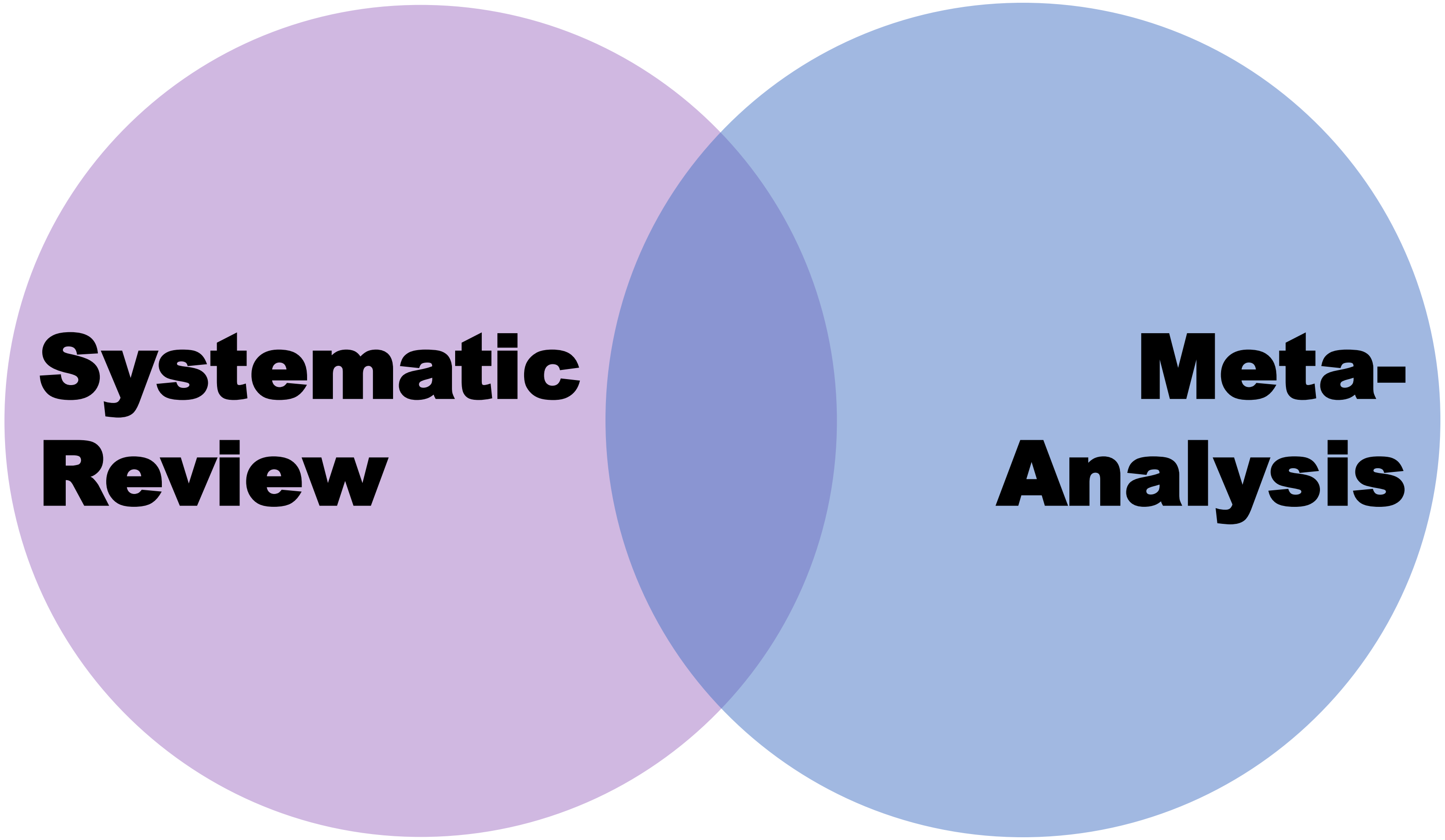
Introduction to Meta-Analysis

Samantha Estrada PhD

College of HEST Quantitative Methods Workshop Series

What is Meta-Analysis?

“the ***statistical analysis*** of a large collection of ***analysis results*** from individual studies for the purpose of integrating the findings”
(Glass, 1976)



Systematic Review

Meta-Analysis

“Data Collection”

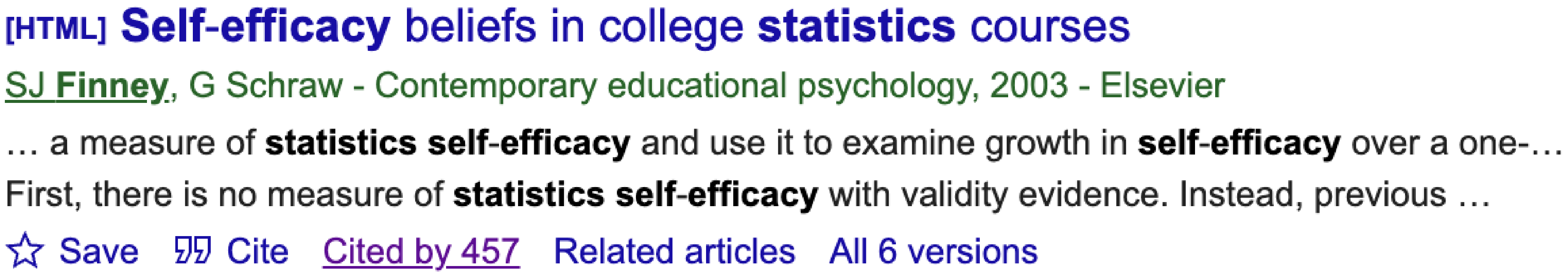
Where to start…

- Identify a topic
  - Be realistic
  - Team
- Keywords
  - Report on the keywords you used, “finney schraw current statistics self-efficacy”, “current statistics self-efficacy CSSE” “CSSE”
- Boolean logic
- *Statistical self-efficacy OR*
- *Statistical confidence OR*
- *Statistical anxiety OR*
- *Statistical self-belief OR*

- *Statistical education* OR*
- *Statistical learning OR*

“Data Collection”

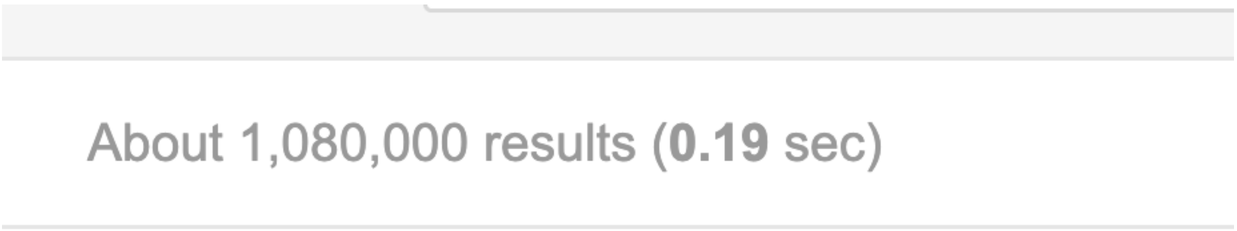
- Identify a popular database within your field to comb through the studies.
  - Google Scholar
  - Web of Science
  - PsycINFO
  - Pubmed/Medline
- Dates
  - (October 2025-December 2025)
  - “Natural” cutoffs



[HTML] **Self-efficacy** beliefs in college **statistics** courses

SJ **Finney**, G Schraw - Contemporary educational psychology, 2003 - Elsevier

… a measure of **statistics self-efficacy** and use it to examine growth in **self-efficacy** over a one-… First, there is no measure of **statistics self-efficacy** with validity evidence. Instead, previous …



About 1,080,000 results (**0.19** sec)

PICOT

- To include the studies there needs to be a similarity in the studies.
- Suggested models for this:
- PICOT (PCR Online, n. d.)
  - Populations
  - Intervention
  - Comparison
  - Outcome
  - Time frame

Meta-Analysis Type	What It Combines	Typical Use
Mean difference	Raw continuous scores	Common outcomes
OR/RR/RD	Binary outcomes	Clinical, epidemiology
Correlation/Fisher’s z	r values	Psych & education
Proportion	Prevalence rates	Public health
Reliability generalization	α, ω, ICC	Measurement studies

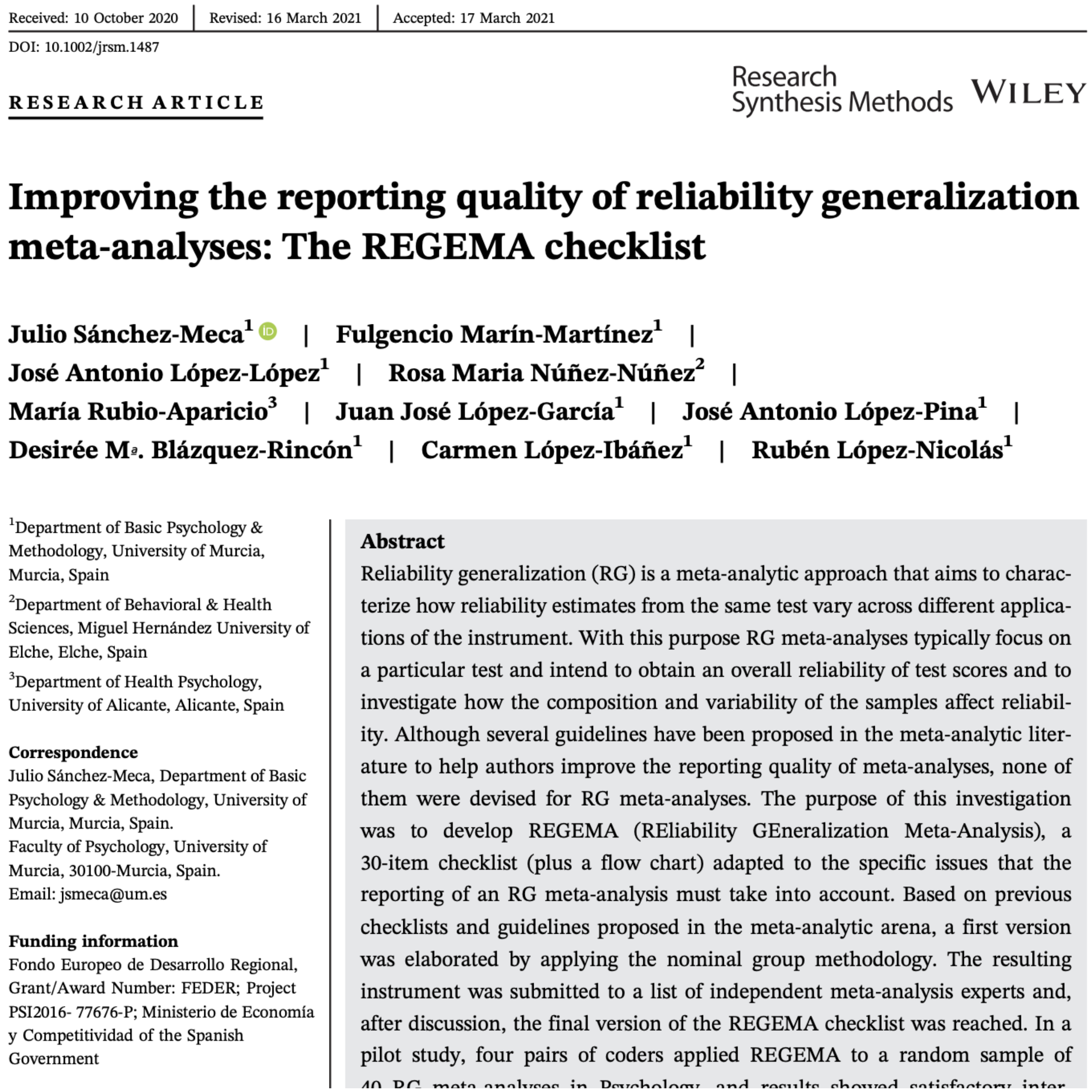
Guidelines

- PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA)
    - Tricco et al. (2018f). PRISMA extension for scoping reviews (PRISMA-ScR)
        - PRISMA-IPD for individual, participant data meta-analyses
        - PRISMA-NMA for network meta-analyses.

Reliability Generalization

- Reliability generalization is a type of meta-analysis
- Focus on the reliability estimates, usually Cronbach’s alpha, vary when the test is applied to different samples (Sanchez-Meca et al., 2019)

- REGEMA: REliability GEneralization Meta-Analysis
  - Sánchez-Meca et al. (2021)
    - Full text empirical references assessed
      - Excluded + reason
    - Records not recovered by ILL
    - Empirical references included in the meta-analysis



Received: 10 October 2020 | Revised: 16 March 2021 | Accepted: 17 March 2021

DOI: 10.1002/jrsm.1487

RESEARCH ARTICLE

Research Synthesis Methods WILEY

Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist

Julio Sánchez-Meca[1] | Fulgencio Marín-Martínez[1] | José Antonio López-López[1] | Rosa Maria Núñez-Núñez[2] | María Rubio-Aparicio[3] | Juan José López-García[1] | José Antonio López-Pina[1] | Desirée Mª. Blázquez-Rincón[1] | Carmen López-Ibáñez[1] | Rubén López-Nicolás[1]

[1] Department of Basic Psychology & Methodology, University of Murcia, Murcia, Spain

[2] Department of Behavioral & Health Sciences, Miguel Hernández University of Elche, Elche, Spain

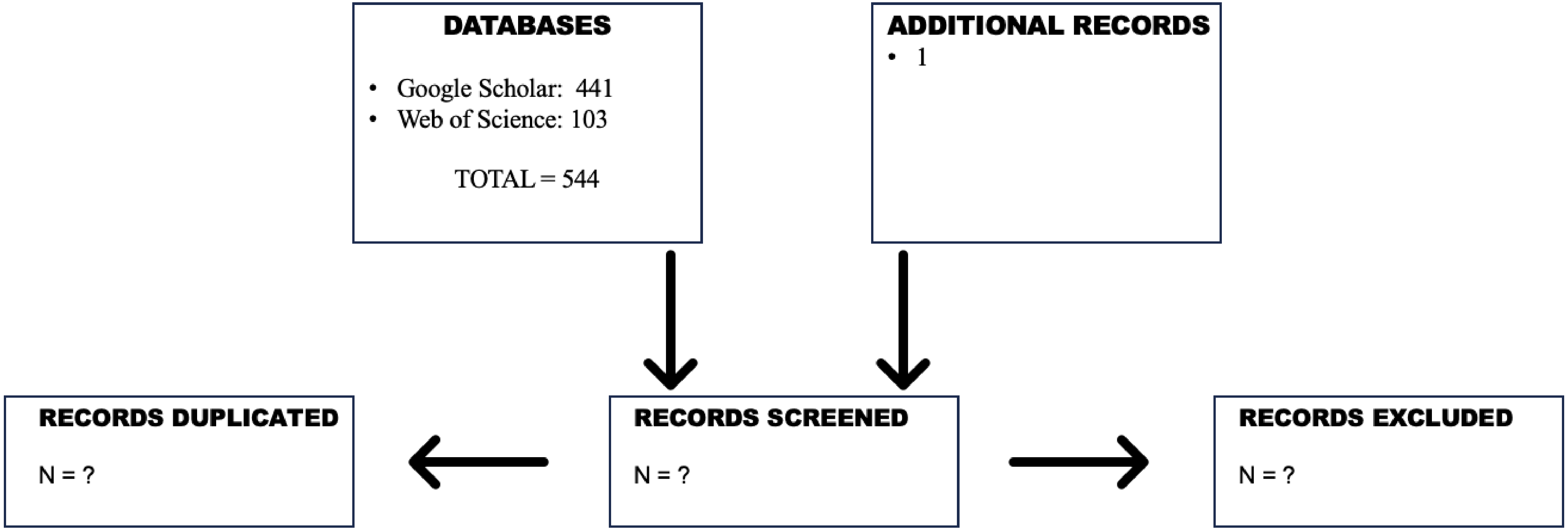
[3] Department of Health Psychology, University of Alicante, Alicante, Spain

Correspondence
Julio Sánchez-Meca, Department of Basic Psychology & Methodology, University of Murcia, Murcia, Spain.
Faculty of Psychology, University of Murcia, 30100-Murcia, Spain.
Email: jsmeca@um.es

Funding information
Fondo Europeo de Desarrollo Regional, Grant/Award Number: FEDER; Project PSI2016- 77676-P; Ministerio de Economía y Competitividad of the Spanish Government

Abstract

Reliability generalization (RG) is a meta-analytic approach that aims to characterize how reliability estimates from the same test vary across different applications of the instrument. With this purpose RG meta-analyses typically focus on a particular test and intend to obtain an overall reliability of test scores and to investigate how the composition and variability of the samples affect reliability. Although several guidelines have been proposed in the meta-analytic literature to help authors improve the reporting quality of meta-analyses, none of them were devised for RG meta-analyses. The purpose of this investigation was to develop REGEMA (REliability GEneralization Meta-Analysis), a 30-item checklist (plus a flow chart) adapted to the specific issues that the reporting of an RG meta-analysis must take into account. Based on previous checklists and guidelines proposed in the meta-analytic arena, a first version was elaborated by applying the nominal group methodology. The resulting instrument was submitted to a list of independent meta-analysis experts and, after discussion, the final version of the REGEMA checklist was reached. In a pilot study, four pairs of coders applied REGEMA to a random sample of 40 RG meta-analyses in Psychology, and results showed satisfactory inter-



DATABASES

- Google Scholar: 441
- Web of Science: 103

TOTAL = 544

ADDITIONAL RECORDS

- 1

RECORDS DUPLICATED

N = ?

RECORDS SCREENED

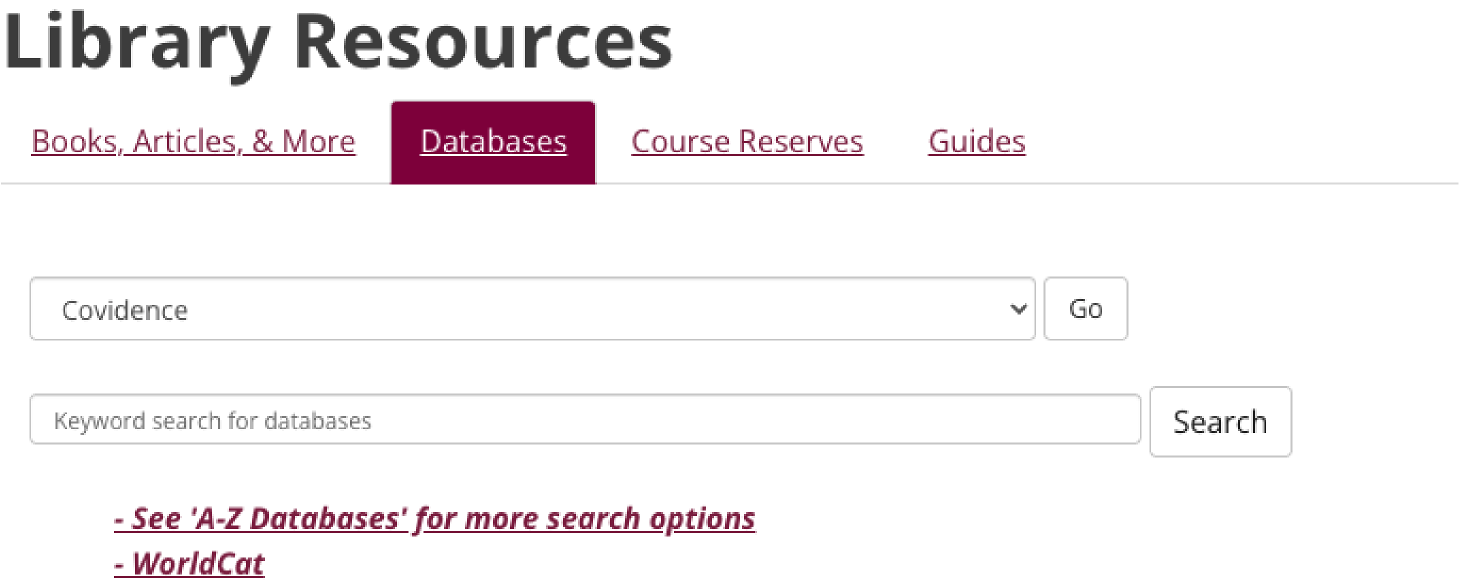
N = ?

RECORDS EXCLUDED

N = ?

Resource

Covidence



Covidence

- See 'A-Z Databases' for more search options

- WorldCat

“Data Collection”

- Emailing authors to share data.
  - Open Science Framework: https://osf.io/
  - Journal and/or universities databases

- Store digital copies in a reliable place. Things disappear from the internet.
  - Journal Articles, conference proceedings, posters, etc.

Effect Size

- Effect sizes are a statistical measure that attempts to represent the magnitude or strength of the relationship (Cohen, 1977).
- Each statistical method will have its own effect size.

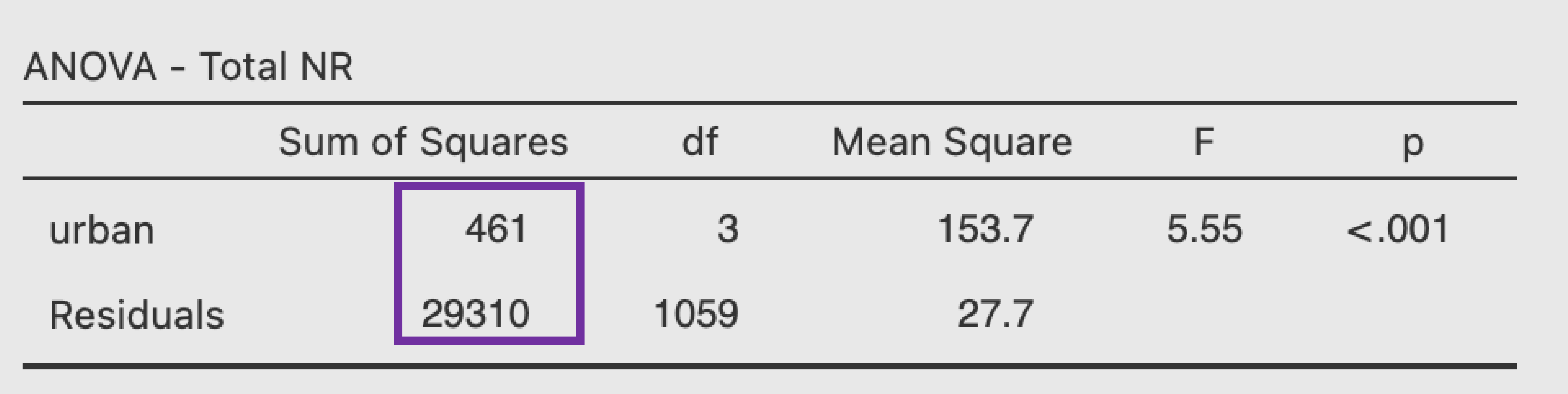
	Effect Size
T-test	Cohen’s d, Hedges’s g
ANOVA	η^2, η_p^2
Correlation/Regression	*r*, R^2
Chi-squares	Cramer’s V, ϕ

Effect sizes: Common Issues

- Not reported or reported from previous study.
- Calculate your own effect size from article information

Example

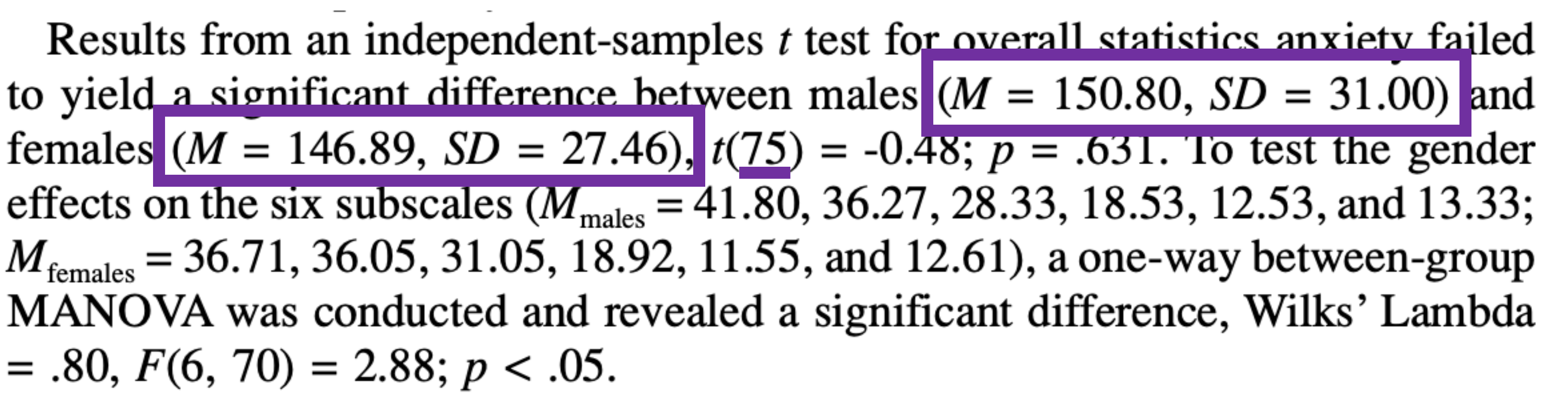
$$\eta_p^2 = \frac{SS_{EFFECT}}{SS_{EFFECT} + SS_{RESIDUALS}} \,, \eta = \frac{SS_{EFFECT}}{SS_{RESIDUALS}}$$



ANOVA - Total NR

	Sum of Squares	df	Mean Square	F	p
urban	461	3	153.7	5.55	<.001
Residuals	29310	1059	27.7		

Example



Results from an independent-samples t test for overall statistics anxiety failed to yield a significant difference between males (M = 150.80, SD = 31.00) and females (M = 146.89, SD = 27.46), $t(\underline{75})$ = -0.48; p = .631. To test the gender effects on the six subscales (M_{males} = 41.80, 36.27, 28.33, 18.53, 12.53, and 13.33; M_{females} = 36.71, 36.05, 31.05, 18.92, 11.55, and 12.61), a one-way between-group MANOVA was conducted and revealed a significant difference, Wilks' Lambda = .80, F(6, 70) = 2.88; p < .05.

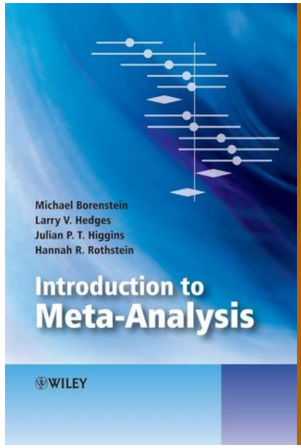
Hsiao & Chiang (2011)

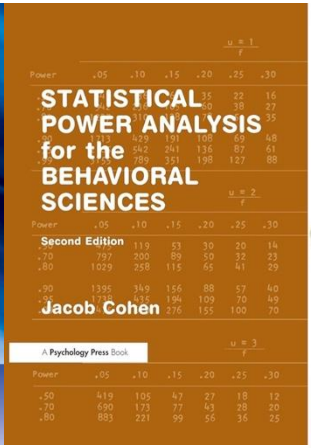
$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

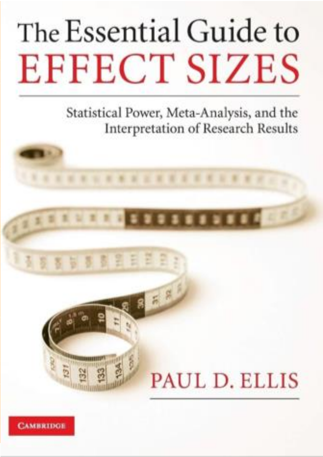
☑ Means and SD for each group

☑ Sample size

Best resource for effect sizes Cohen (1988) book, but I also recommend this (open access) paper by Durlak (2009)






The Essential Guide to EFFECT SIZES
Statistical Power, Meta-Analysis, and the Interpretation of Research Results
PAUL D. ELLIS
CAMBRIDGE

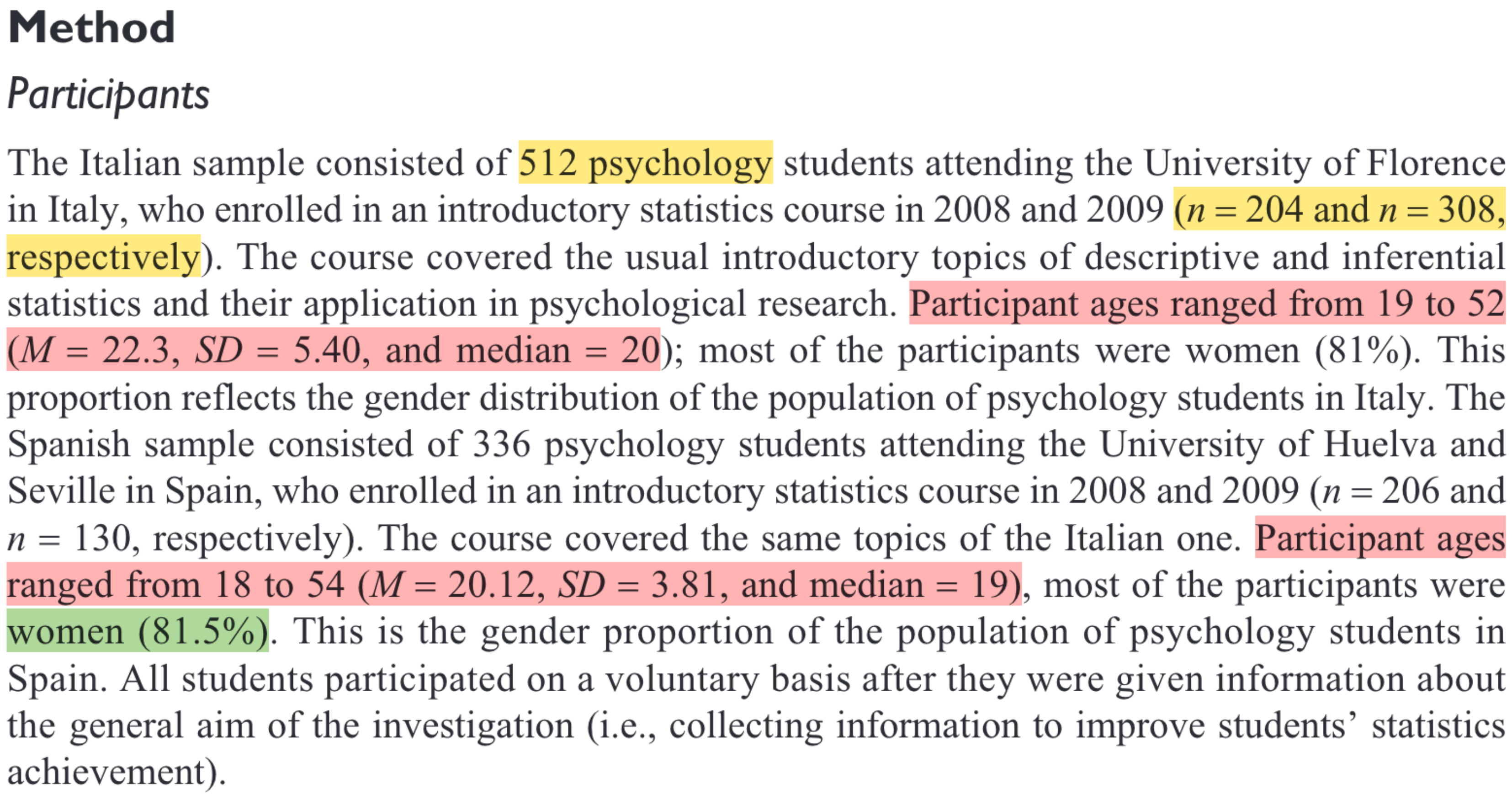
Resource

Effect Size

- Books
  - Borenstein et al. (2021) .
  - Cohen (1988).
  - Ellis (2010).
- Article
  - Durlak (2009)

Coding Process

Descriptive Information

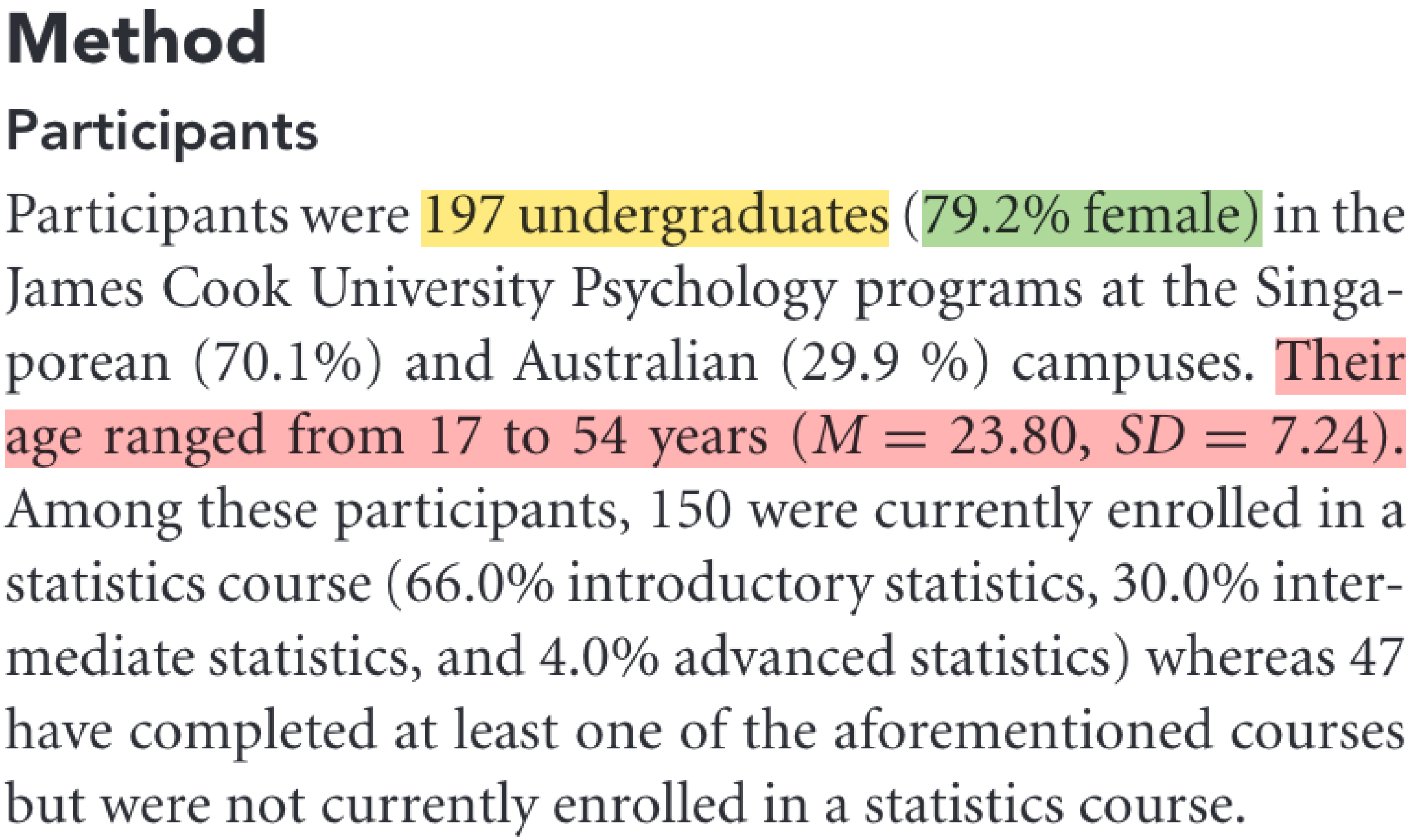


Method

Participants

The Italian sample consisted of 512 psychology students attending the University of Florence in Italy, who enrolled in an introductory statistics course in 2008 and 2009 ($n = 204$ and $n = 308$, respectively). The course covered the usual introductory topics of descriptive and inferential statistics and their application in psychological research. Participant ages ranged from 19 to 52 ($M = 22.3$, $SD = 5.40$, and median = 20); most of the participants were women (81%). This proportion reflects the gender distribution of the population of psychology students in Italy. The Spanish sample consisted of 336 psychology students attending the University of Huelva and Seville in Spain, who enrolled in an introductory statistics course in 2008 and 2009 ($n = 206$ and $n = 130$, respectively). The course covered the same topics of the Italian one. Participant ages ranged from 18 to 54 ($M = 20.12$, $SD = 3.81$, and median = 19), most of the participants were women (81.5%). This is the gender proportion of the population of psychology students in Spain. All students participated on a voluntary basis after they were given information about the general aim of the investigation (i.e., collecting information to improve students' statistics achievement).

Chiesi et al. (2011)



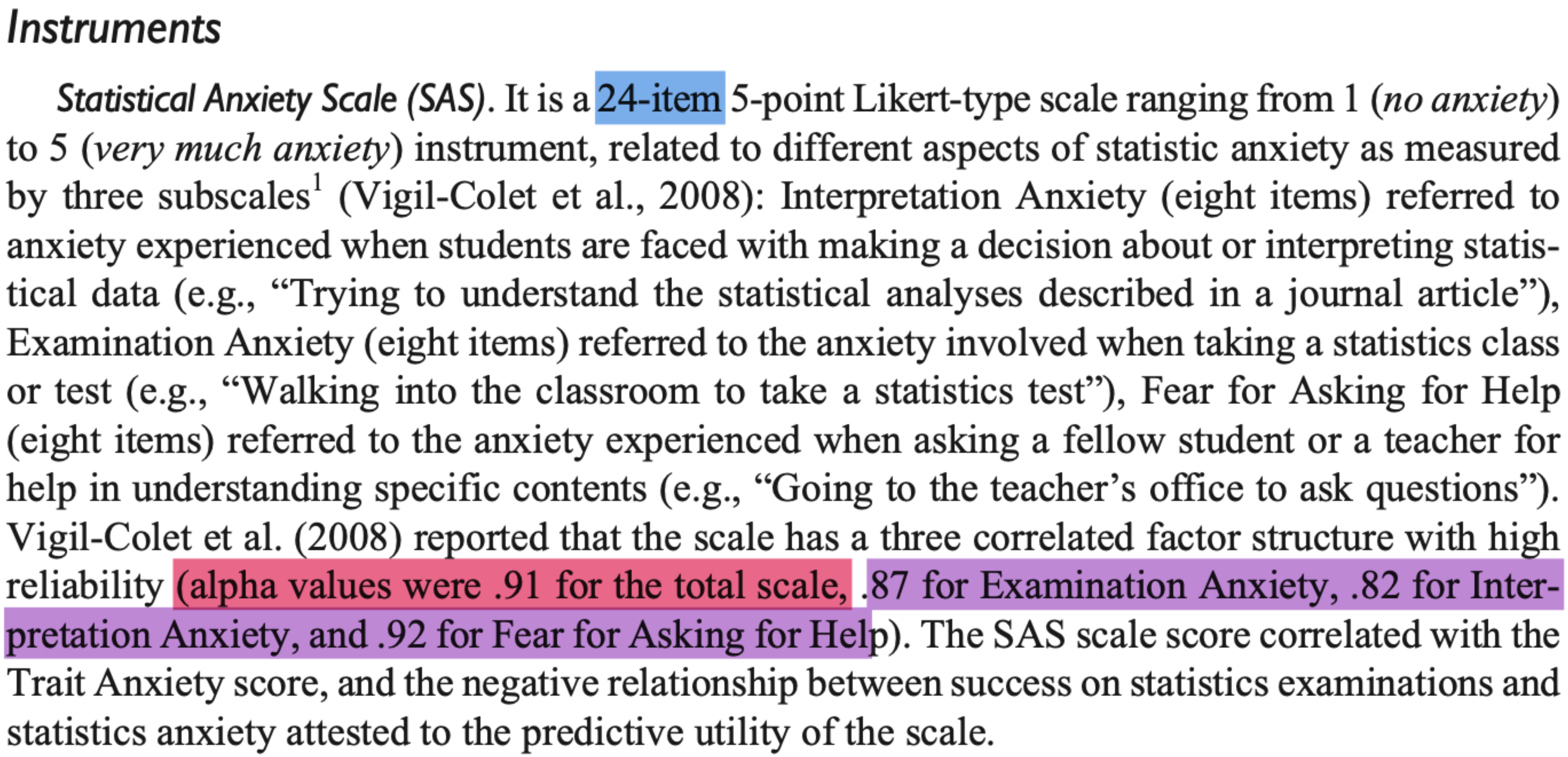
Method

Participants

Participants were 197 undergraduates (79.2% female) in the James Cook University Psychology programs at the Singaporean (70.1%) and Australian (29.9 %) campuses. Their age ranged from 17 to 54 years ($M = 23.80$, $SD = 7.24$). Among these participants, 150 were currently enrolled in a statistics course (66.0% introductory statistics, 30.0% intermediate statistics, and 4.0% advanced statistics) whereas 47 have completed at least one of the aforementioned courses but were not currently enrolled in a statistics course.

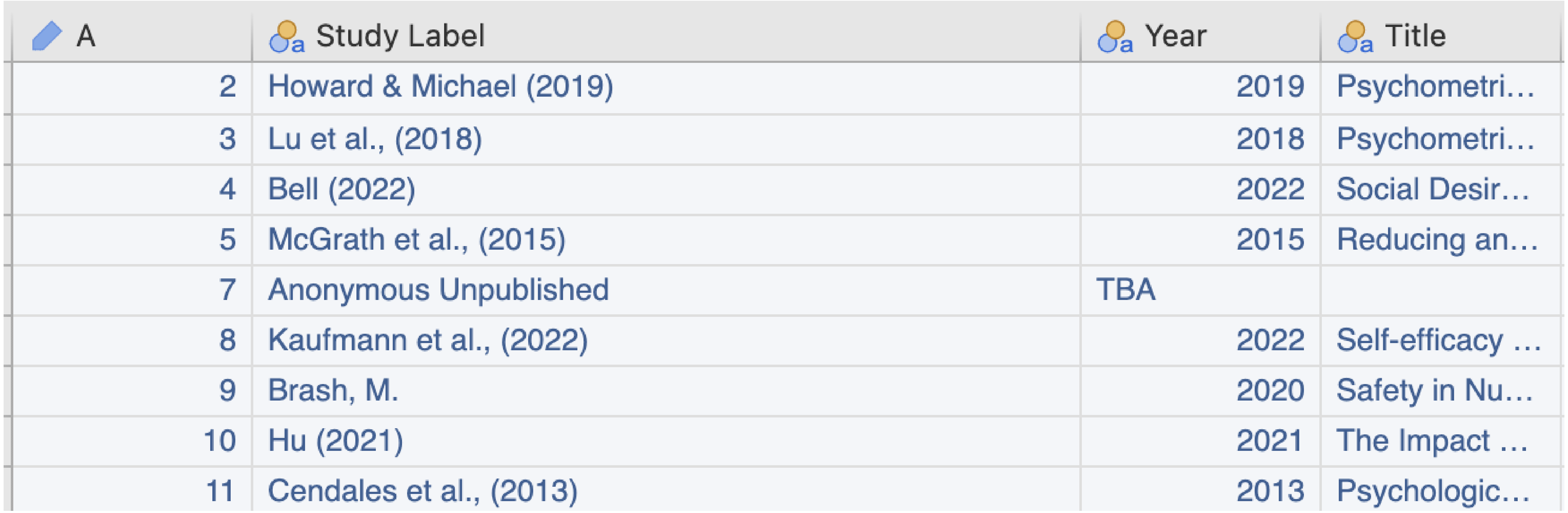
Drew & Chillon, 2014

Effect Size



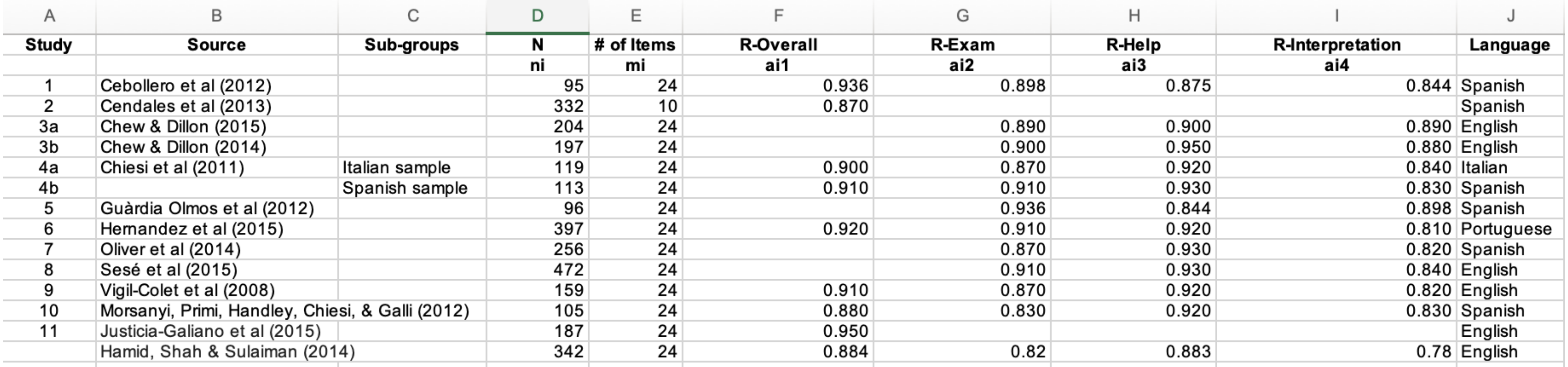
Instruments

Statistical Anxiety Scale (SAS). It is a 24-item 5-point Likert-type scale ranging from 1 (*no anxiety*) to 5 (*very much anxiety*) instrument, related to different aspects of statistic anxiety as measured by three subscales[1] (Vigil-Colet et al., 2008): Interpretation Anxiety (eight items) referred to anxiety experienced when students are faced with making a decision about or interpreting statistical data (e.g., "Trying to understand the statistical analyses described in a journal article"), Examination Anxiety (eight items) referred to the anxiety involved when taking a statistics class or test (e.g., "Walking into the classroom to take a statistics test"), Fear for Asking for Help (eight items) referred to the anxiety experienced when asking a fellow student or a teacher for help in understanding specific contents (e.g., "Going to the teacher's office to ask questions"). Vigil-Colet et al. (2008) reported that the scale has a three correlated factor structure with high reliability (alpha values were .91 for the total scale, .87 for Examination Anxiety, .82 for Interpretation Anxiety, and .92 for Fear for Asking for Help). The SAS scale score correlated with the Trait Anxiety score, and the negative relationship between success on statistics examinations and statistics anxiety attested to the predictive utility of the scale.

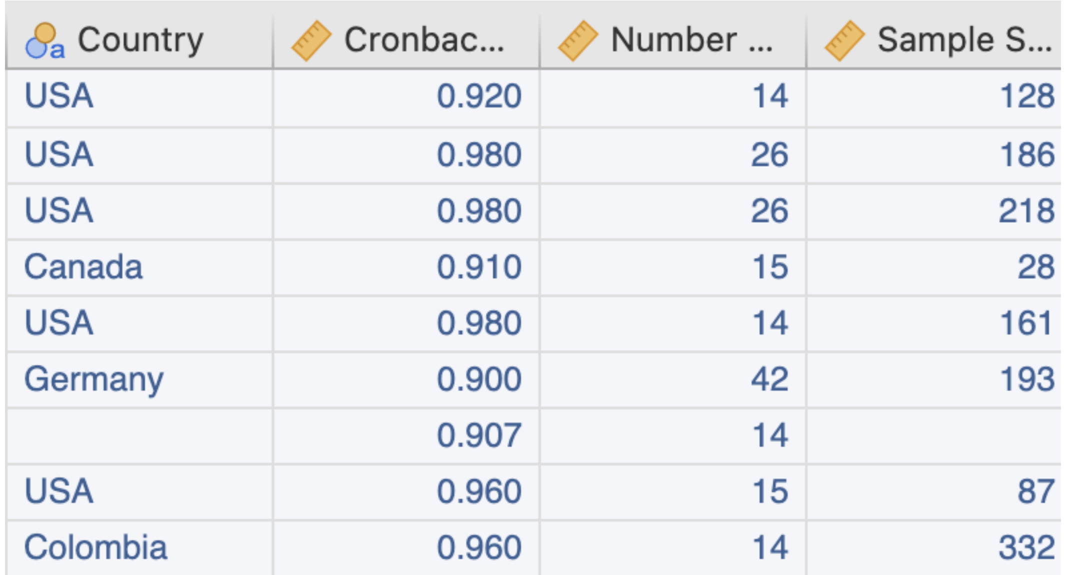


A	Study Label	Year	Title
2	Howard & Michael (2019)	2019	Psychometri…
3	Lu et al., (2018)	2018	Psychometri…
4	Bell (2022)	2022	Social Desir…
5	McGrath et al., (2015)	2015	Reducing an…
7	Anonymous Unpublished	TBA	
8	Kaufmann et al., (2022)	2022	Self-efficacy …
9	Brash, M.	2020	Safety in Nu…
10	Hu (2021)	2021	The Impact …
11	Cendales et al., (2013)	2013	Psychologic…

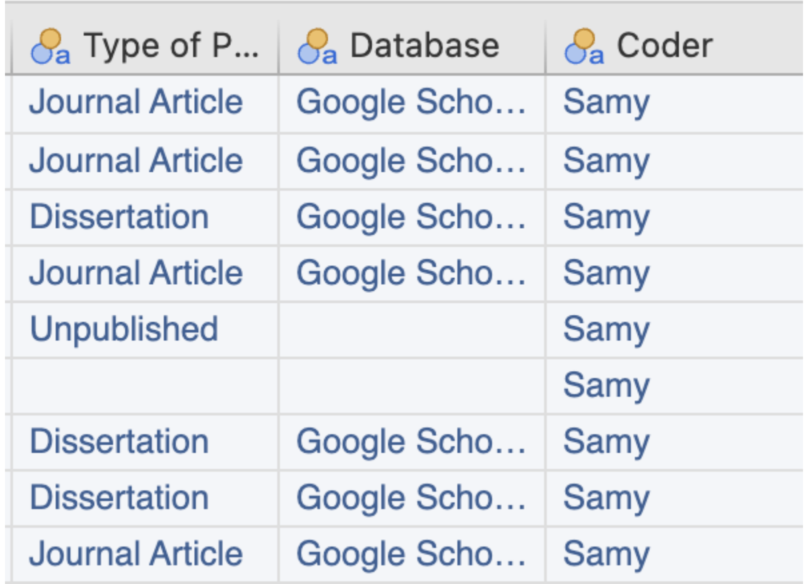
Coding Sheet Example



A	B	C	D	E	F	G	H	I	J
Study	**Source**	**Sub-groups**	**N**	**# of Items**	**R-Overall**	**R-Exam**	**R-Help**	**R-Interpretation**	**Language**
			ni	**mi**	**ai1**	**ai2**	**ai3**	**ai4**	
1	Cebollero et al (2012)		95	24	0.936	0.898	0.875	0.844	Spanish
2	Cendales et al (2013)		332	10	0.870				Spanish
3a	Chew & Dillon (2015)		204	24		0.890	0.900	0.890	English
3b	Chew & Dillon (2014)		197	24		0.900	0.950	0.880	English
4a	Chiesi et al (2011)	Italian sample	119	24	0.900	0.870	0.920	0.840	Italian
4b		Spanish sample	113	24	0.910	0.910	0.930	0.830	Spanish
5	Guàrdia Olmos et al (2012)		96	24		0.936	0.844	0.898	Spanish
6	Hernandez et al (2015)		397	24	0.920	0.910	0.920	0.810	Portuguese
7	Oliver et al (2014)		256	24		0.870	0.930	0.820	Spanish
8	Sesé et al (2015)		472	24		0.910	0.930	0.840	English
9	Vigil-Colet et al (2008)		159	24	0.910	0.870	0.920	0.820	English
10	Morsanyi, Primi, Handley, Chiesi, & Galli (2012)		105	24	0.880	0.830	0.920	0.830	Spanish
11	Justicia-Galiano et al (2015)		187	24	0.950				English
	Hamid, Shah & Sulaiman (2014)		342	24	0.884	0.82	0.883	0.78	English



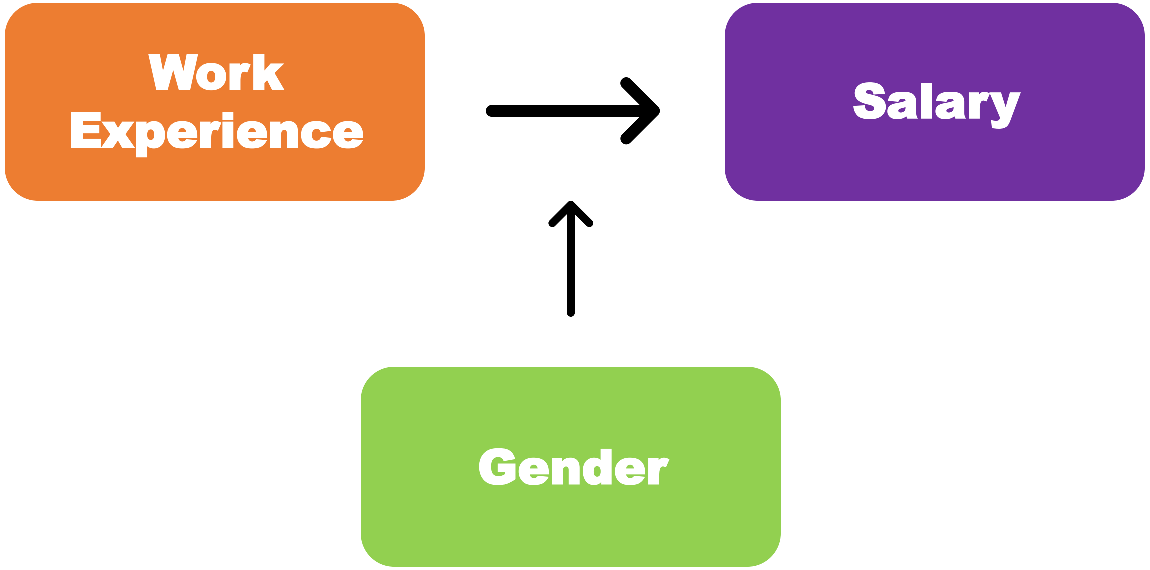
Country	Cronbac...	Number ...	Sample S...
USA	0.920	14	128
USA	0.980	26	186
USA	0.980	26	218
Canada	0.910	15	28
USA	0.980	14	161
Germany	0.900	42	193
	0.907	14	
USA	0.960	15	87
Colombia	0.960	14	332



Type of P...	Database	Coder
Journal Article	Google Scho...	Samy
Journal Article	Google Scho...	Samy
Dissertation	Google Scho...	Samy
Journal Article	Google Scho...	Samy
Unpublished		Samy
		Samy
Dissertation	Google Scho...	Samy
Dissertation	Google Scho...	Samy
Journal Article	Google Scho...	Samy

Moderator Variables

- Explains *when* or *for whom* X affects Y.



Conducting a Meta-Analysis

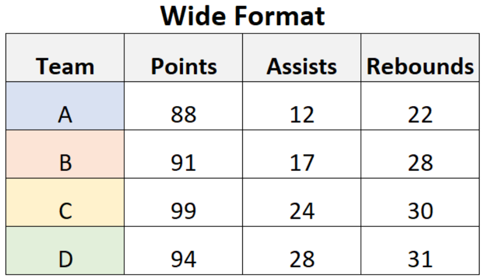
Resource

Software

- There is specialized software for meta-analysis:
  - Open Meta
  - Meta-Essentials
  - R, it’s free and open source
    - Meta, metafor, meta-package
  - Jamovi. Also free and open source.
    - Uses the same package, metafor.
  - SPSS

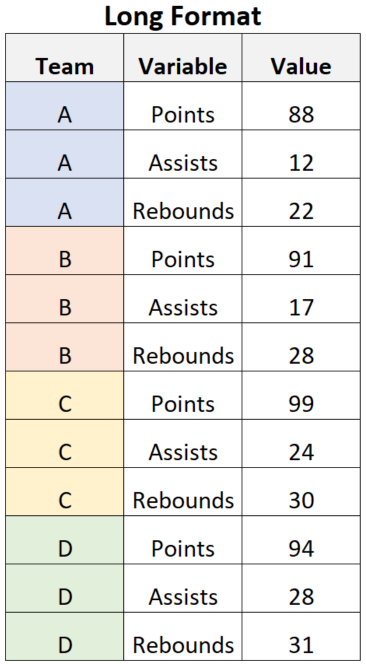
Interrater Agreement

- Measure of consistency between two (or more) raters.
- How to assess it? Cohen’s Kappa
- For there or more raters? Fleiss Kappa
- Data need to be in long format



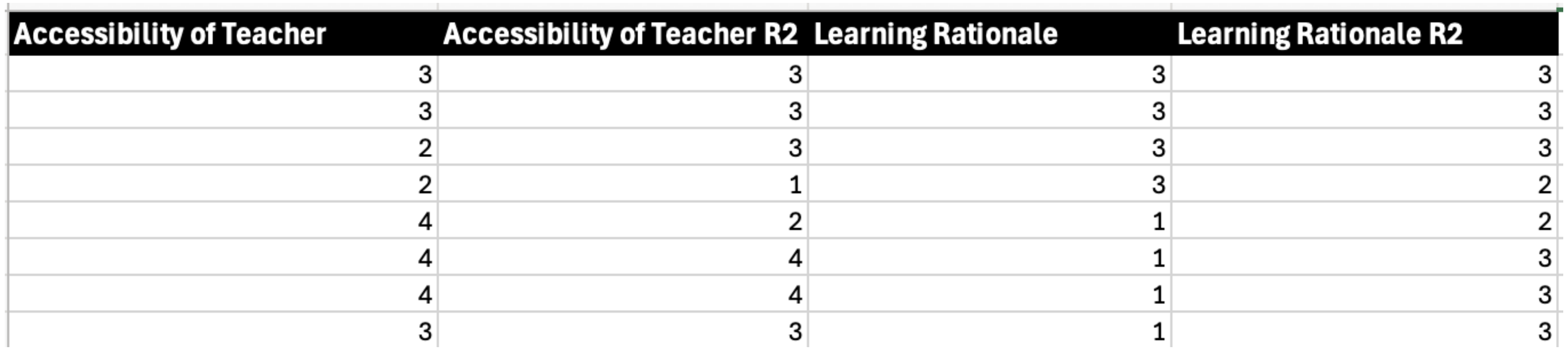
Wide Format

Team	Points	Assists	Rebounds
A	88	12	22
B	91	17	28
C	99	24	30
D	94	28	31

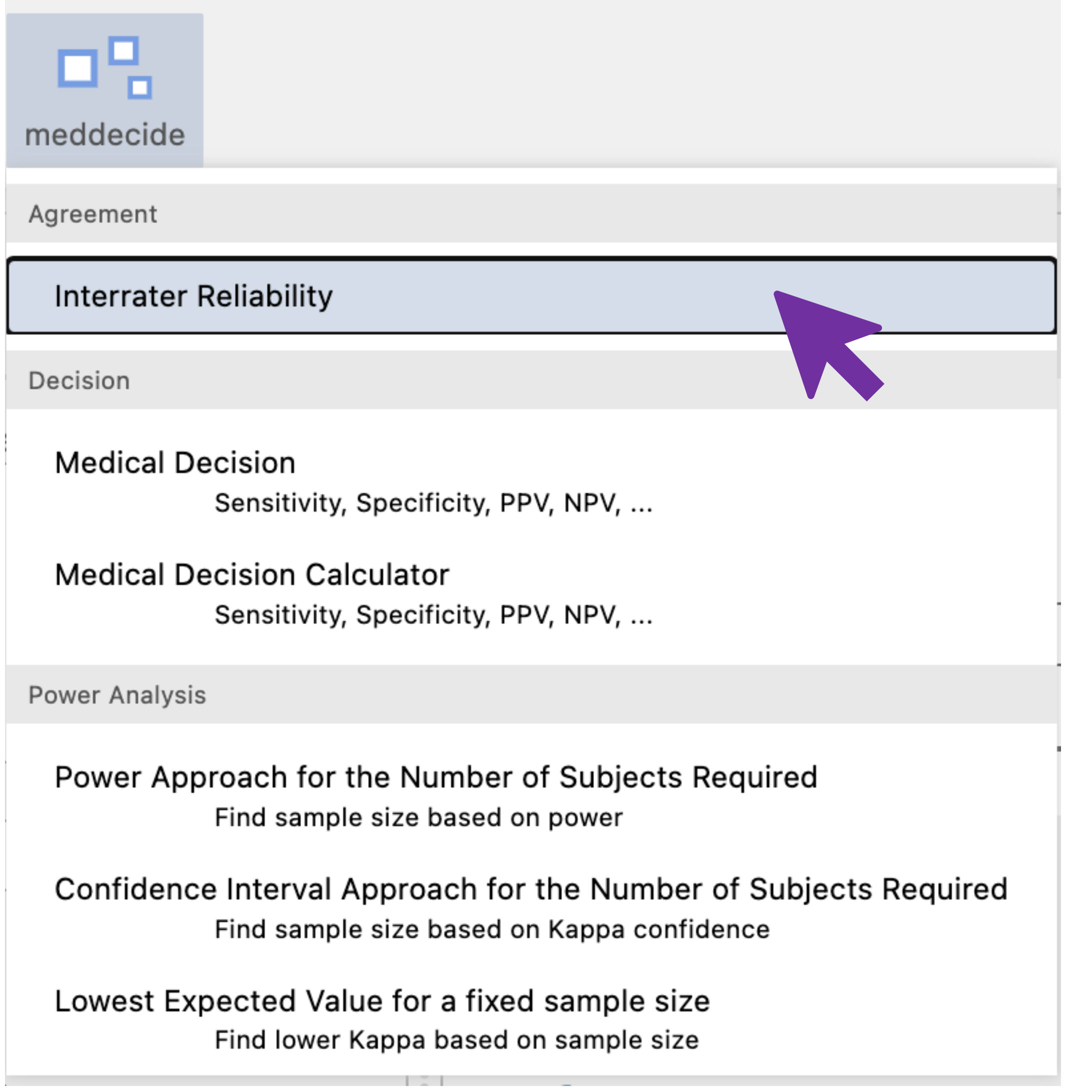


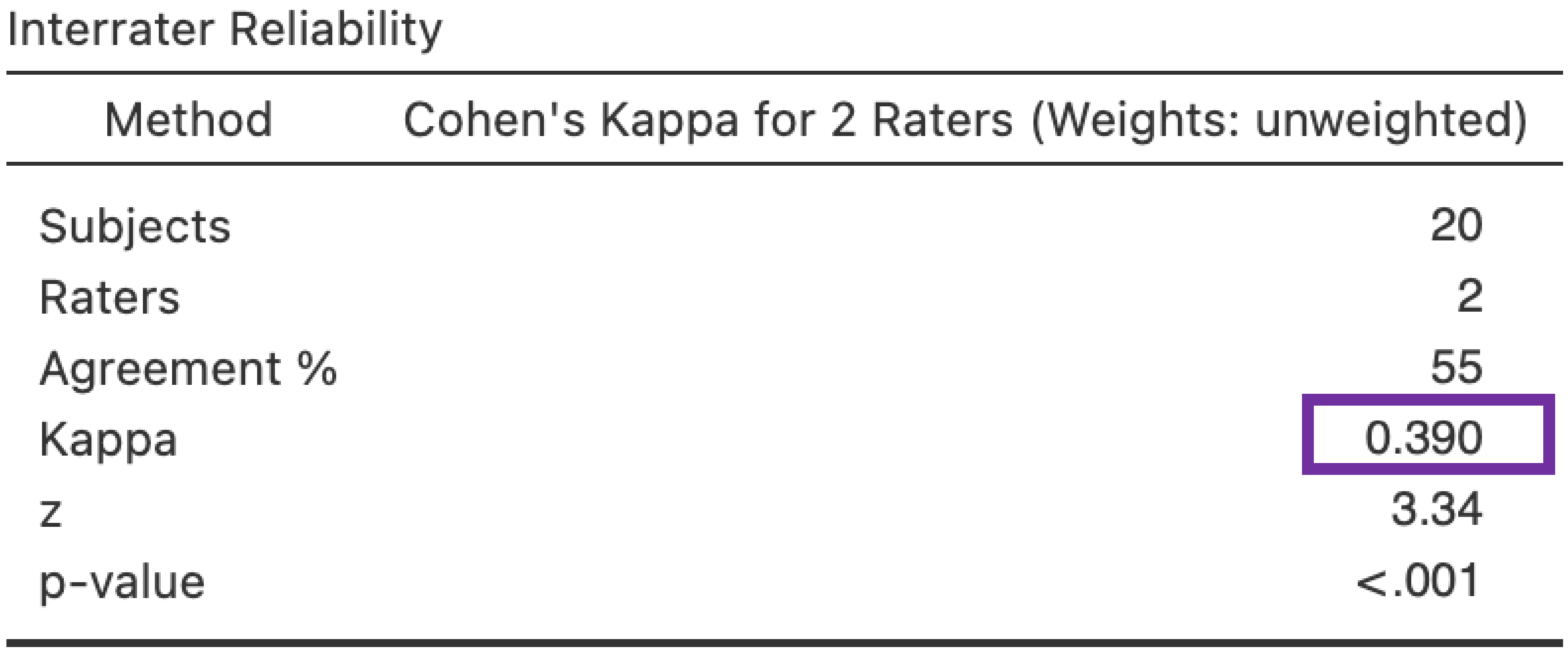
Long Format

Team	Variable	Value
A	Points	88
A	Assists	12
A	Rebounds	22
B	Points	91
B	Assists	17
B	Rebounds	28
C	Points	99
C	Assists	24
C	Rebounds	30
D	Points	94
D	Assists	28
D	Rebounds	31



Accessibility of Teacher	Accessibility of Teacher R2	Learning Rationale	Learning Rationale R2
3	3	3	3
3	3	3	3
2	3	3	3
2	1	3	2
4	2	1	2
4	4	1	3
4	4	1	3
3	3	1	3


meddecide
Agreement
Interrater Reliability
Decision
Medical Decision
Sensitivity, Specificity, PPV, NPV, ...
Medical Decision Calculator
Sensitivity, Specificity, PPV, NPV, ...
Power Analysis
Power Approach for the Number of Subjects Required
Find sample size based on power
Confidence Interval Approach for the Number of Subjects Required
Find sample size based on Kappa confidence
Lowest Expected Value for a fixed sample size
Find lower Kappa based on sample size



Interrater Reliability

Method	Cohen's Kappa for 2 Raters (Weights: unweighted)
Subjects	20
Raters	2
Agreement %	55
Kappa	0.390
z	3.34
p-value	<.001

Kappa Interpretation

Kappa Value	Interpretation
-1.00,0.0	No agreement
0.00, 0.20	Poor agreement
0.21, 0.40	Fair agreement
0.41, 0.75	Moderate agreement
0.76, 0.80	Excellent agreement

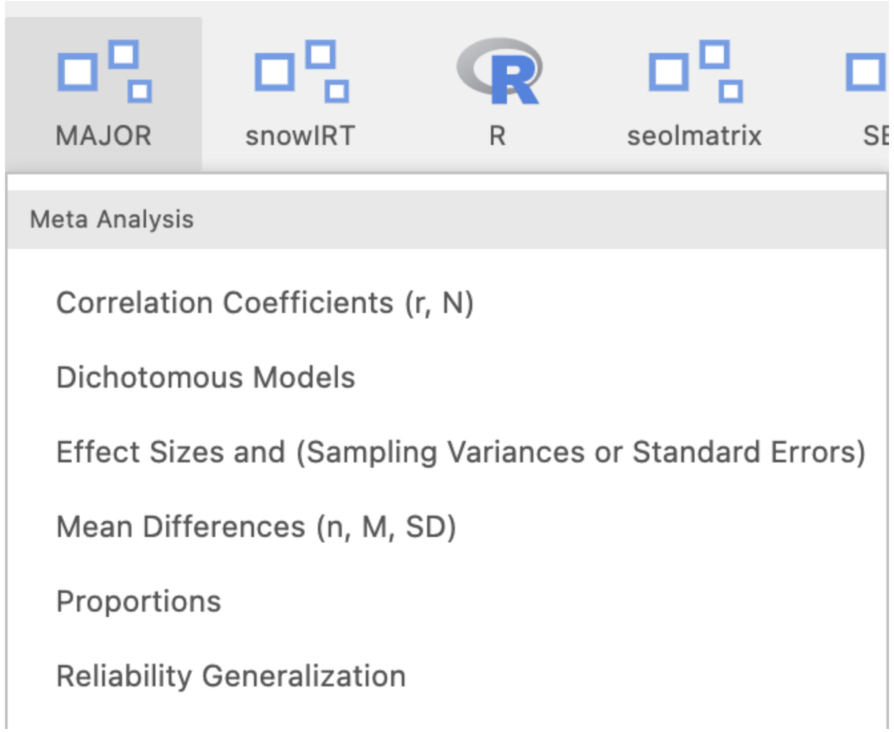
Landis and Koch (1977)

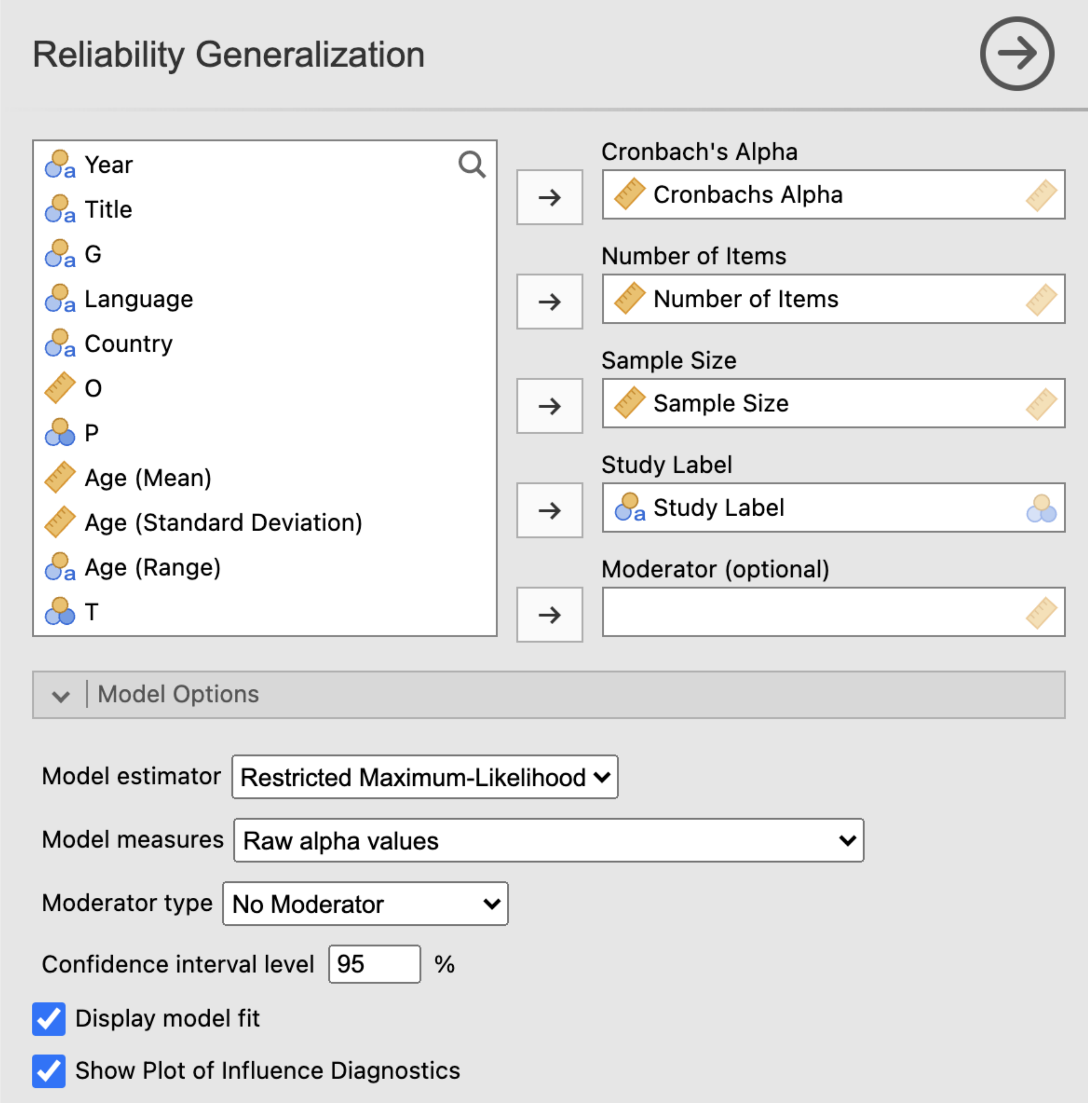
Resource

R Code

- Kappa Calculation & Interpretation
  - McHugh (2012)
  - Landis and Koch (1977)

```
install.packages("irr")

library(irr)

# Example data
ratings <- data.frame(
  rater1 = c(3, 3, 2, 3, 4, 4),
  rater2 = c(3, 3, 3, 1, 2, 4)
)

# Cohen's Kappa
kappa2(ratings)
```


MAJOR
snowIRT
R
seolmatrix
SE
Meta Analysis
Correlation Coefficients (r, N)
Dichotomous Models
Effect Sizes and (Sampling Variances or Standard Errors)
Mean Differences (n, M, SD)
Proportions
Reliability Generalization



Reliability Generalization

Cronbach's Alpha: Cronbachs Alpha

Number of Items: Number of Items

Sample Size: Sample Size

Study Label: Study Label

Moderator (optional):

Model estimator: Restricted Maximum-Likelihood

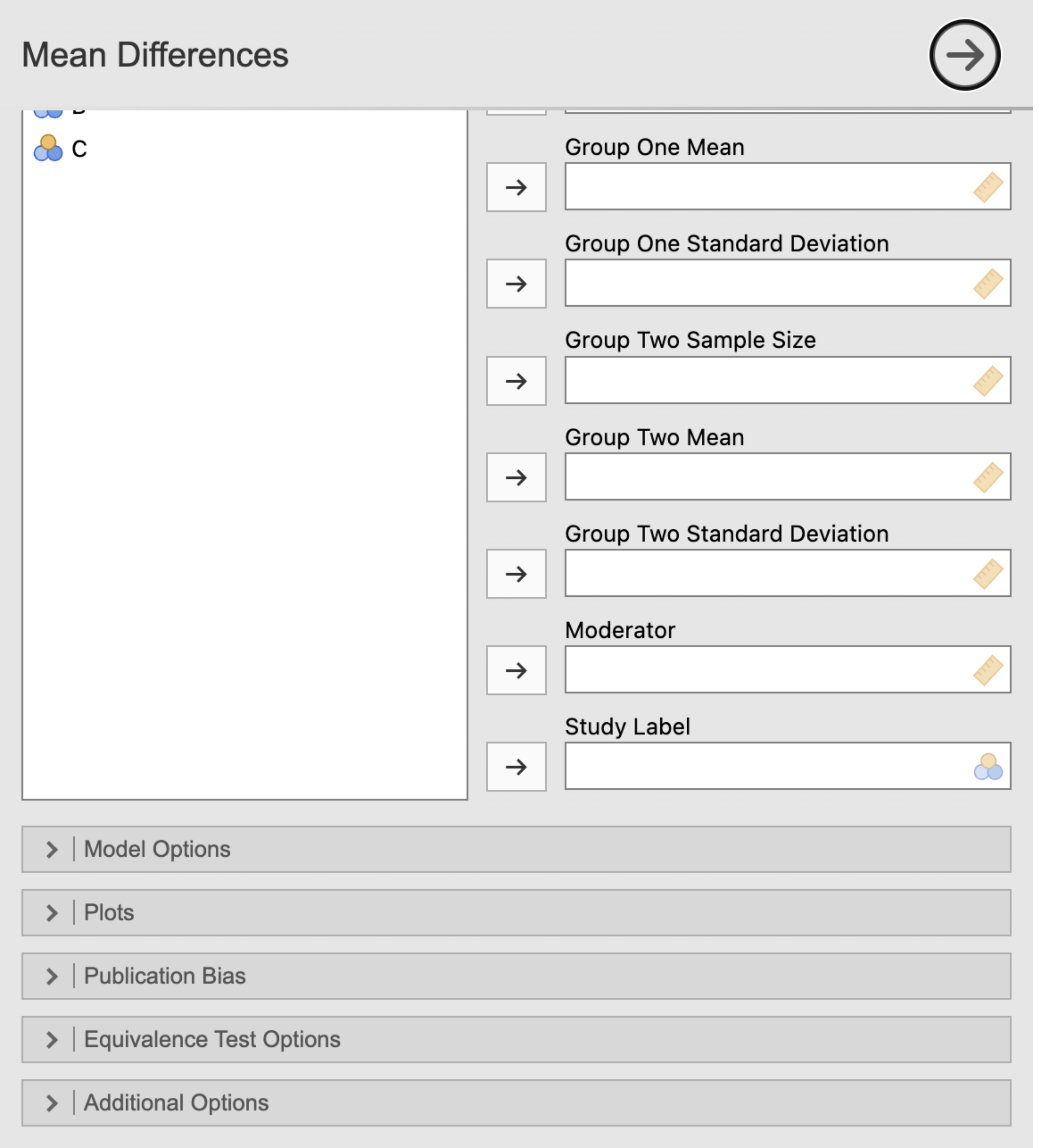
Model measures: Raw alpha values

Moderator type: No Moderator

Confidence interval level: 95 %

☑ Display model fit

☑ Show Plot of Influence Diagnostics



Mean Differences

C

Group One Mean

Group One Standard Deviation

Group Two Sample Size

Group Two Mean

Group Two Standard Deviation

Moderator

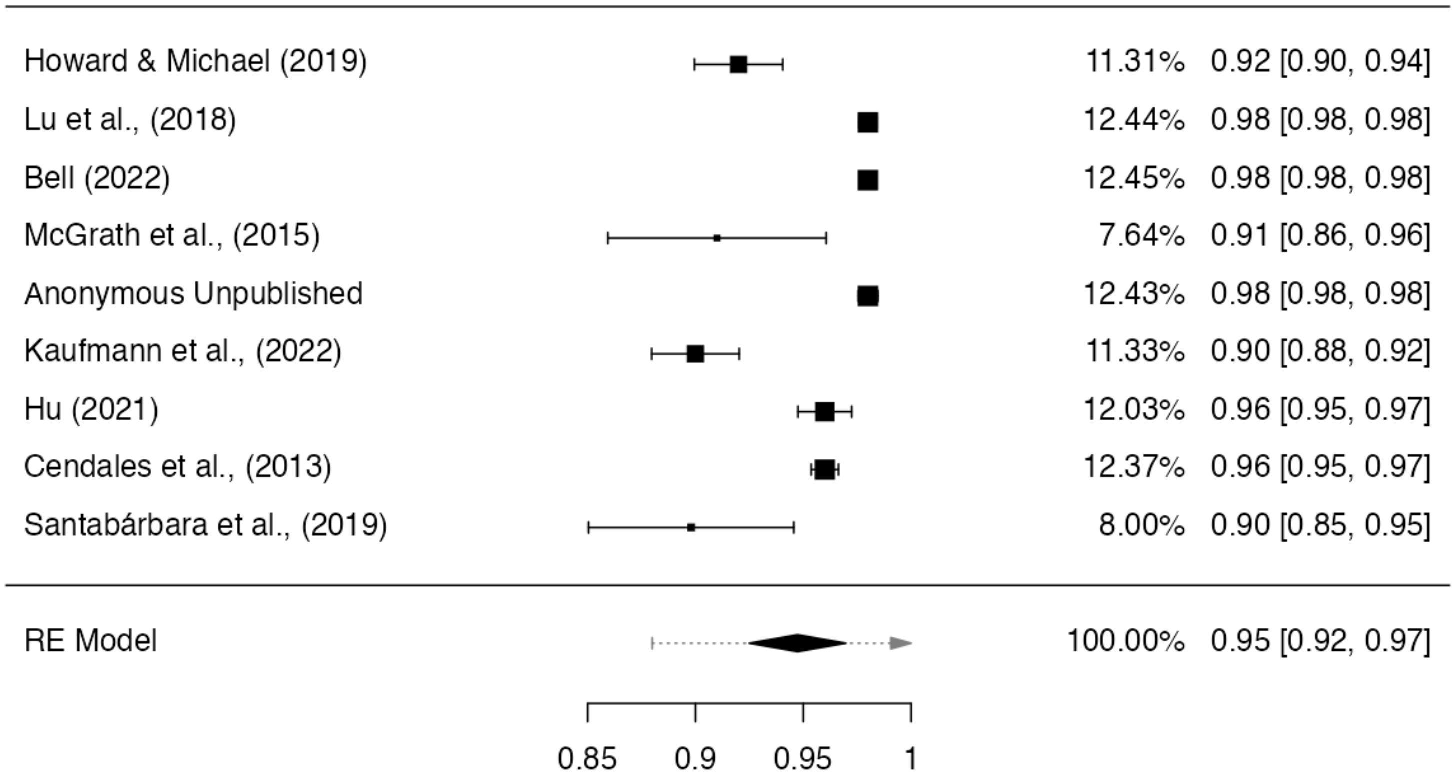
Study Label

- Model Options
- Plots
- Publication Bias
- Equivalence Test Options
- Additional Options

Forest Plot

Forest Plot

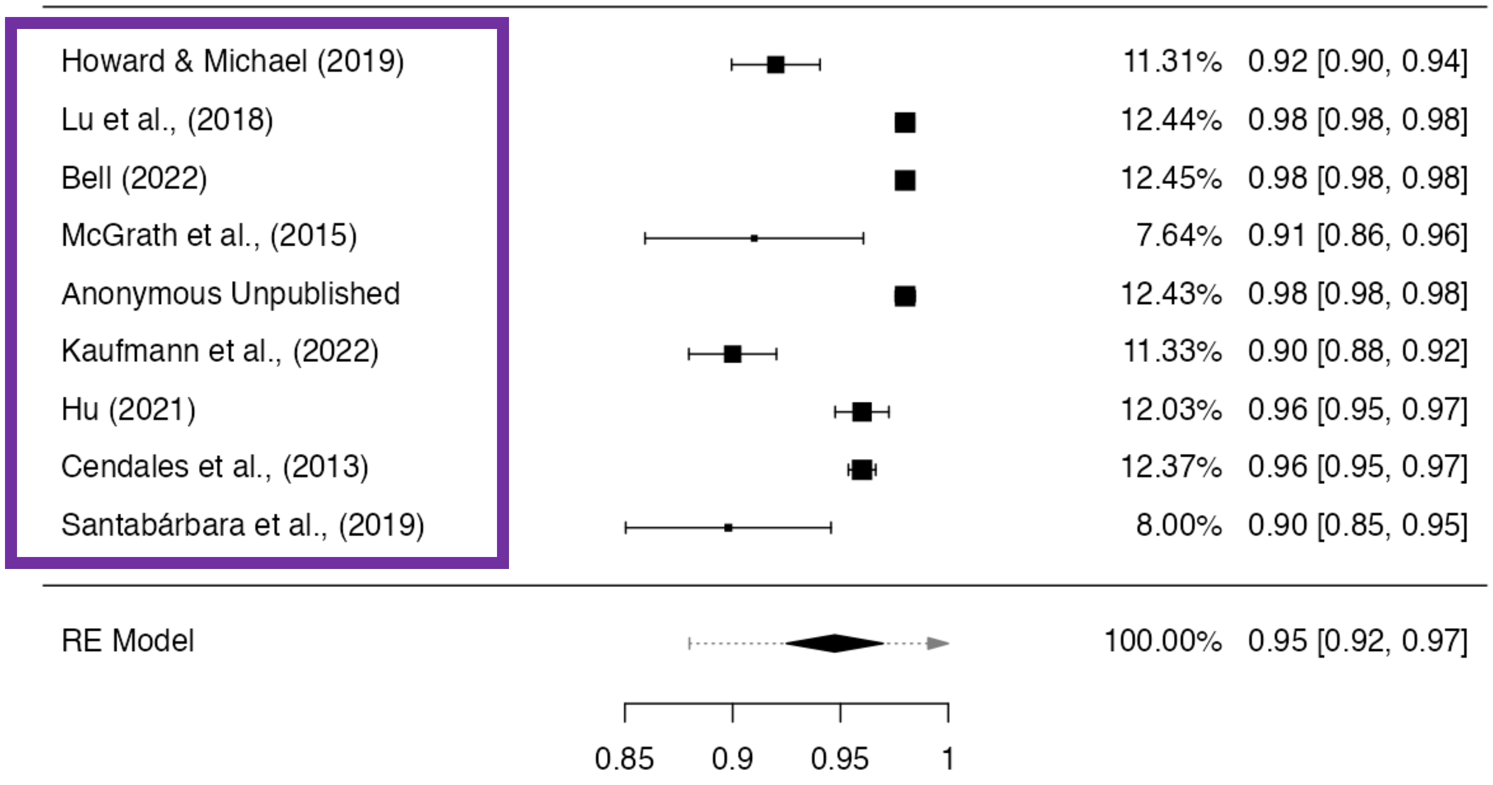
[3]



Forest Plot

[3]

Study Name



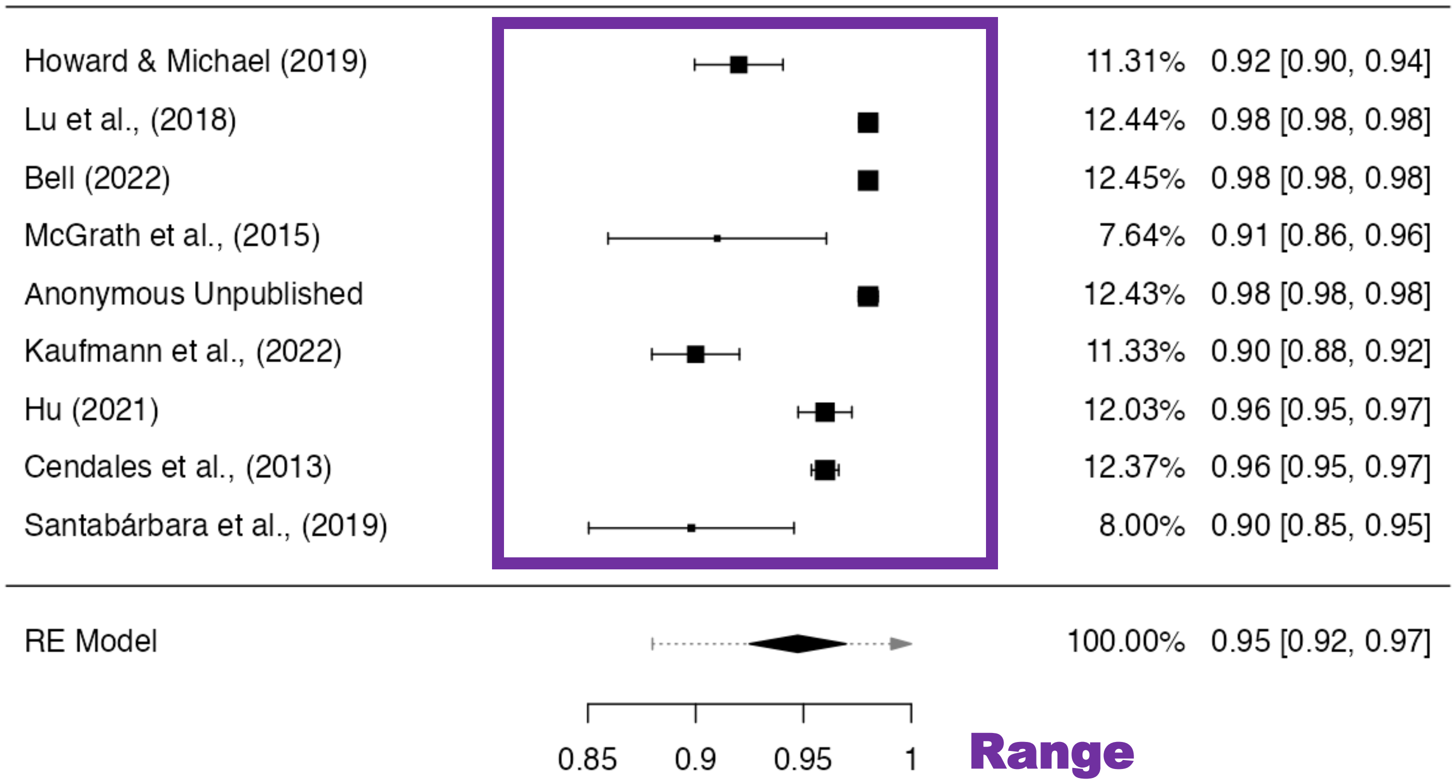
Howard & Michael (2019)	11.31%	0.92 [0.90, 0.94]
Lu et al., (2018)	12.44%	0.98 [0.98, 0.98]
Bell (2022)	12.45%	0.98 [0.98, 0.98]
McGrath et al., (2015)	7.64%	0.91 [0.86, 0.96]
Anonymous Unpublished	12.43%	0.98 [0.98, 0.98]
Kaufmann et al., (2022)	11.33%	0.90 [0.88, 0.92]
Hu (2021)	12.03%	0.96 [0.95, 0.97]
Cendales et al., (2013)	12.37%	0.96 [0.95, 0.97]
Santabárbara et al., (2019)	8.00%	0.90 [0.85, 0.95]
RE Model	100.00%	0.95 [0.92, 0.97]

Forest Plot

[3]

Effect Size

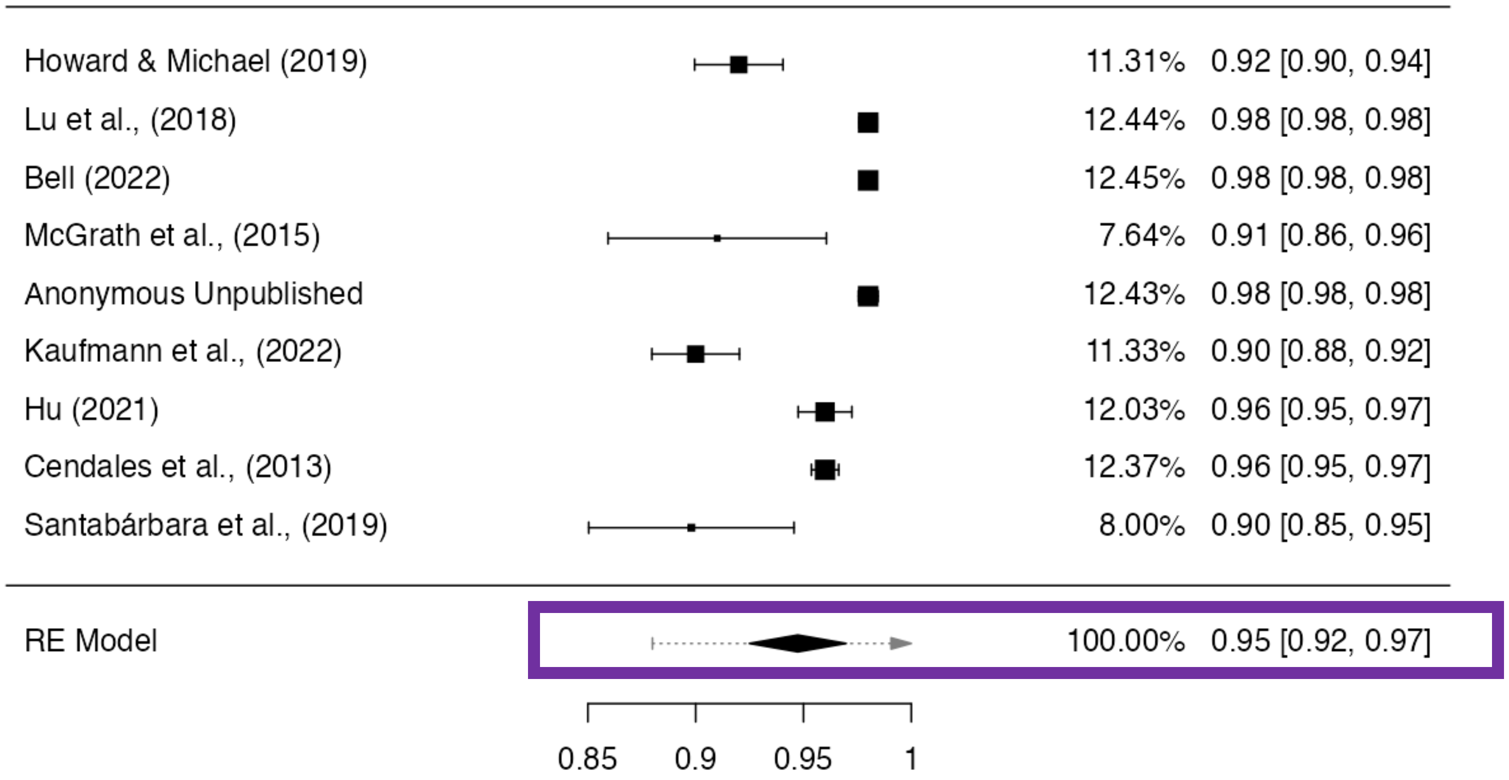
● ■



Range

Forest Plot

[3]

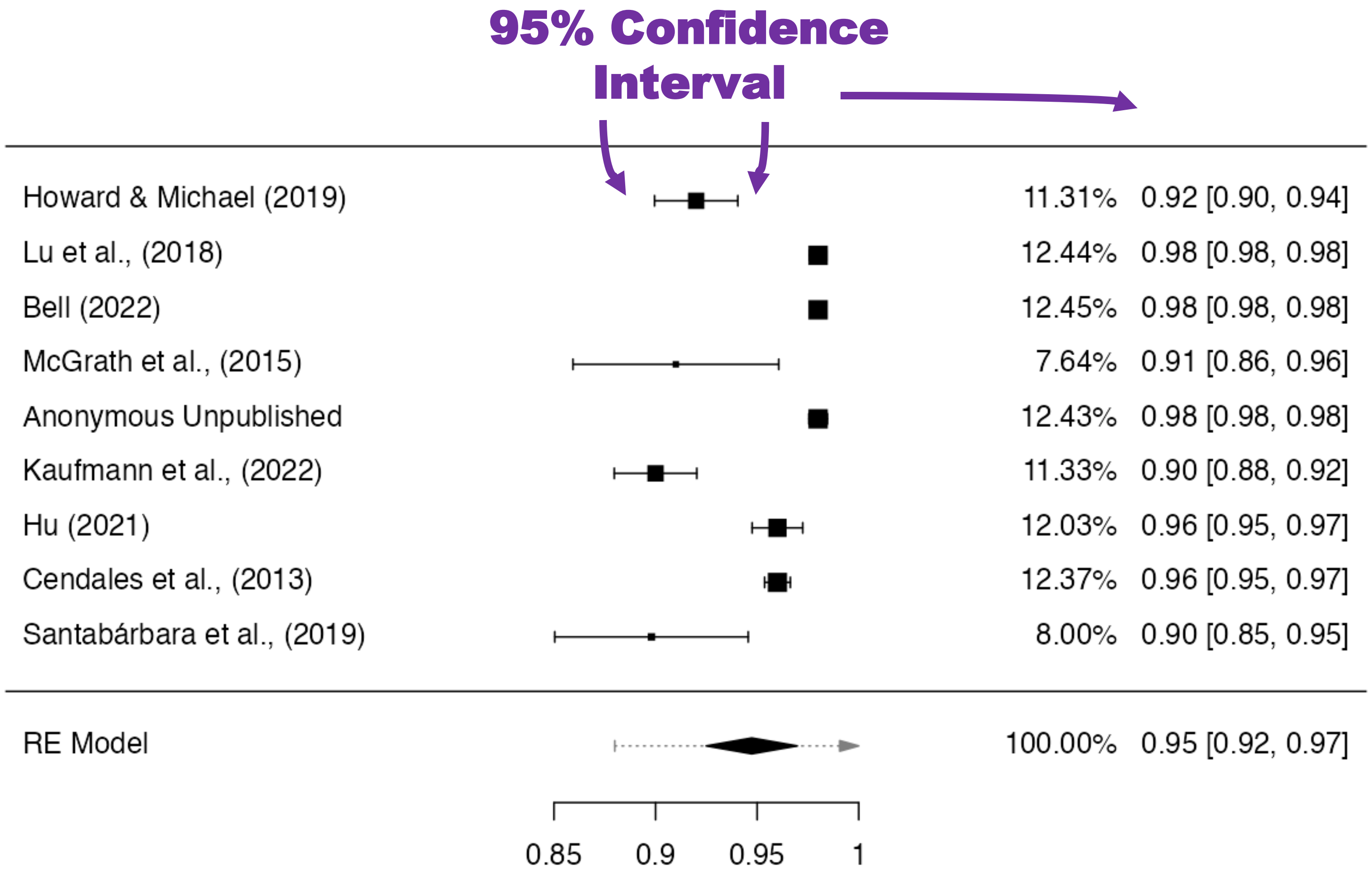


Howard & Michael (2019)	11.31%	0.92 [0.90, 0.94]
Lu et al., (2018)	12.44%	0.98 [0.98, 0.98]
Bell (2022)	12.45%	0.98 [0.98, 0.98]
McGrath et al., (2015)	7.64%	0.91 [0.86, 0.96]
Anonymous Unpublished	12.43%	0.98 [0.98, 0.98]
Kaufmann et al., (2022)	11.33%	0.90 [0.88, 0.92]
Hu (2021)	12.03%	0.96 [0.95, 0.97]
Cendales et al., (2013)	12.37%	0.96 [0.95, 0.97]
Santabárbara et al., (2019)	8.00%	0.90 [0.85, 0.95]
RE Model	100.00%	0.95 [0.92, 0.97]

Pooled Effect

Forest Plot

[3]



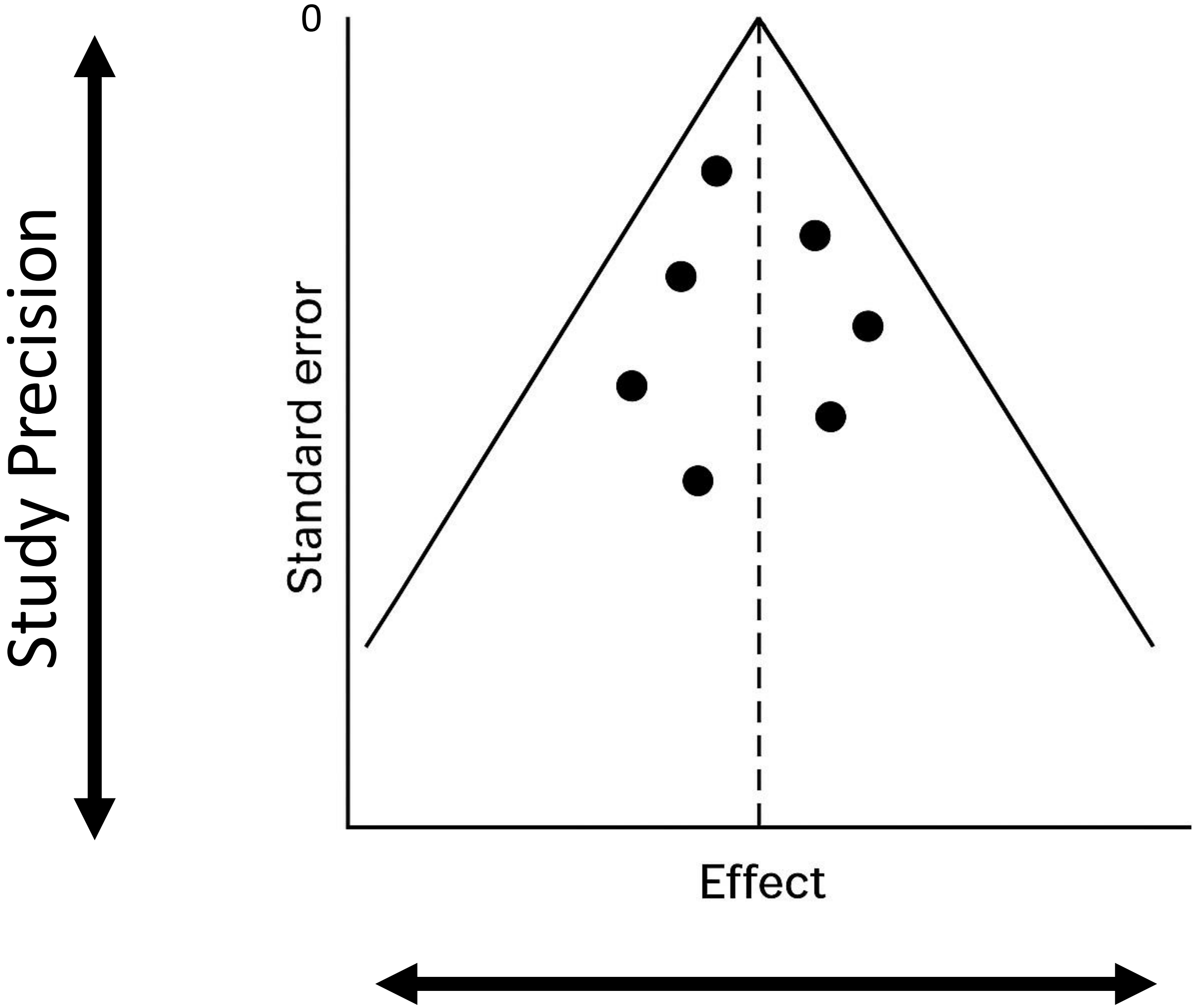
95% Confidence Interval

Howard & Michael (2019)	11.31%	0.92 [0.90, 0.94]
Lu et al., (2018)	12.44%	0.98 [0.98, 0.98]
Bell (2022)	12.45%	0.98 [0.98, 0.98]
McGrath et al., (2015)	7.64%	0.91 [0.86, 0.96]
Anonymous Unpublished	12.43%	0.98 [0.98, 0.98]
Kaufmann et al., (2022)	11.33%	0.90 [0.88, 0.92]
Hu (2021)	12.03%	0.96 [0.95, 0.97]
Cendales et al., (2013)	12.37%	0.96 [0.95, 0.97]
Santabárbara et al., (2019)	8.00%	0.90 [0.85, 0.95]
RE Model	100.00%	0.95 [0.92, 0.97]

Funnel Plot

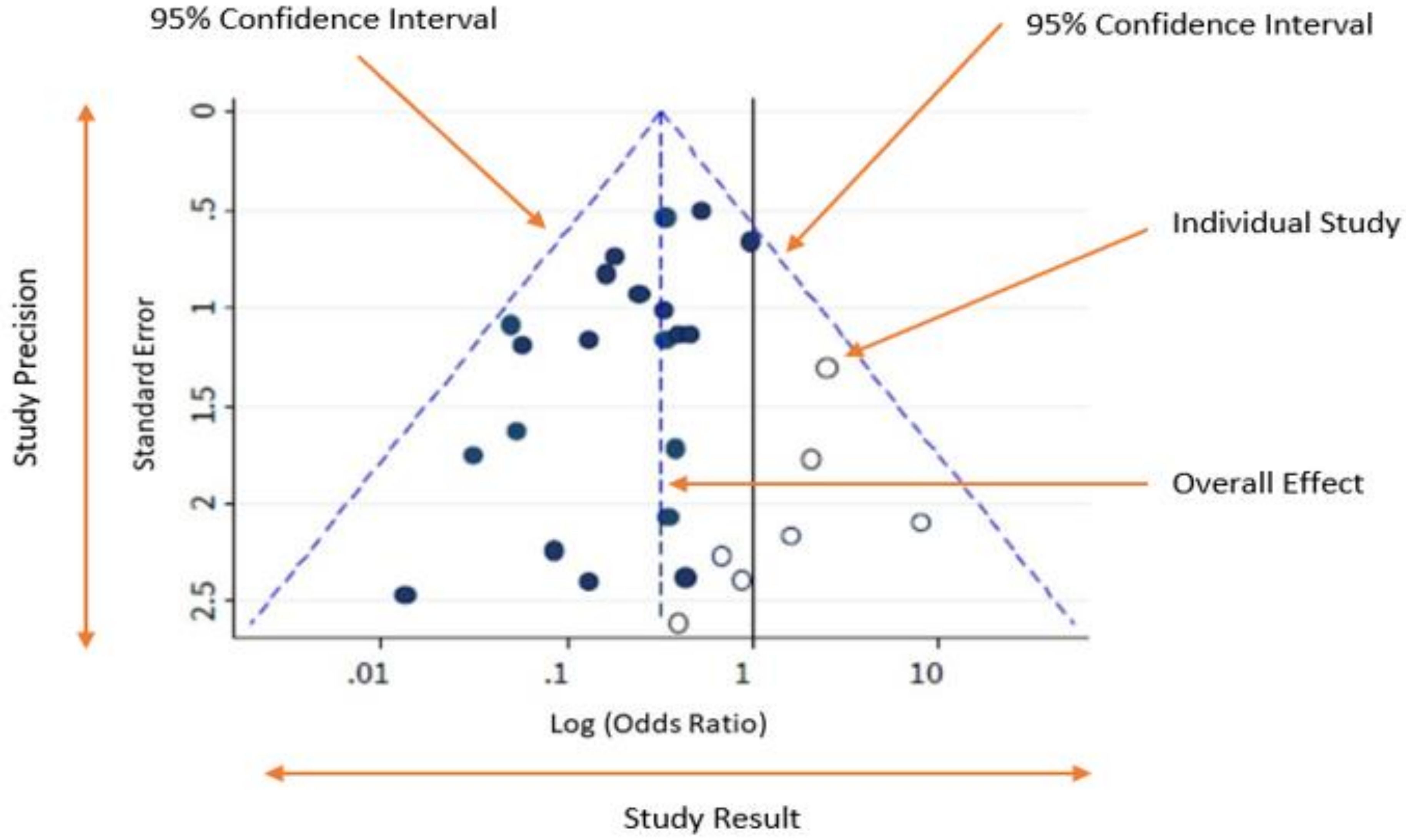
Funnel Plot

- A funnel plot tells you about the variability (standard error) of the individual studies against the mean effect size.
- As the study size increases the SE approaches zero.
- Assumes the plot should be symmetrical (that there are as many studies above / below the mean effect size)
    - Lack of symmetry can suggest publication bias, or ”small study” bias



Study Precision

Study Result



95% Confidence Interval

95% Confidence Interval

Individual Study

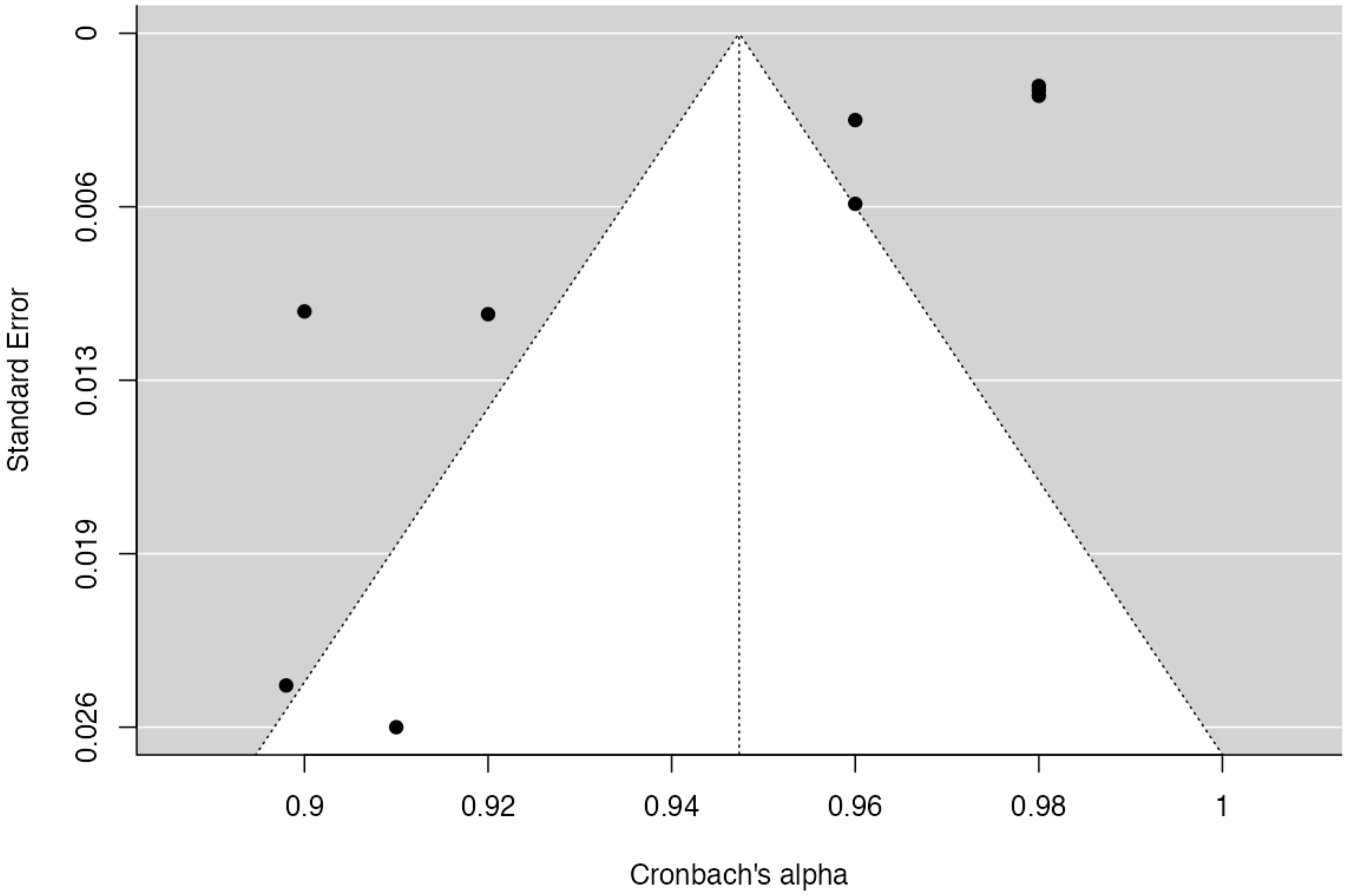
Overall Effect

Study Precision

Study Result

Funnel Plot

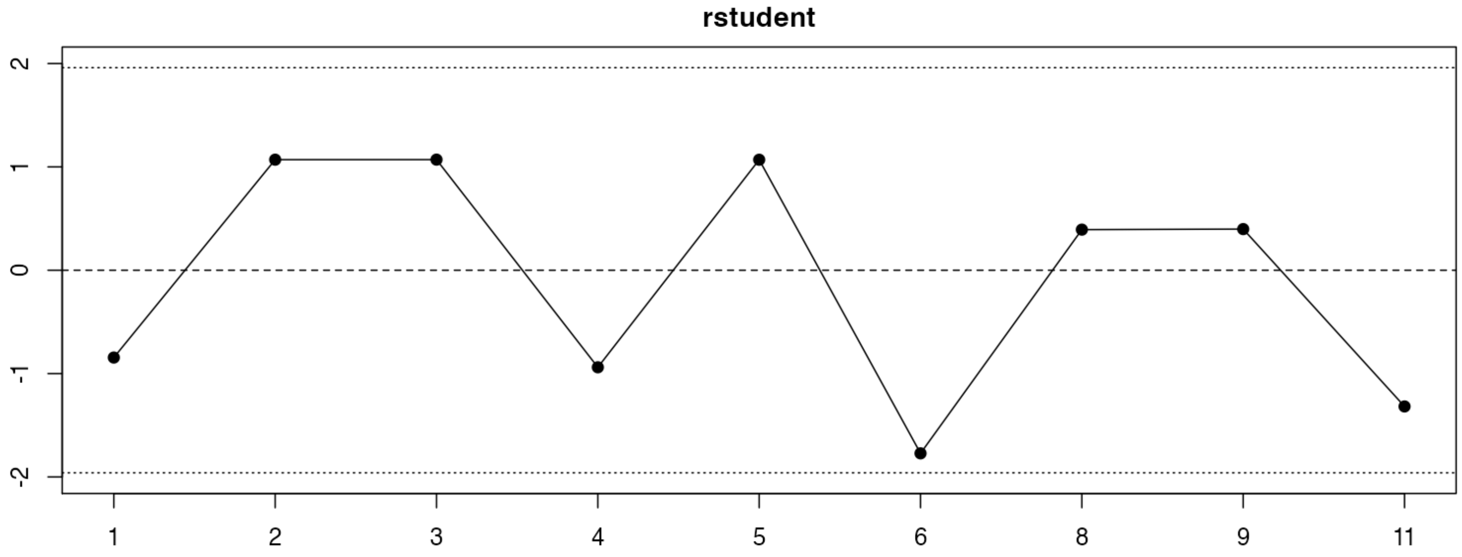
[3]



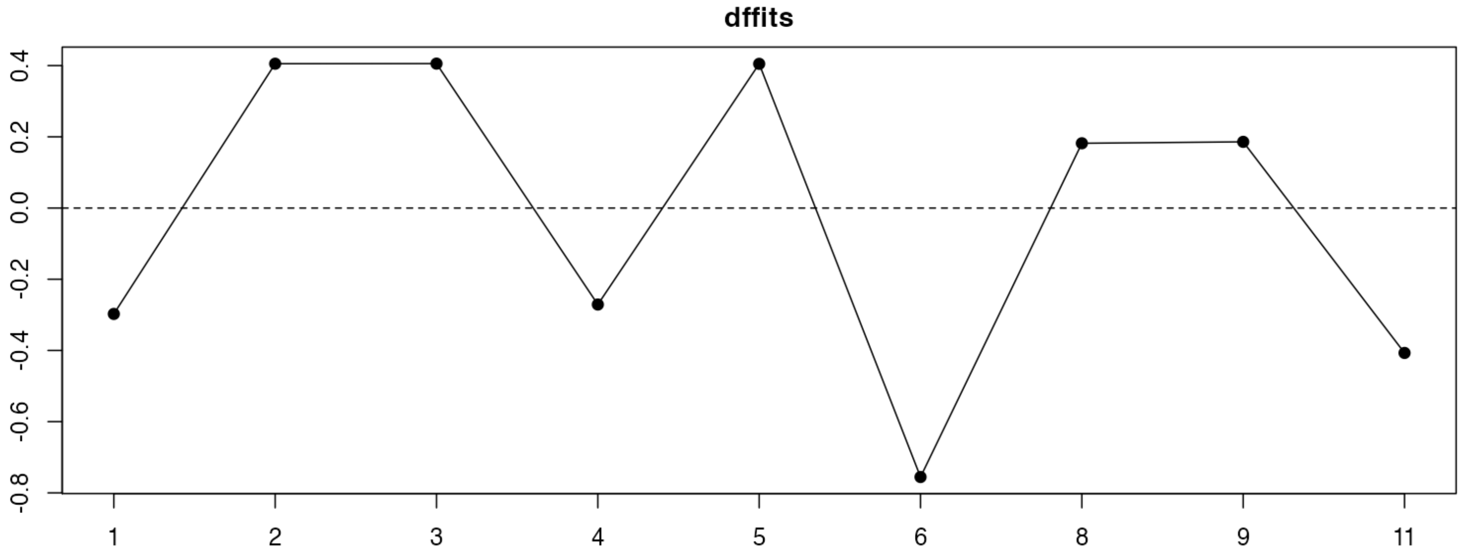
Outliers

Outlier and Influential Case Diagnostics

Externally Standardized Residual

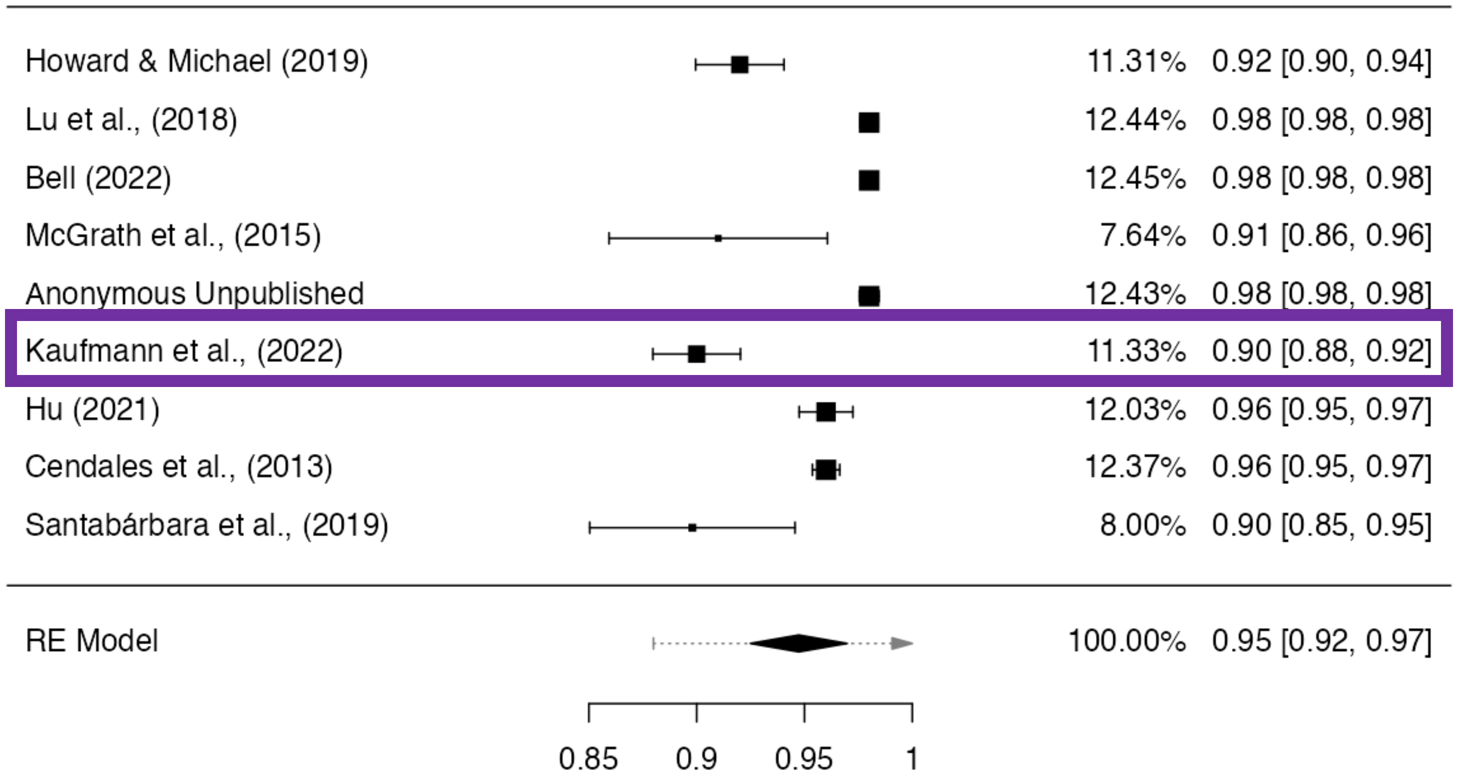


DFFITS Values

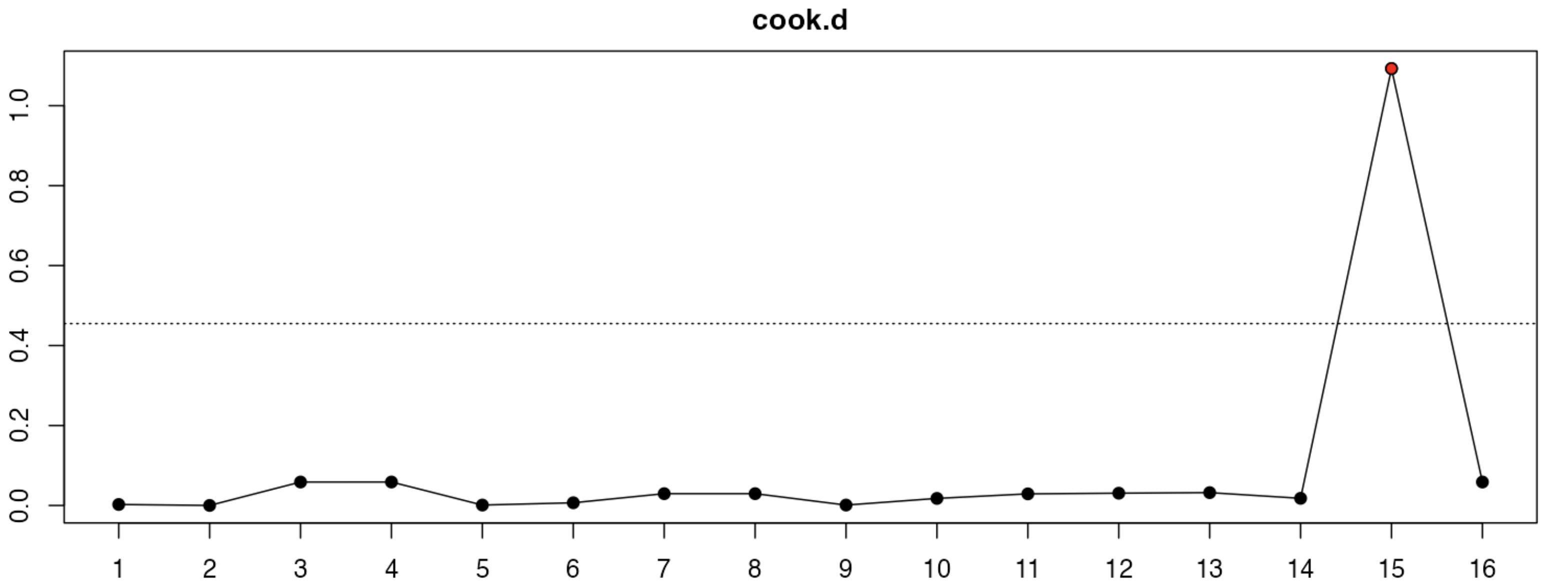


Forest Plot

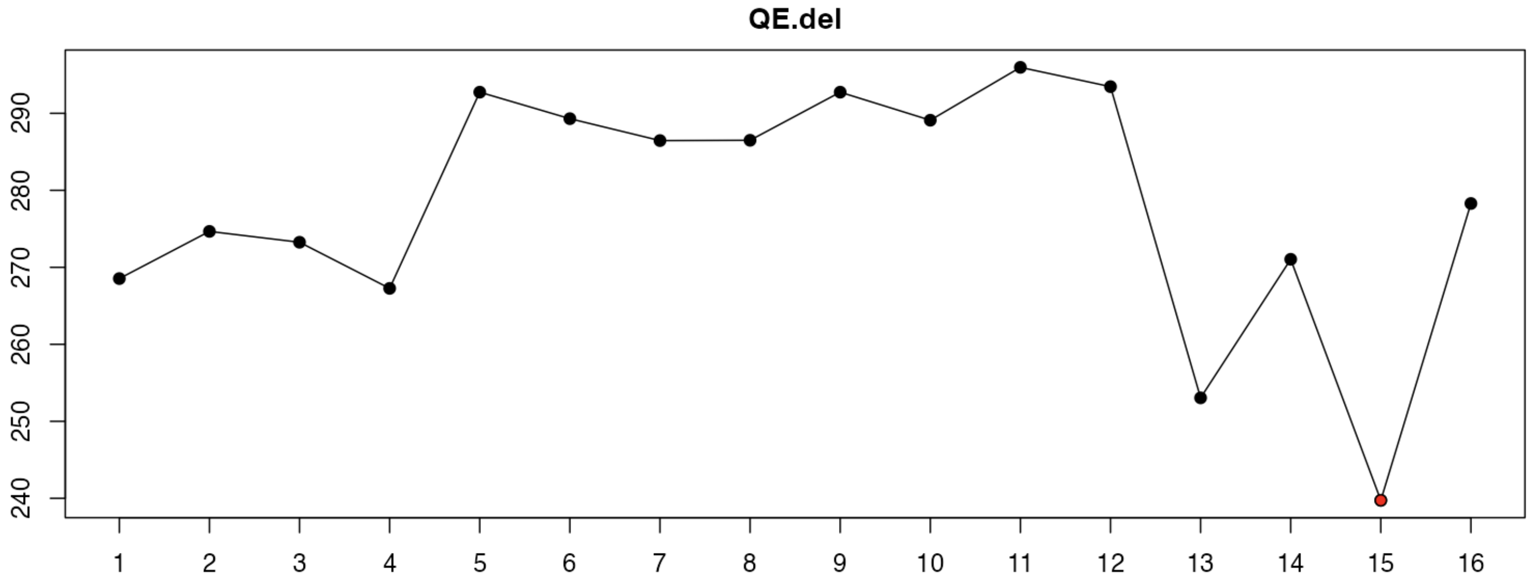
[3]

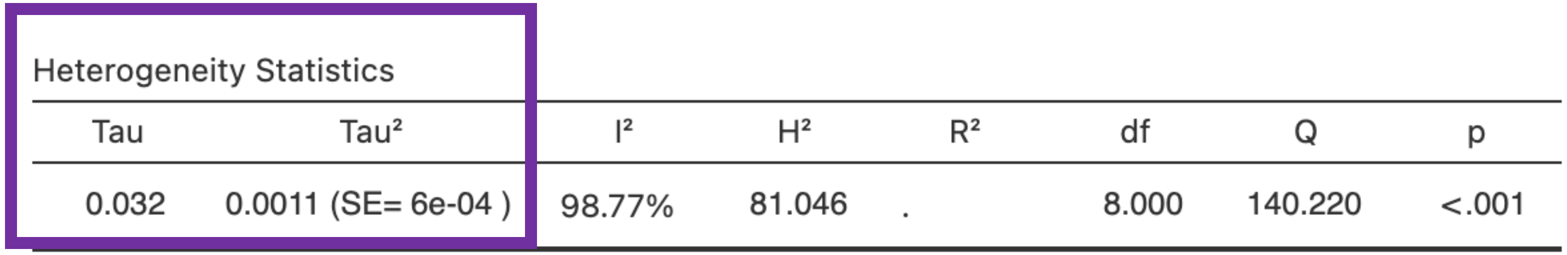


Cook's Distances



Leave-one-out (residual) Heterogeneity Test Statistics

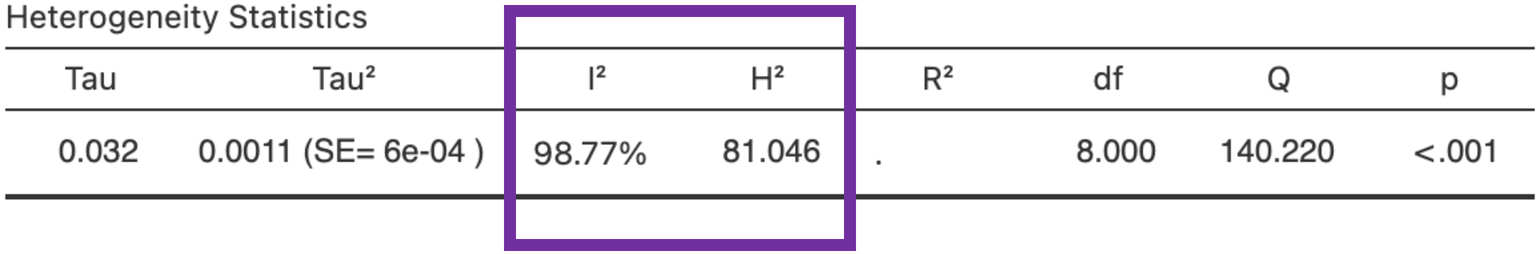




Heterogeneity Statistics

Tau	Tau^2	I^2	H^2	R^2	df	Q	p
0.032	0.0011 (SE= 6e-04)	98.77%	81.046	.	8.000	140.220	<.001

- Tau
  - Estimated SD of the population effect sizes.
    - τ = 0 means all studies share the same population effect sizes
    - Larger τ greater spread in the population effect sizes
    - τ expressed in the units of the effect size.
- Tau Squared
  - The between-study variance
  - .0011 is VERY SMALL indicates that the effect size variability is minimal



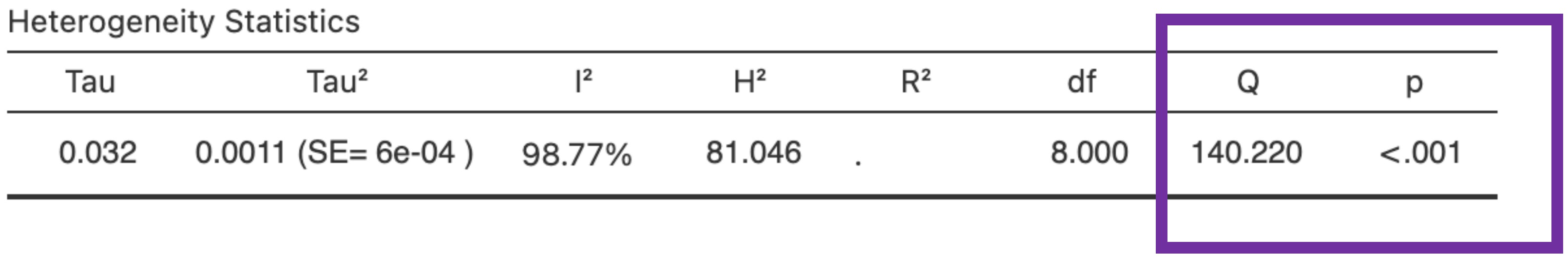
Heterogeneity Statistics

Tau	Tau^2	I^2	H^2	R^2	df	Q	p
0.032	0.0011 (SE= 6e-04)	98.77%	81.046	.	8.000	140.220	<.001

- I^2 statistic
  - The percentage of variation across studies that is due to heterogeneity rather than chance
    - *"How much of the variability in effect sizes is because studies are actually different, and not just random noise?"*
  - I^2 is an intuitive and simple expression of the inconsistency of studies' results.
- H^2 statistic
  - The observed variance among studies is 81 times greater than what we would expect if all studies shared the same population effect size.
  - $H^2 = 1$ No excess heterogeneity

I^2 Value	Interpretation
0–25%	Low heterogeneity
25–50%	Moderate heterogeneity
50–75%	Substantial heterogeneity
75–100%	Considerable heterogeneity

Higgins et al. (2003)



Heterogeneity Statistics

Tau	Tau^2	I^2	H^2	R^2	df	Q	p
0.032	0.0011 (SE= 6e-04)	98.77%	81.046	.	8.000	140.220	<.001

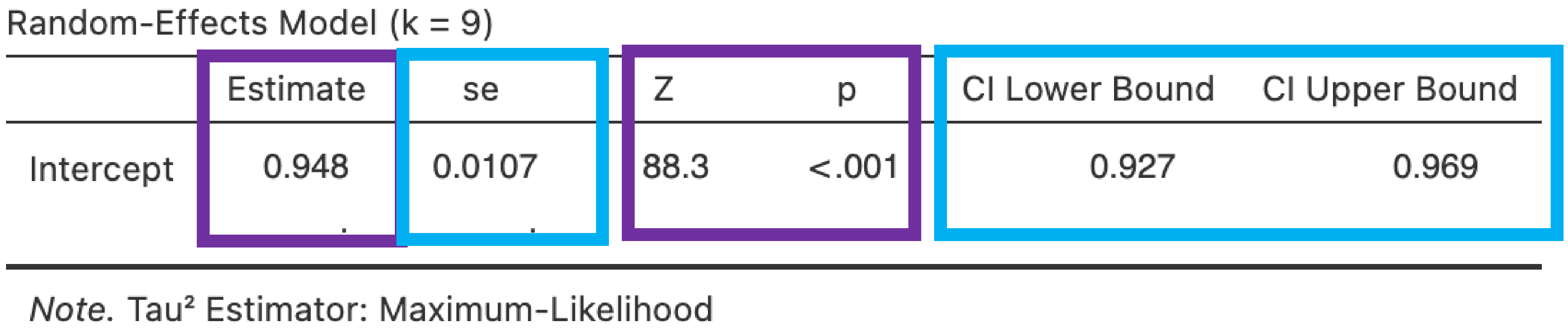
- Cochran’s Q
  - Classical measure of heterogeneity
  - Ho: the true effect size is the same across studies and variations are simply caused by chance.
- Small Q means low heterogeneity
  - effects are similar across studies
- Large Q means heterogeneity present
  - studies differ more than expected

Fixed Effects vs Random Effects

- **Fixed Effects**
  - conduct if it is reasonable to assume underlying effect size is SAME for all studies
- Test: test of heterogeneity
  - Pooling
  - If significant, go for random effects model
- If there is very little variation between trials then I^2 will be low and a fixed effects model might be appropriate.

- **Random Effects**
  - Conduct if test of heterogeneity is significant.
  - Q: $p < .05$
- Assumes outcome comes from a normal distribution
- More practical

Ho: $\theta = 0$
Ha: $\theta \neq 0$



Random-Effects Model (k = 9)

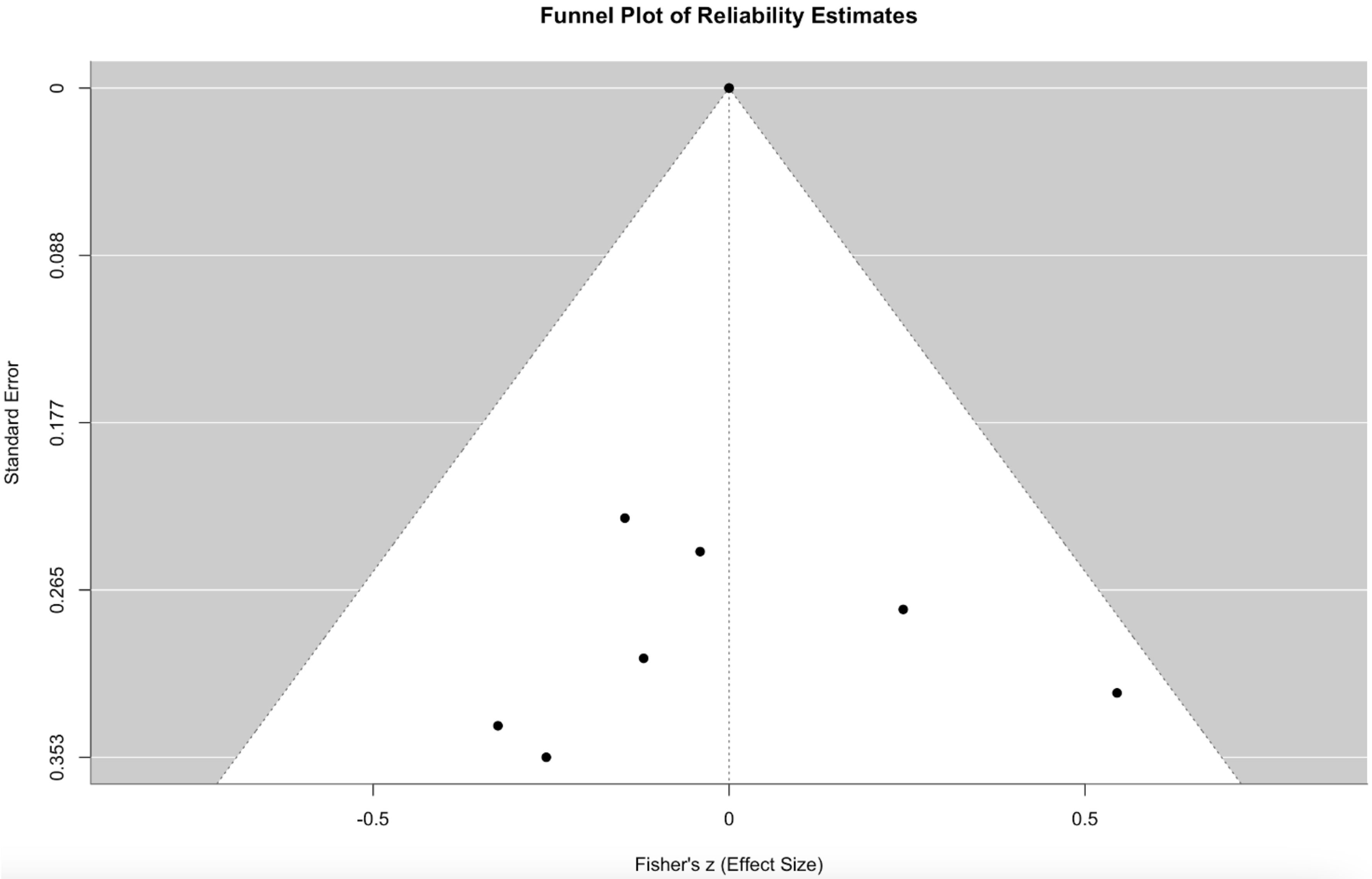
	Estimate	se	Z	p	CI Lower Bound	CI Upper Bound
Intercept	0.948	0.0107	88.3	<.001	0.927	0.969

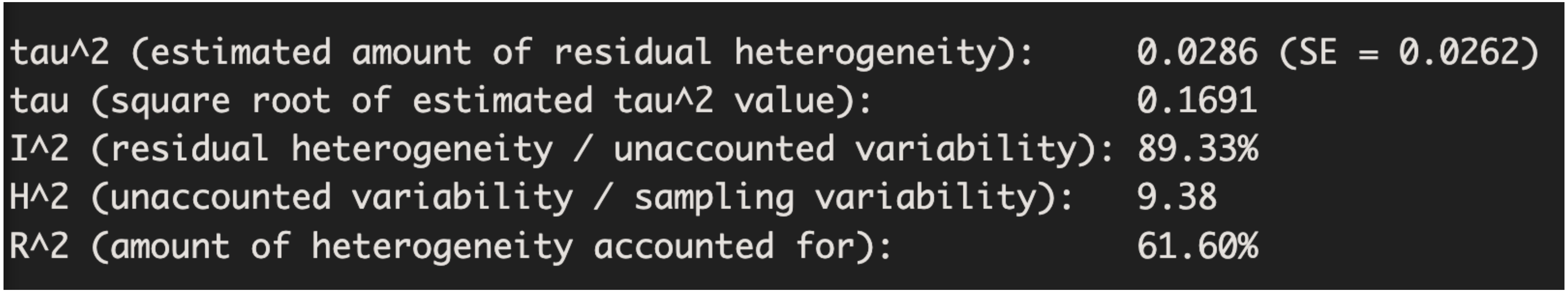
Note. Tau² Estimator: Maximum-Likelihood

R Code

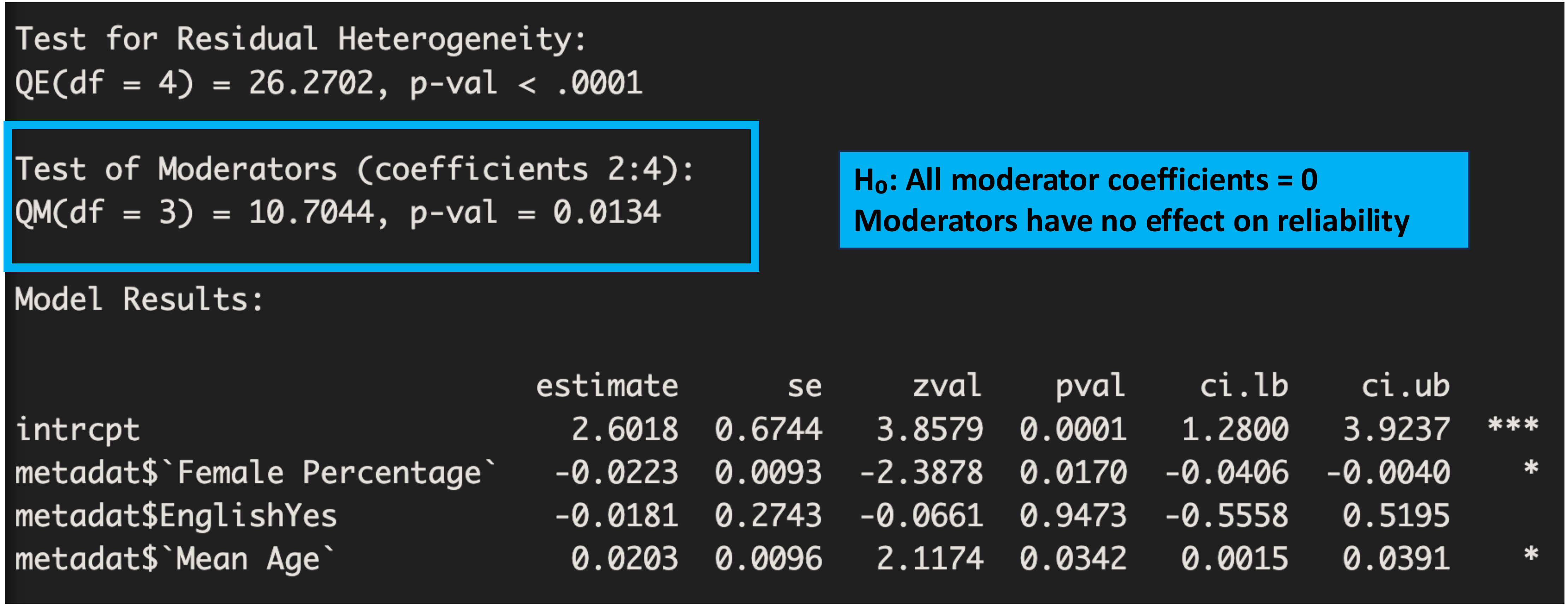
R Code

```
library(metafor)
metadat <- read.csv("~/Library/CloudStorage/OneDrive-NewMexicoStateUniversity/Research/CSSE Meta-Analysis/APA 2025/APA 2025 Poster.csv")
metadat$z_alpha <- atanh(metadat$Reliabilitity.Overall)
metadat$var_z_alpha <- 1 / (metadat$N - 3)
res <- rma(yi = z_alpha, vi = var_z_alpha,
        mods = ~metadat$Gender+ metadat$Language + metadat$Age,
        data = metadat, method = "REML")
```


Funnel Plot of Reliability Estimates
Standard Error
0
0.088
0.177
0.265
0.353
-0.5
0
0.5
Fisher's z (Effect Size)



```
tau^2 (estimated amount of residual heterogeneity):     0.0286 (SE = 0.0262)
tau (square root of estimated tau^2 value):             0.1691
I^2 (residual heterogeneity / unaccounted variability): 89.33%
H^2 (unaccounted variability / sampling variability):   9.38
R^2 (amount of heterogeneity accounted for):            61.60%
```



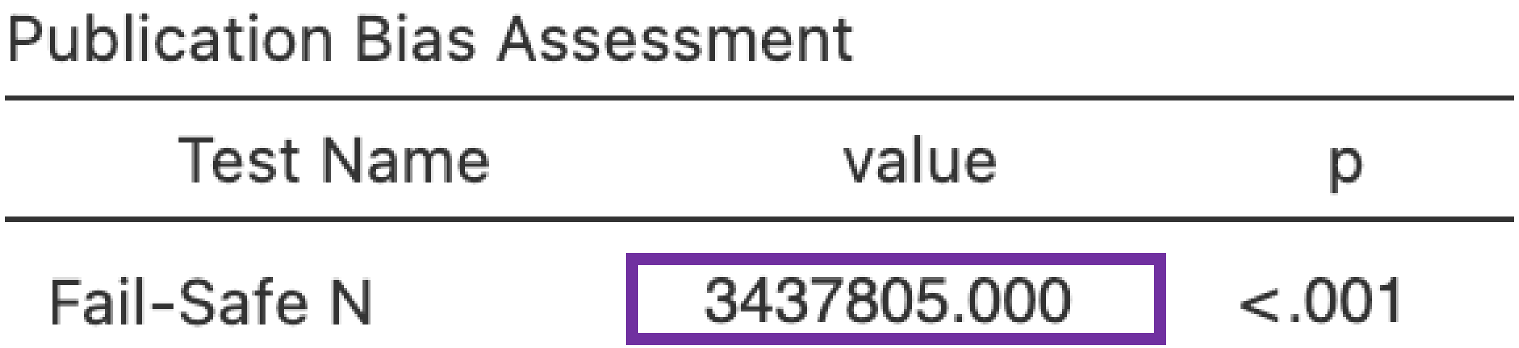
```
Test for Residual Heterogeneity:
QE(df = 4) = 26.2702, p-val < .0001

Test of Moderators (coefficients 2:4):
QM(df = 3) = 10.7044, p-val = 0.0134

Model Results:

                             estimate      se     zval    pval    ci.lb    ci.ub
intrcpt                        2.6018  0.6744   3.8579  0.0001   1.2800   3.9237  ***
metadat$`Female Percentage`   -0.0223  0.0093  -2.3878  0.0170  -0.0406  -0.0040    *
metadat$EnglishYes            -0.0181  0.2743  -0.0661  0.9473  -0.5558   0.5195
metadat$`Mean Age`             0.0203  0.0096   2.1174  0.0342   0.0015   0.0391    *
```

H_0: All moderator coefficients = 0
Moderators have no effect on reliability

Issues in Meta-Analyses

- Published studies
- The drawer problem
- Fail Safe N

Fail Safe N



Publication Bias Assessment

Test Name	value	p
Fail-Safe N	3437805.000	<.001

References & Slides