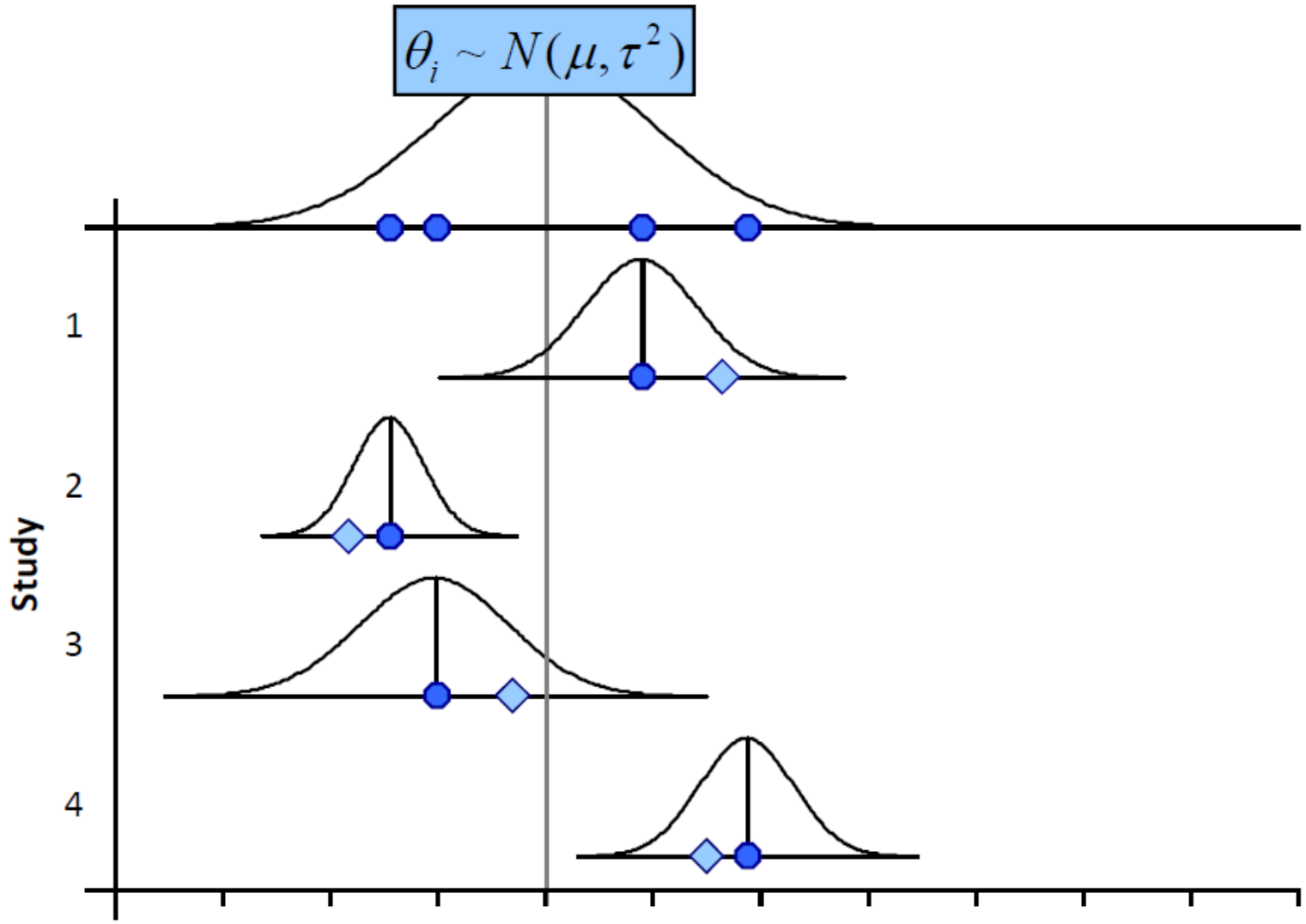


We assume that all of studies (blue circles) share a common effect-size (blue triangle; θ). Thus, the only reason our studies differ in their results is sampling error (ε_i).

Thus, the observed effect (G_i) is assumed to be a function of:

$$G_i = \theta + \varepsilon_i$$

Random effects model



Random effects model

The Random-Effects Model

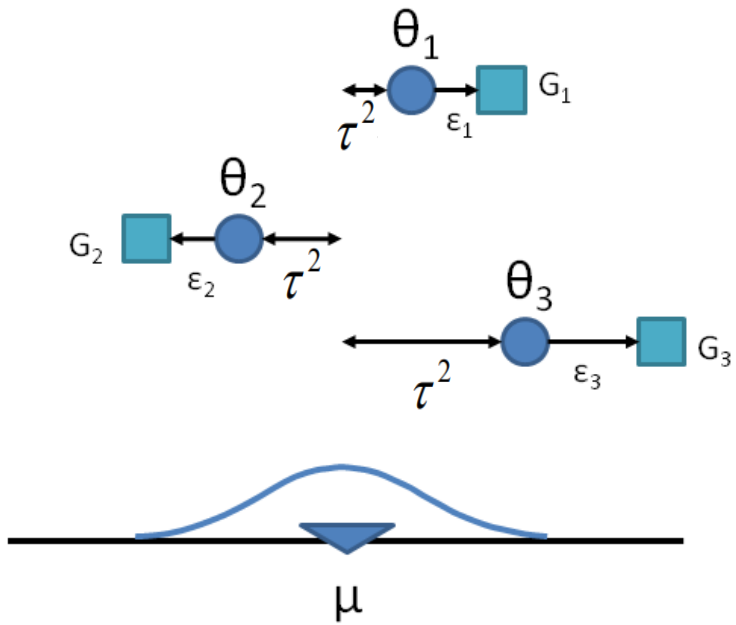
Now we have multiple sources of error:

τ^2 is the true variation in effect sized. Or, the distance between the population effect for each study (θ_i) and summary effect (μ).

ε_i, the sampling error for each observed effect (G_i).

Thus:

$$G_i = \mu + \tau^2 + \varepsilon_i$$



Underlying effects (θ) are normally distributed around some average effect (μ).

Random effects model

Model

$$y_i = \overbrace{\mu + u_i}^{\theta_i} + \varepsilon_i \qquad u_i \sim N(0, \tau^2)$$

Parameter Estimate

$$\hat{\mu} = \frac{\sum w_i y_i}{\sum w_i} \qquad w_i = \frac{1}{v_i + \hat{\tau}^2}$$

Var and SE of the Estimate

$$Var[\hat{\mu}] = \frac{1}{\sum w_i} \qquad SE[\hat{\theta}] = \sqrt{\frac{1}{\sum w_i}}$$

Inference

$$z = \frac{\hat{\mu}}{SE[\hat{\mu}]} \qquad \hat{\mu} \pm 1.96 SE[\hat{\mu}]$$

Random effects model

- Is the standard nowadays, default in metafor
- Estimator: Restricted maximum likelihood (REML), default in metafor
- Confidence intervals: Via method by Knapp-Hartung, set in metafor via test = "knha"

5a Exercise

R Markdown Script

Run Meta-Analysis

- **rma()**
- Keep in mind to specify **test = „knha“**

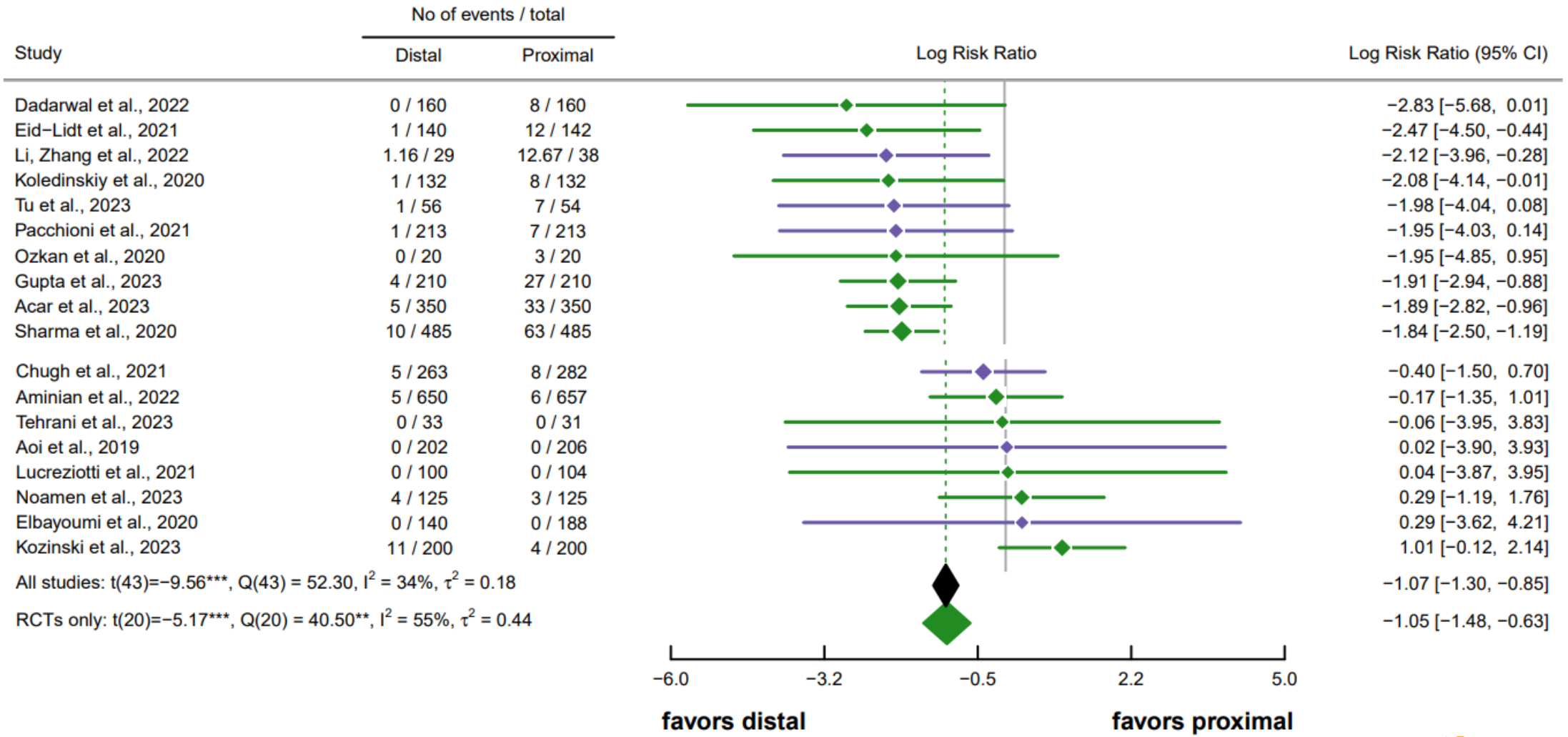
Random effects model

```
##
## Random-Effects Model (k = 90; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -137.0998   274.1997   278.1997   283.1769   278.3392
##
## tau^2 (estimated amount of total heterogeneity): 0.7711 (SE = 0.1664)
## tau (square root of estimated tau^2 value):      0.8781
## I^2 (total heterogeneity / total variability):   73.38%
## H^2 (total variability / sampling variability):  3.76
##
## Test for Heterogeneity:
## Q(df = 89) = 311.7118, p-val < .0001
##
## Model Results:
##
## estimate      se     tval  df    pval    ci.lb    ci.ub
##  -1.1448  0.1176  -9.7344  89  <.0001  -1.3785  -0.9111  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6

Forest Plot

Forest plot


No of events / total
Study
Distal
Proximal
Log Risk Ratio
Log Risk Ratio (95% CI)
Dadarwal et al., 2022 0 / 160 8 / 160 −2.83 [−5.68, 0.01]
Eid−Lidt et al., 2021 1 / 140 12 / 142 −2.47 [−4.50, −0.44]
Li, Zhang et al., 2022 1.16 / 29 12.67 / 38 −2.12 [−3.96, −0.28]
Koledinskiy et al., 2020 1 / 132 8 / 132 −2.08 [−4.14, −0.01]
Tu et al., 2023 1 / 56 7 / 54 −1.98 [−4.04, 0.08]
Pacchioni et al., 2021 1 / 213 7 / 213 −1.95 [−4.03, 0.14]
Ozkan et al., 2020 0 / 20 3 / 20 −1.95 [−4.85, 0.95]
Gupta et al., 2023 4 / 210 27 / 210 −1.91 [−2.94, −0.88]
Acar et al., 2023 5 / 350 33 / 350 −1.89 [−2.82, −0.96]
Sharma et al., 2020 10 / 485 63 / 485 −1.84 [−2.50, −1.19]
Chugh et al., 2021 5 / 263 8 / 282 −0.40 [−1.50, 0.70]
Aminian et al., 2022 5 / 650 6 / 657 −0.17 [−1.35, 1.01]
Tehrani et al., 2023 0 / 33 0 / 31 −0.06 [−3.95, 3.83]
Aoi et al., 2019 0 / 202 0 / 206 0.02 [−3.90, 3.93]
Lucreziotti et al., 2021 0 / 100 0 / 104 0.04 [−3.87, 3.95]
Noamen et al., 2023 4 / 125 3 / 125 0.29 [−1.19, 1.76]
Elbayoumi et al., 2020 0 / 140 0 / 188 0.29 [−3.62, 4.21]
Kozinski et al., 2023 11 / 200 4 / 200 1.01 [−0.12, 2.14]
All studies: t(43)=−9.56***, Q(43) = 52.30, I² = 34%, τ² = 0.18 −1.07 [−1.30, −0.85]
RCTs only: t(20)=−5.17***, Q(20) = 40.50**, I² = 55%, τ² = 0.44 −1.05 [−1.48, −0.63]
−6.0
−3.2
−0.5
2.2
5.0
favors distal
favors proximal

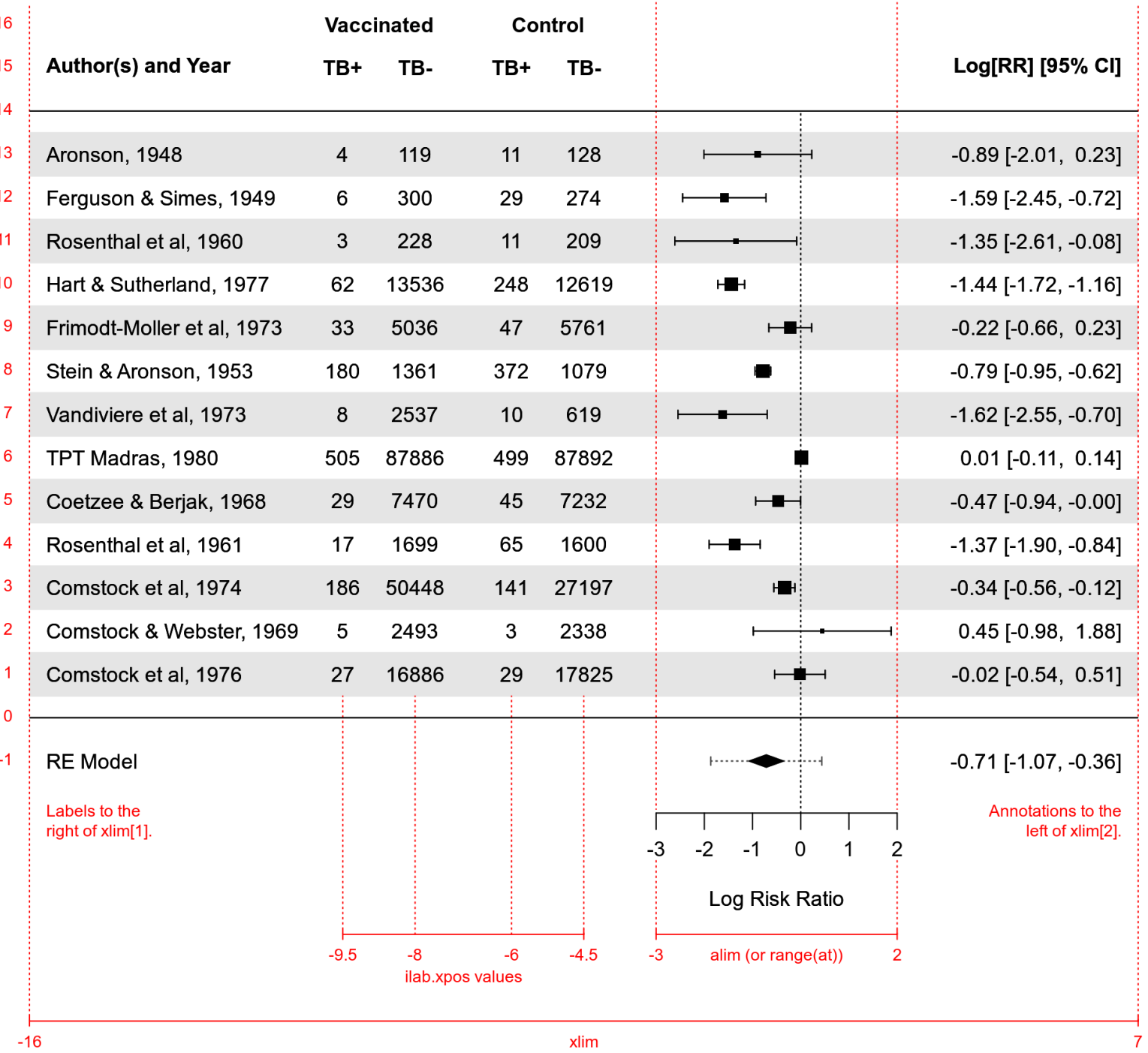
6a Exercise

R Markdown Script

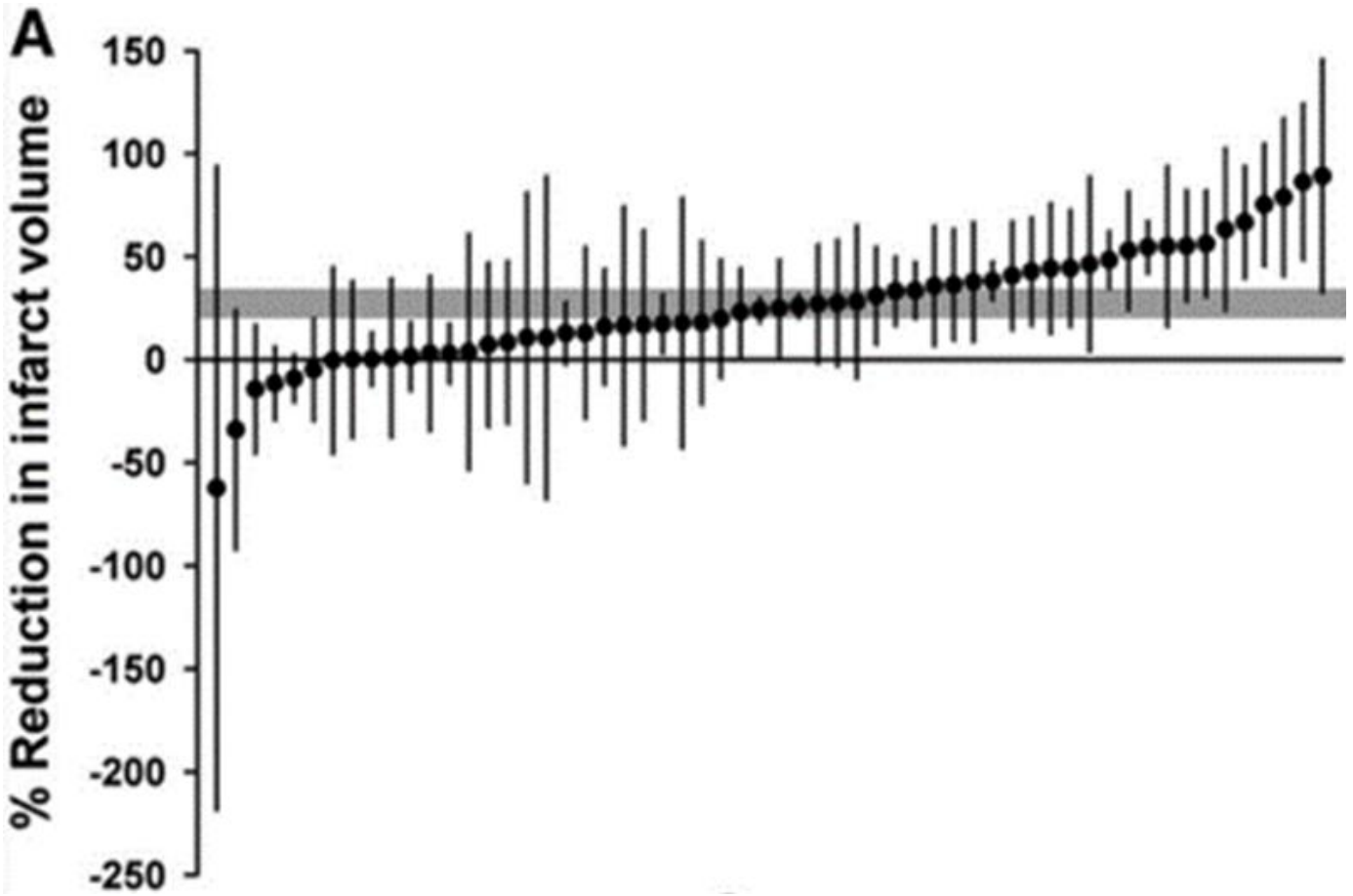
Create a forest plot

- **forest()**
- The **order** argument gives control over the sequence of studies
- The **slab** argument allows the inclusion of study labels

Modifying fores



Timber/caterpillar plot



Timber/caterpillar plot

```
forest(yi, vi, xlim=c(-2.5,3.5), ### adjust horizontal plot region limits
       order="obs", ### order by size of yi
       slab=NA, annotate=FALSE, ### remove study labels and annotations
       efac=0, ### remove vertical bars at end of CIs
       pch=19, ### changing point symbol to filled circle
       col="gray40", ### change color of points/CIs
       psize=2, ### increase point size
       cex.lab=1, cex.axis=1, ### increase size of x-axis title/labels
       lty=c("solid","blank")) ### remove horizontal line at top of plot
```

Welcome back!

Course structure

2nd Day
• *Heterogeneity in meta-analysis of animal studies*
• *Meta regression:*
• *Regression model & Explaining heterogeneity*
BREAK
• *Running models in R + Plotting*
• *Obstacles*
BREAK
• *Exploring the impacts of Risks of Bias in R*
• *Funnel plotting*
• *Workshop Conclusion & Wrap-up*

7

Heterogeneity

Heterogeneity

- "Statistical heterogeneity manifests itself in the… [study] effects being more different from each other than one would expect due to random error (chance) alone"

  Cochrane Handbook

  = unexpected amount of variation in effect sizes between studies

- We can calculate the amount of variation we would expect due to sampling error (ε)

- Heterogeneity is measured by:
  - Q-test statistic
  - τ^2
  - I^2

Heterogeneity: Q-test

- If the effect sizes are really homogeneous, then Q follows a chi-square distribution
- -> p-value of a significance test

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k$$

$$w_i = 1 / v_i$$

$$Q = \sum w_i (y_i - \hat{\theta})^2$$

Heterogeneity: τ^2

- τ^2 estimates the total amount of variability (heterogeneity) among the effect sizes
- heterogeneity may be due to random or systematic differences between the θi ’s
- τ^2 does not differentiate between sources
- Hard to interpret (is a raw variance): When is it large?

Heterogeneity: I^2

- I^2 estimates (in %) how much of the total variability in the effect size estimates is due to heterogeneity among the true effects
- Rough guidelines (Cochrane):
  - 0% to 40%: might not be important
  - 30% to 60%: moderate heterogeneity
  - 50% to 90%: substantial heterogeneity
  - 75% to 100%: considerable heterogeneity

$$I^2 = 100\% \times \frac{\hat{\tau}^2_{RE}}{\hat{\tau}^2_{RE} + s^2}$$

$$= 100\% \times \frac{Q - (k-1)}{Q}$$

$$s^2 = \frac{(k-1)\sum w_i}{\left(\sum w_i\right)^2 - \sum w_i^2}$$

7a Exercise

R Markdown Script

Check heterogeneity

- Run **rma()** again
- Adding heterogeneity measures to the forest plot

Example: Heterogeneity

```
##
## Random-Effects Model (k = 90; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -137.0998   274.1997   278.1997   283.1769   278.3392
##
## tau^2 (estimated amount of total heterogeneity): 0.7711 (SE = 0.1664)
## tau (square root of estimated tau^2 value):      0.8781
## I^2 (total heterogeneity / total variability):   73.38%
## H^2 (total variability / sampling variability):  3.76
##
## Test for Heterogeneity:
## Q(df = 89) = 311.7118, p-val < .0001
##
## Model Results:
##
## estimate      se     tval  df    pval    ci.lb    ci.ub
##  -1.1448  0.1176  -9.7344  89  <.0001  -1.3785  -0.9111  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Heterogeneity – a bad thing?

- In clinical MAs, heterogeneity is traditionally a nuisance

- By understanding the sources of heterogeneity, we can better understand mechanisms at work
- Especially important in the light of translational failure in preclinical research

- Heterogeneity might thus be a desirable property of our MA
- Heterogeneity in experimental setups increases replicability (Usui et al, 2021)

Usui T, Macleod MR, McCann SK, Senior AM, Nakagawa S. Meta-analysis of variation suggests that embracing variability improves both replicability and generalizability in preclinical research. PLoS Biology. 2021 May 19;19(5):e3001009.

8

Exploring Heterogenity - Meta-Regression

Heterogeneity in preclinical studies

- Heterogeneity is typically greater in preclinical than in clinical MAs

- Study-level variables that may influence the direction/size of the outcome
  - substantive variables (characteristics of the treatment, context, subjects)
  - methodological variables (e.g., randomized versus non-randomized study)
  - extrinsic variables (e.g., publication year, published/unpublished)

- In preclinical data:
  - Study samples (e.g. species, sex)
  - Interventions or outcomes (dose, outcome types)
  - Methodology: outcome measures used, quality etc.

Meta-Regression

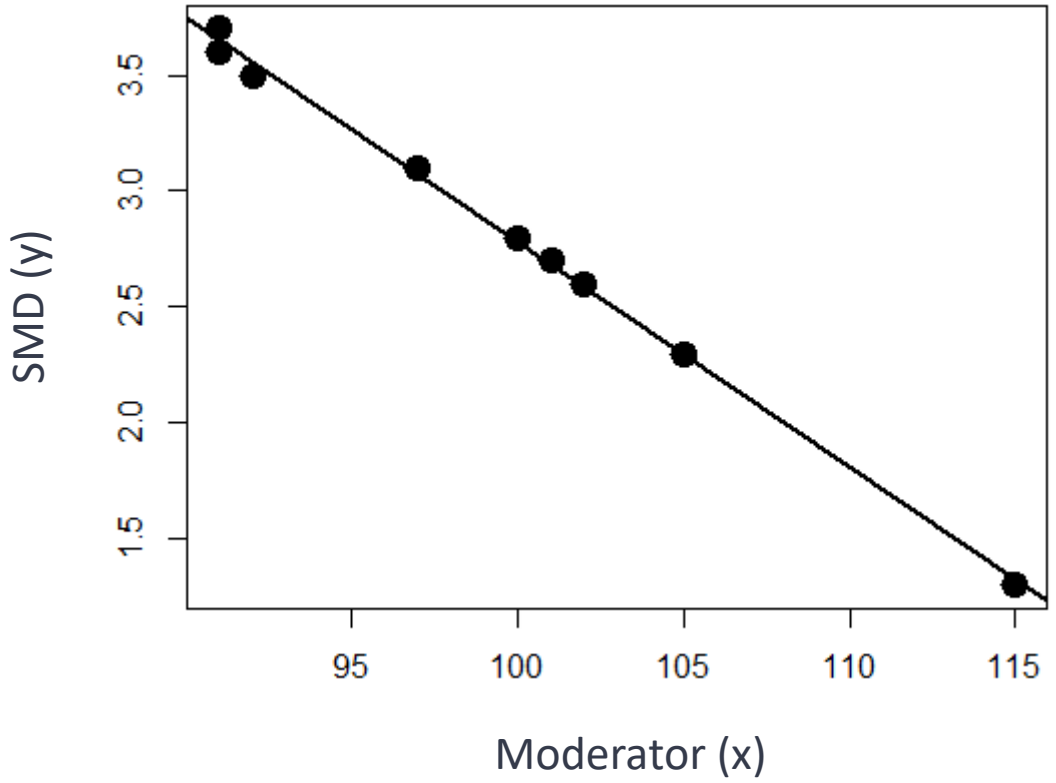
- Using regression analysis to evaluate the relevance of these study-specific features for the found effect
- The variables are also called “**moderators**” of the effect

Intro to Regression

Describing a scatter plot:

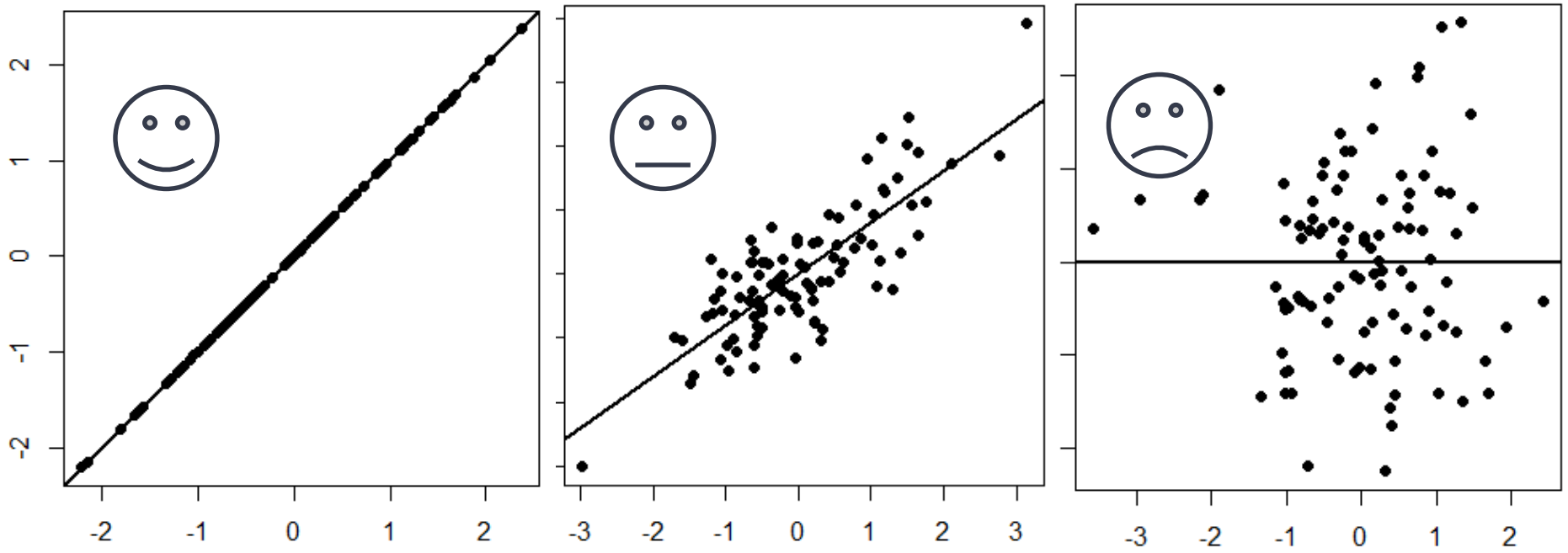
$$\hat{y} = a + b \times x$$

- a= intercept
- b = slope



Intro to Regression

$R^2 = 1$ $R^2 = 0.64$ $R^2 = 0$



Meta-Regression

A linear regression is described by:

- An equation for the line: $\hat{y} = a + b \times x$
- R^2 as measure of quality: How much variance in y can be predicted by x?

- I^2 is also a variance proportion!
- -> How much variance (I^2) can be explained by the moderator?

8a Exercise

R Markdown Script

Meta-regression

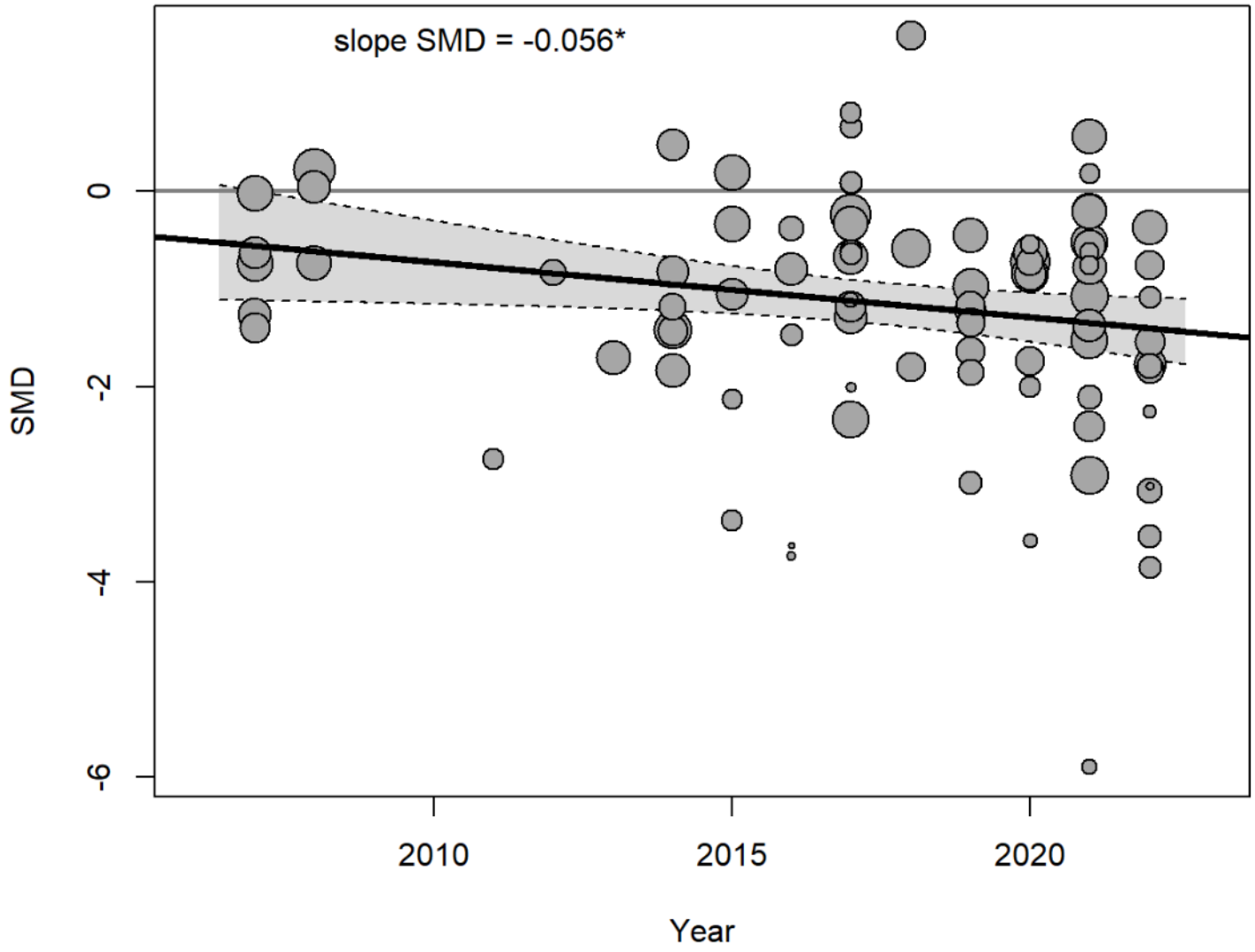
- rma(), once again
- Adding a moderator: year of publication

Meta-Regression

```
##
## Mixed-Effects Model (k = 90; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -133.7690   267.5381   273.5381   280.9701   273.8238
##
## tau^2 (estimated amount of residual heterogeneity):     0.7199 (SE = 0.1594)
## tau (square root of estimated tau^2 value):             0.8484
## I^2 (residual heterogeneity / unaccounted variability): 71.85%
## H^2 (unaccounted variability / sampling variability):   3.55
## R^2 (amount of heterogeneity accounted for):            6.64%
##
## Test for Residual Heterogeneity:
## QE(df = 88) = 289.5407, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 4.9019, p-val = 0.0268
##
## Model Results:
##
##          estimate       se     zval    pval    ci.lb     ci.ub
## intrcpt  112.6325  51.3878   2.1918  0.0284  11.9143  213.3507  *
## year      -0.0564   0.0255  -2.2140  0.0268  -0.1063   -0.0065  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plotting Meta-Regression

- Bubble plot:



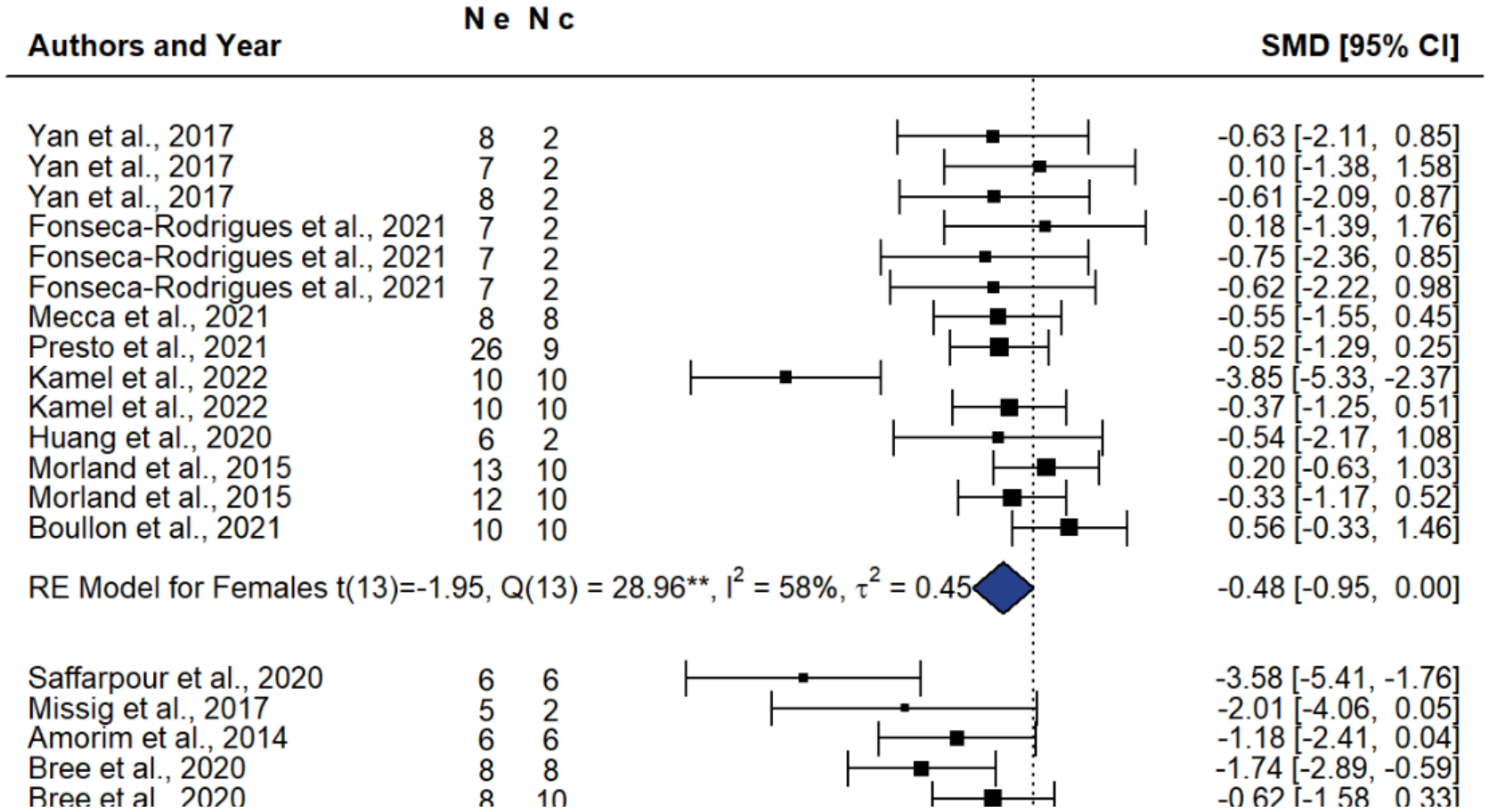
Meta-Regression with binary predictors

- Nothing changes, however, predictors have to be coded as factor (or 0 and 1) in R

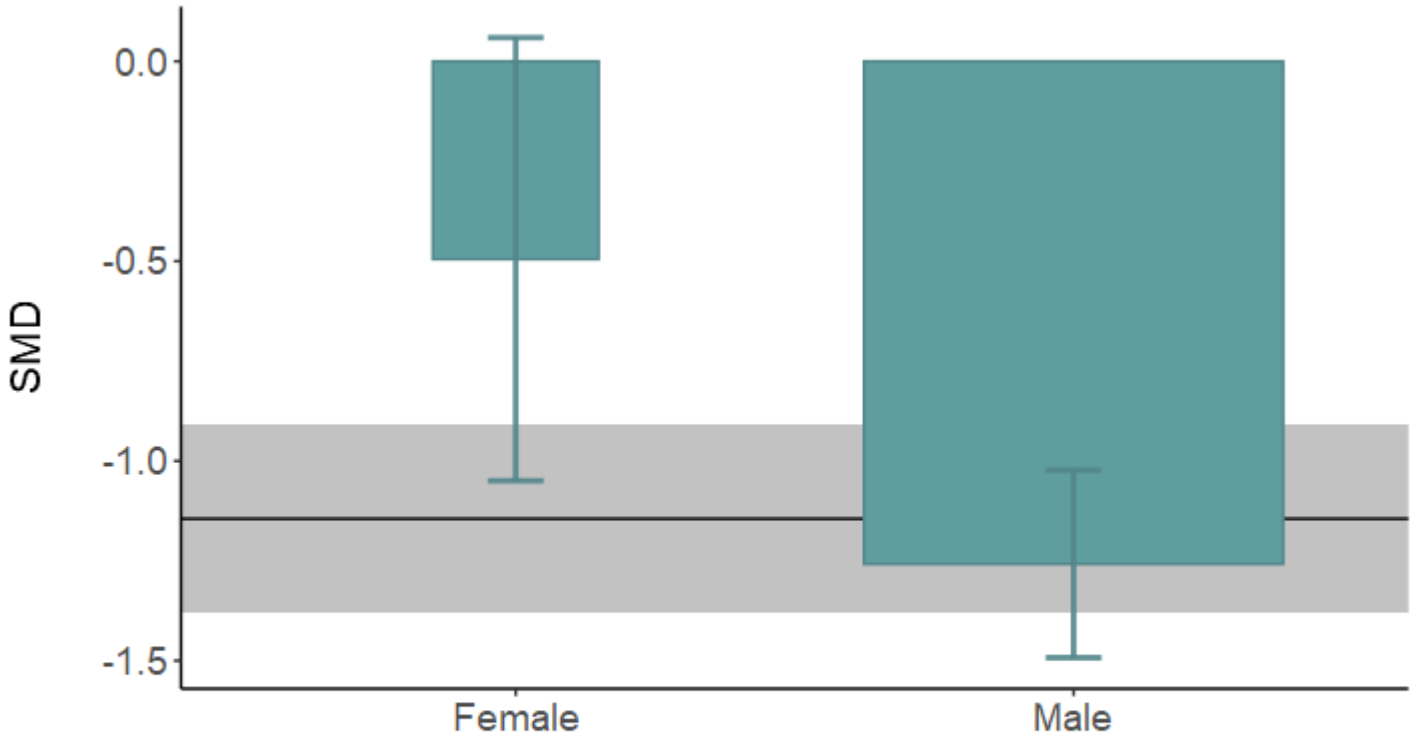
```
##
## Mixed-Effects Model (k = 89; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -132.2240   264.4480   270.4480   277.8458   270.7372
##
## tau^2 (estimated amount of residual heterogeneity):     0.7227 (SE = 0.1611)
## tau (square root of estimated tau^2 value):             0.8501
## I^2 (residual heterogeneity / unaccounted variability): 71.86%
## H^2 (unaccounted variability / sampling variability):   3.55
## R^2 (amount of heterogeneity accounted for):            8.43%
##
## Test for Residual Heterogeneity:
## QE(df = 87) = 293.4408, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 6.1679, p-val = 0.0130
##
## Model Results:
##
##          estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt   -0.4952  0.2829  -1.7503  0.0801  -1.0497   0.0593  .
## sexMale   -0.7629  0.3072  -2.4835  0.0130  -1.3649  -0.1608  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sub-group Analysis (less used in preclinical)

- Forest plot (and meta-analysis) separated by group:



Plotting categorical moderators



The horizontal grey bar represents the 95% CIs of the global estimate.
The width of the bars is proportional to the number of animals in each subgroup.

8b Exercise

R Markdown Script

Meta-regression

- Run meta-regression for the **strain** variable

Obstacles in Meta-Regression

- Small number of studies: Run only if there are > 10 studies (?)
- Grouping categories: reduce number of categories

8

Detecting Small Study Effects (publication bias)

Publication Bias

- Publication bias occurs when the results of published and unpublished studies differ systematically
- Neutral and negative studies:
    - Take longer to publish
    - Remain unpublished
    - Less likely to be identified in systematic review
    - Leads to the overstatement of efficacy in meta-analysis
- Selective outcome reporting
- Selective analysis reporting

Publication Bias

Small study effects assumptions:

- Larger studies with more commitment of resources and time → more likely to be published
- Smaller studies have greater risk of generating non-significant findings
- Smaller studies therefore are more likely to "remain in the file drawer"

Funnel plot

- In the absence of bias and between study heterogeneity, the scatter will be due to sampling variation alone and the plot will resemble a symmetrical inverted funnel

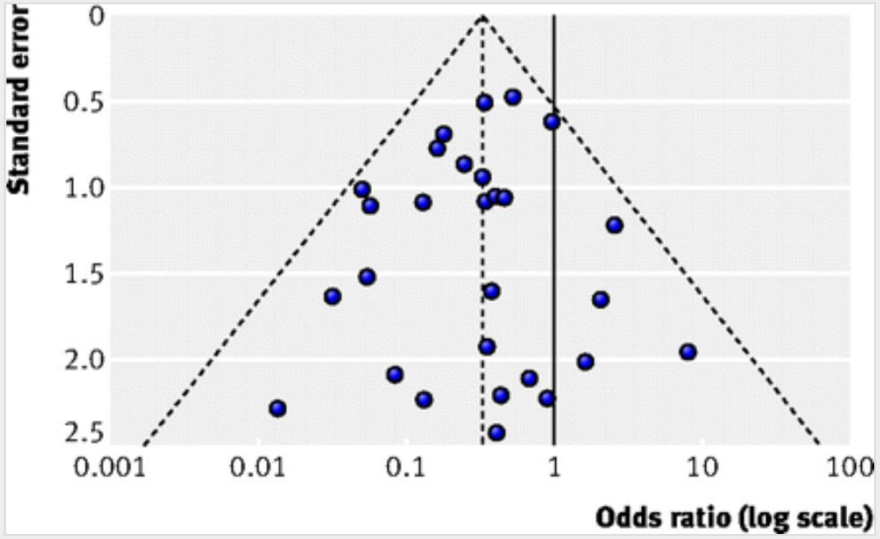
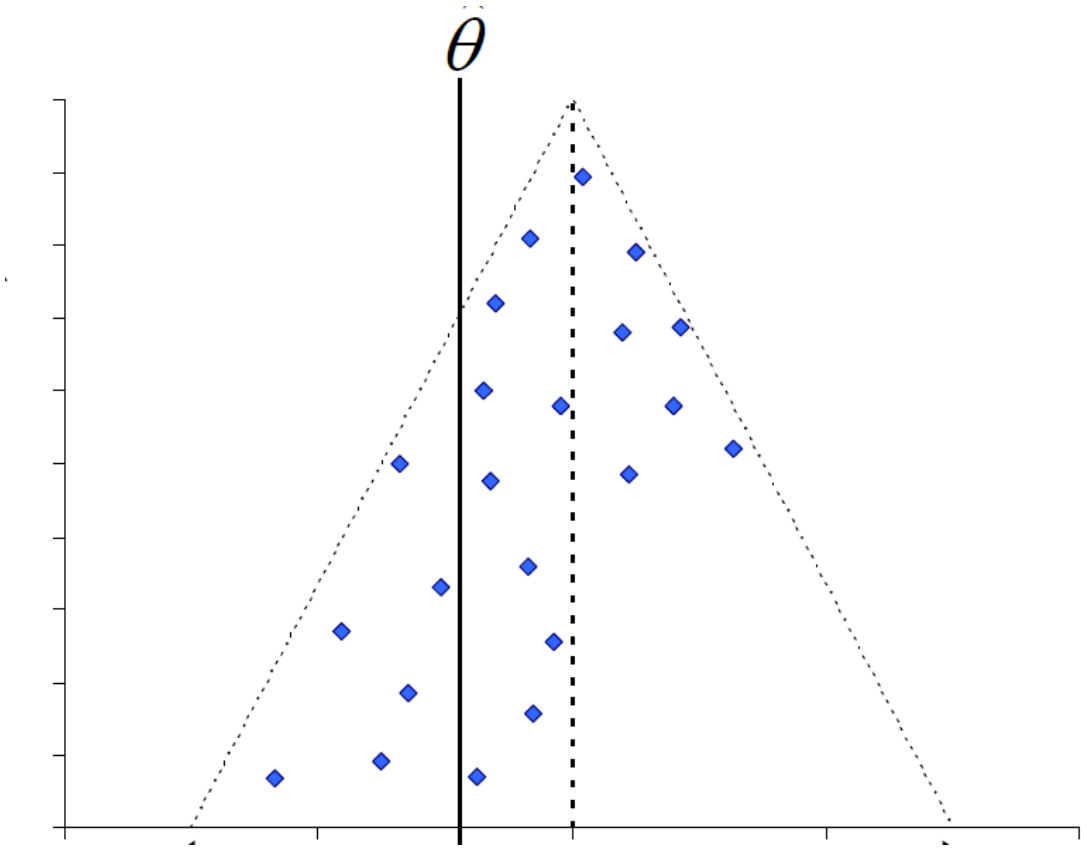


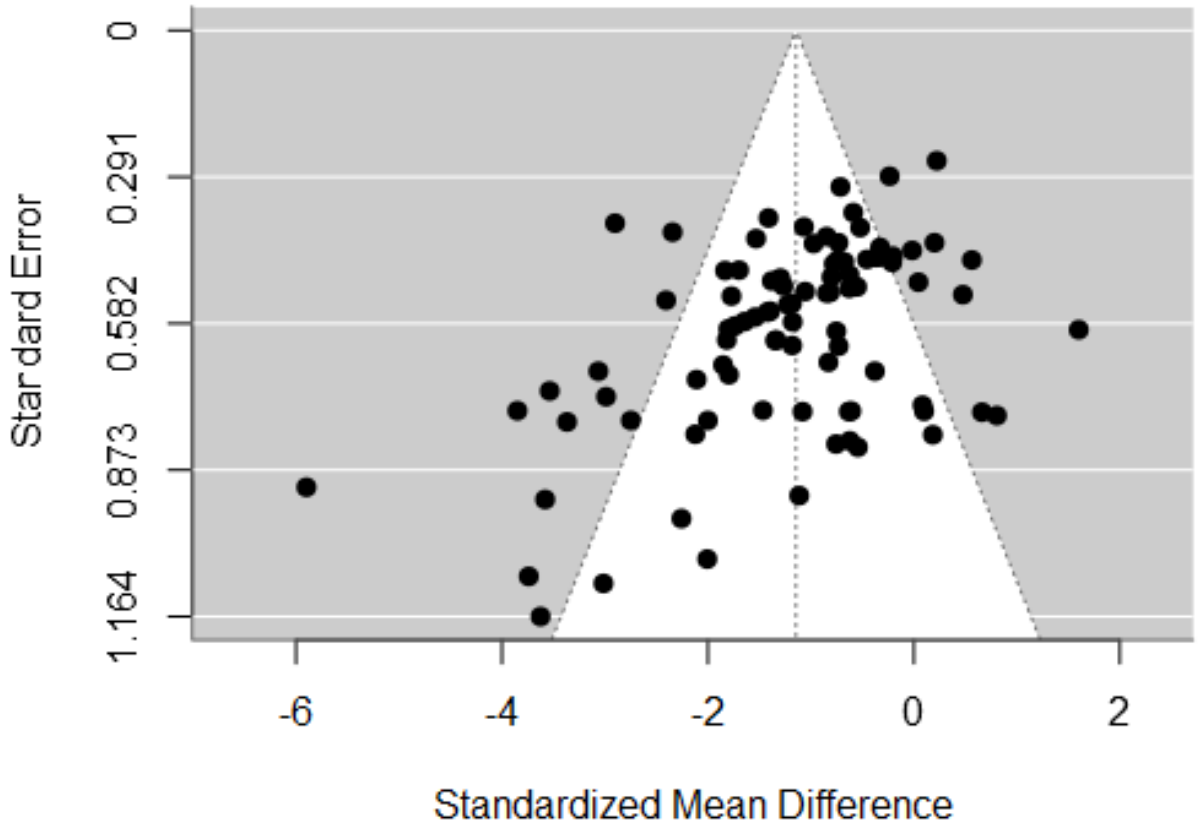
Fig 1 Example of symmetrical funnel plot. The outer dashed lines indicate the triangular region within which 95% of studies are expected to lie in the absence of both biases and heterogeneity (fixed effect summary log odds ratio±1.96×standard error of summary log odds ratio). The solid vertical line corresponds to no intervention effect

Funnel plot

- Problematic example:



Funnel plot for our data



Funnel plot in preclinical data

- In small samples and when using SMD, funnel plots tend to overrepresent differences (Zwetsloot et al.)
- Artificial biases

- Alternative: use other measure of precision but SE
- E.g. sample size (-> inverse of N)

Zwetsloot PP, Van Der Naald M, Sena ES, Howells DW, IntHout J, De Groot JA, Chamuleau SA, MacLeod MR, Wever KE. Standardized mean differences cause funnel plot distortion in publication bias assessments. elife. 2017 Sep 8;6:e24260.

9a Exercise

R Markdown Script

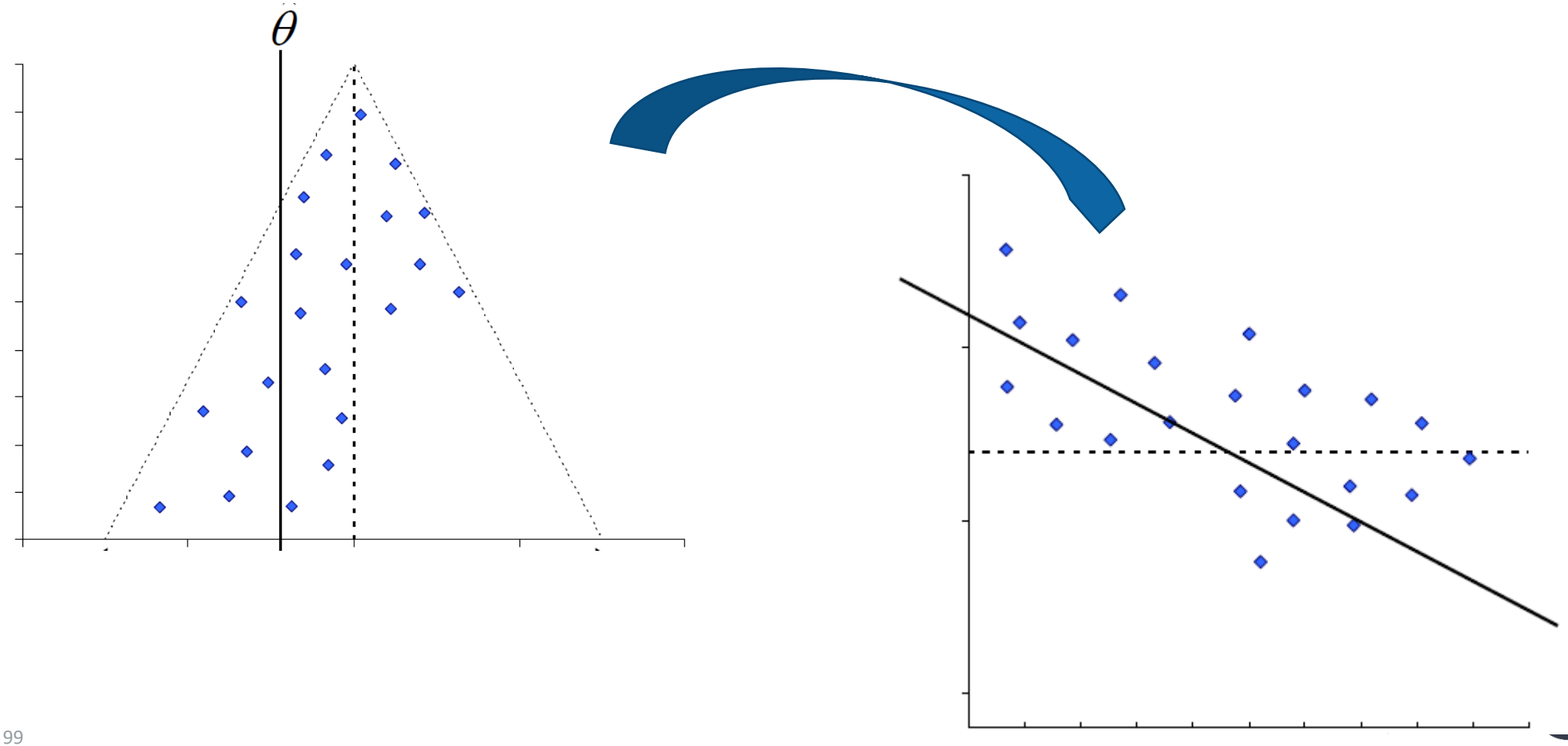
Funnel plotting

- Use the inverse:

  yaxis = "sqrtninv"

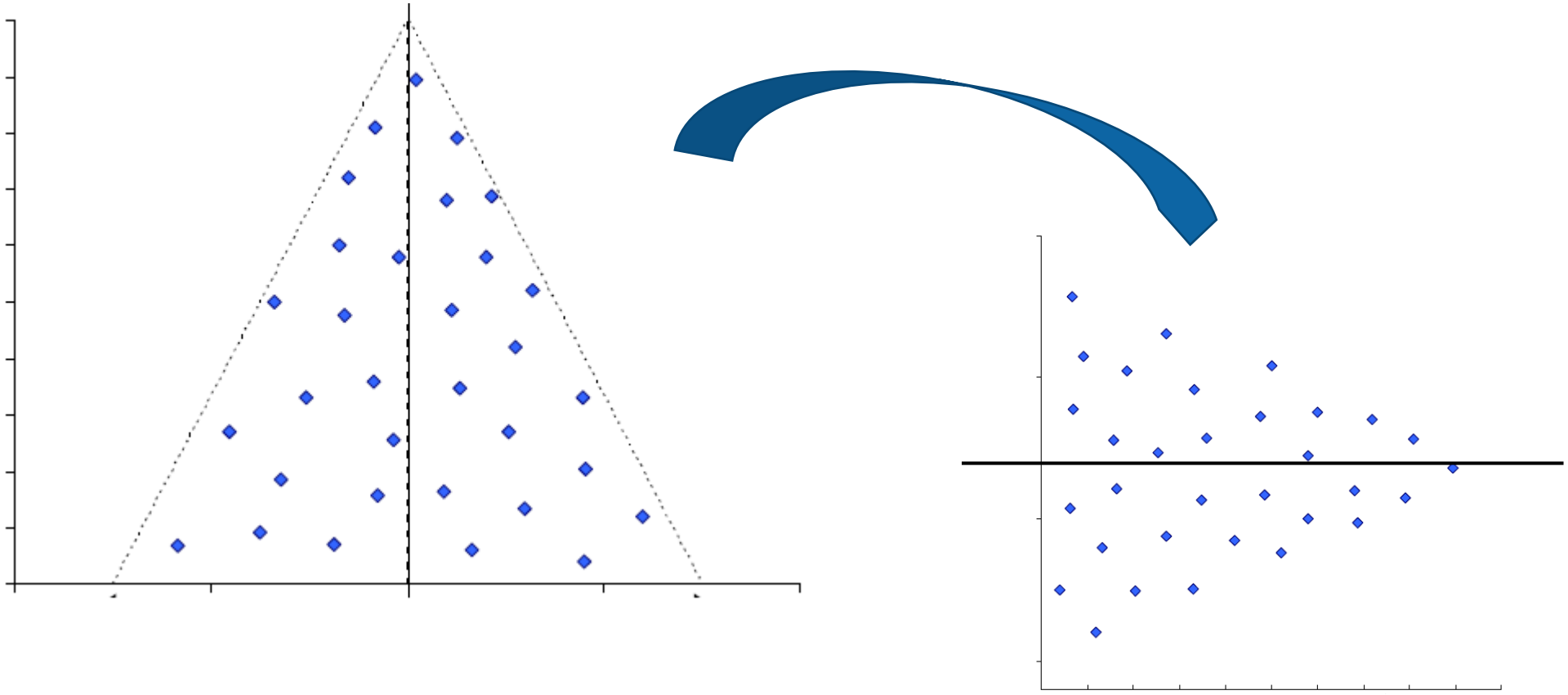
Egger's Regression

A regression line thru the flipped funnel plot (y = effect size, x = precision)



Egger's Regression

Without bias, the regression line is flat (= no slope, slope = 0)



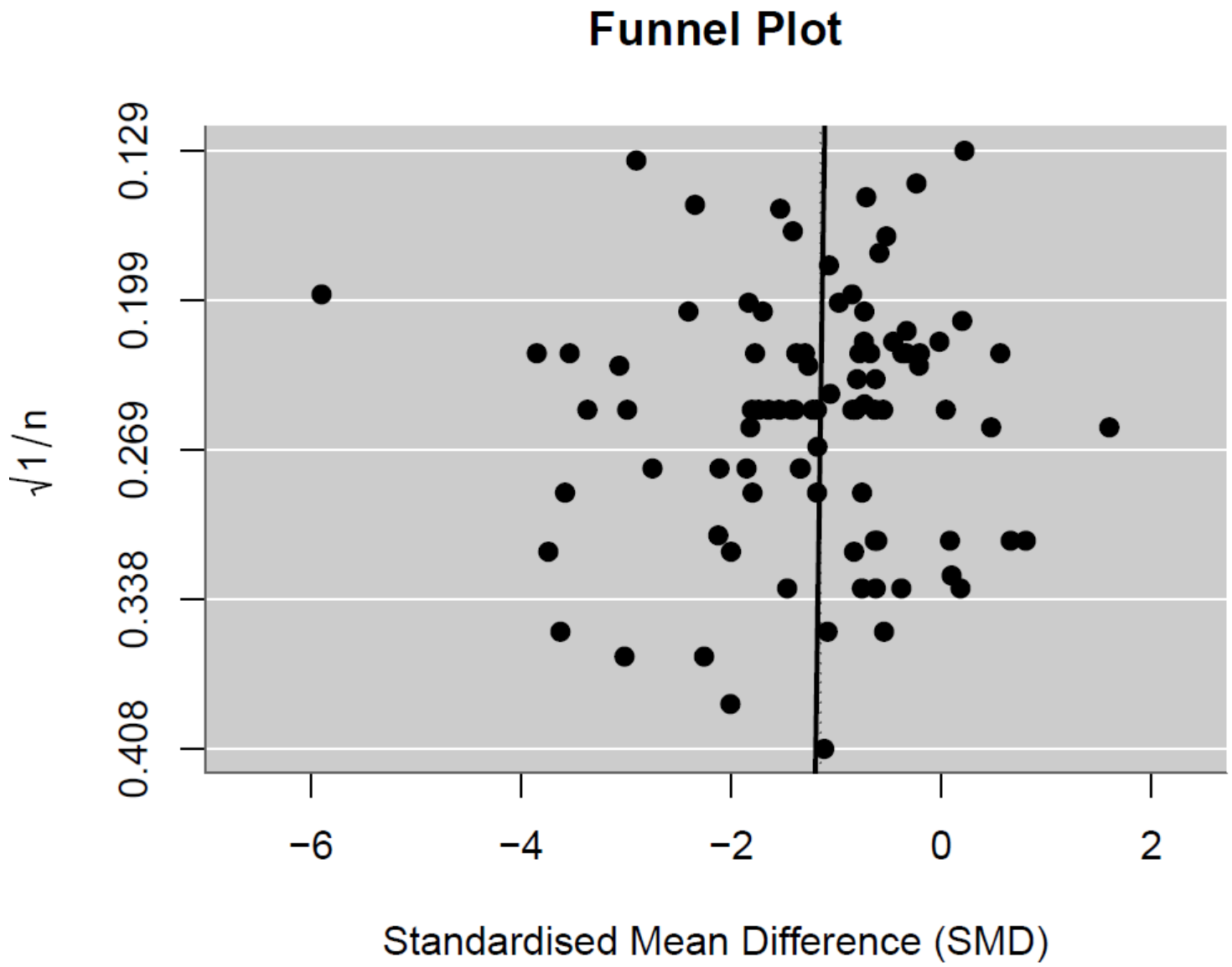
Exercise

R Markdown Script

Calculate an Egger‘s regression

- **regtest()**
- Add the line to the funnel (code is prepared)

Egger's Regression



Trim & Fill

- How many studies would be needed to balance an unbalanced funnel plot?
- -> simulate additional study results to fill the gaps