REPRODUCIBILITY OF PSYCHOLOGICAL META-ANALYSES

Esther Maassen

e.maassen@uvt.nl

@esthermaassen

osf.io/7nsmd/

metaresearch.nl

TILBURG ✦ UNIVERSITY

Data Extraction Errors in Meta-analyses That Use Standardized Mean Differences

Peter C. Gøtzsche, MD, DrMedSci

Asbjørn Hróbjartsson, MD, PhD

Katja Marić, MSc

Britta Tendal, MSc

Results Of the 27 meta-analyses included in this study, we could not replicate the result for at least 1 of the 2 trials within 0.1 in 10 of the meta-analyses (37%), and in 4 cases, the discrepancy was 0.6 or more for the point estimate. Common problems were erroneous number of patients, means, standard deviations, and sign for the effect estimate. In total, 17 meta-analyses (63%) had errors for at least 1 of the 2 trials examined. For the 10 meta-analyses with errors of at least 0.1, we checked the data from all the trials and conducted our own meta-analysis, using the authors' methods. Seven of these 10 meta-analyses were erroneous (70%); 1 was subsequently retracted, and in 2 a significant difference disappeared or appeared.

osf.io/7nsmd/

Revealed or Concealed? Transparency of Procedures, Decisions, and Judgment Calls in Meta-Analyses

Zeynep G. Aytug[1], Hannah R. Rothstein[1], Wencang Zhou[1], and Mary C. Kern[1]

Abstract

The authors examined the degree to which meta-analyses in the organizational sciences transparently report procedures, decisions, and judgment calls by systematically reviewing all (198) meta-analyses published between 1995 and 2008 in 11 top journals that publish meta-analyses in industrial and organizational psychology and organizational behavior. The authors extracted information on 54 features of each meta-analysis. On average, the meta-analyses in the sample provided 52.8% of the information needed to replicate the meta-analysis or to assess its validity and 67.6% of the information considered to be most important according to expert meta-analysts. More recently published meta-analyses exhibited somewhat more transparent reporting practices than older ones did. Overall transparency of reporting (but not reporting of the most important items) was associated with higher ranked journals; transparency was not significantly related to number of citations. The authors discuss the implications of inadequate reporting of meta-analyses for development of cumulative knowledge and effective practice and make suggestions for improving the current

Examining the Reproducibility of Meta-Analyses in Psychology: A Preliminary Report

published meta-analyses. Reproducing published meta-analyses was surprisingly difficult. 96% of meta-analyses published in 2013-2014 did not adhere to reporting guidelines. A third of these meta-analyses did not contain a table specifying all individual effect sizes. Five of the 20 randomly selected meta-analyses we attempted to reproduce could not be reproduced at all due to lack of access to raw data, no details about the effect sizes extracted from each study, or a lack of information about how effect sizes were coded. In the

Daniel Lakens, Marcel van Assen, Farid Anvari, Katherine S. Corker, James. A Grange, Heike Gerger, Fred Hasselman, Jacklyn Koyama et al.

osf.io/7nsmd/

Our study

Part 1: individual study effect size reproducibility

Part 2: meta-analysis reproducibility

Part 1: Method

- 33 psychological meta-analyses
- 1979 individual effect sizes → 500

<u>Table 3: Mean differences between groups on variables</u>

	European American		Asian Indian	
	M	**SD**	**M**	**SD**
Parent Ethnic Identity	3.50	0.67	4.31	0.45
Parent other group orientation	3.49	0.53	3.42	0.60
Grades	3.12	0.64	3.51	0.36
Self esteem	3.90	0.53	3.97	0.47
SES	4.60	0.90	5.25	0.76

Table 1

Descriptions of the 184 Studies Included in the Meta-Analysis

Study	N	Effect size (r)
Adams (1997)	73	.31
Adelabu (2008)	661	.39
Ali (2006)	300	.24
Asner (1999)	86	.01
Bacho (1997)	120	−.10
Barry (2000)	150	−.10
Basurto (1995)	99	.17
Beiser & Hou (2006)	647	.02
Bhadha (2001)	360	.07
Bhargava (2007)	147	−.03

Part 1: Outcomes

• No discrepancy

Part 1: Outcomes

- No discrepancy
- Not enough information

Pre- and Post-Treatment Depressive Symptom Intensity Based on Beck Depressive Inventory (BDI) Scores

BDI Scores	AACBT Group (n = 8)	CBT Group (n = 10)	Average BDI scores in Both Groups
Pre-Treatment BDI	34.4	30.3	32.3
Post-Treatment BDI	21.8	24.4	23.1
Average Decrease	−12.6	−5.9	−9.2

Note: BDI scores can range from 0–60.

$\eta_p^2 = .15$

osf.io/7nsmd/

Part 1: Outcomes

- No discrepancy
- Not enough information
- Different effect

Subsequently, repeated-measures analyses were conducted within each age group to evaluate the Valence × Subtype interaction. Among the young group, a significant Valence × Subtype interaction emerged, $F(2, 64) = 5.80, p < .01, \eta_p^2 = .13$. Simple effects analyses within each of the two levels of valence were conducted, revealing a significant main effect of subtype upon the proportion of positive words falsely recalled, $F(2, 65) = 3.02, p = .05, \eta_p^2 = .09$, and the proportion of
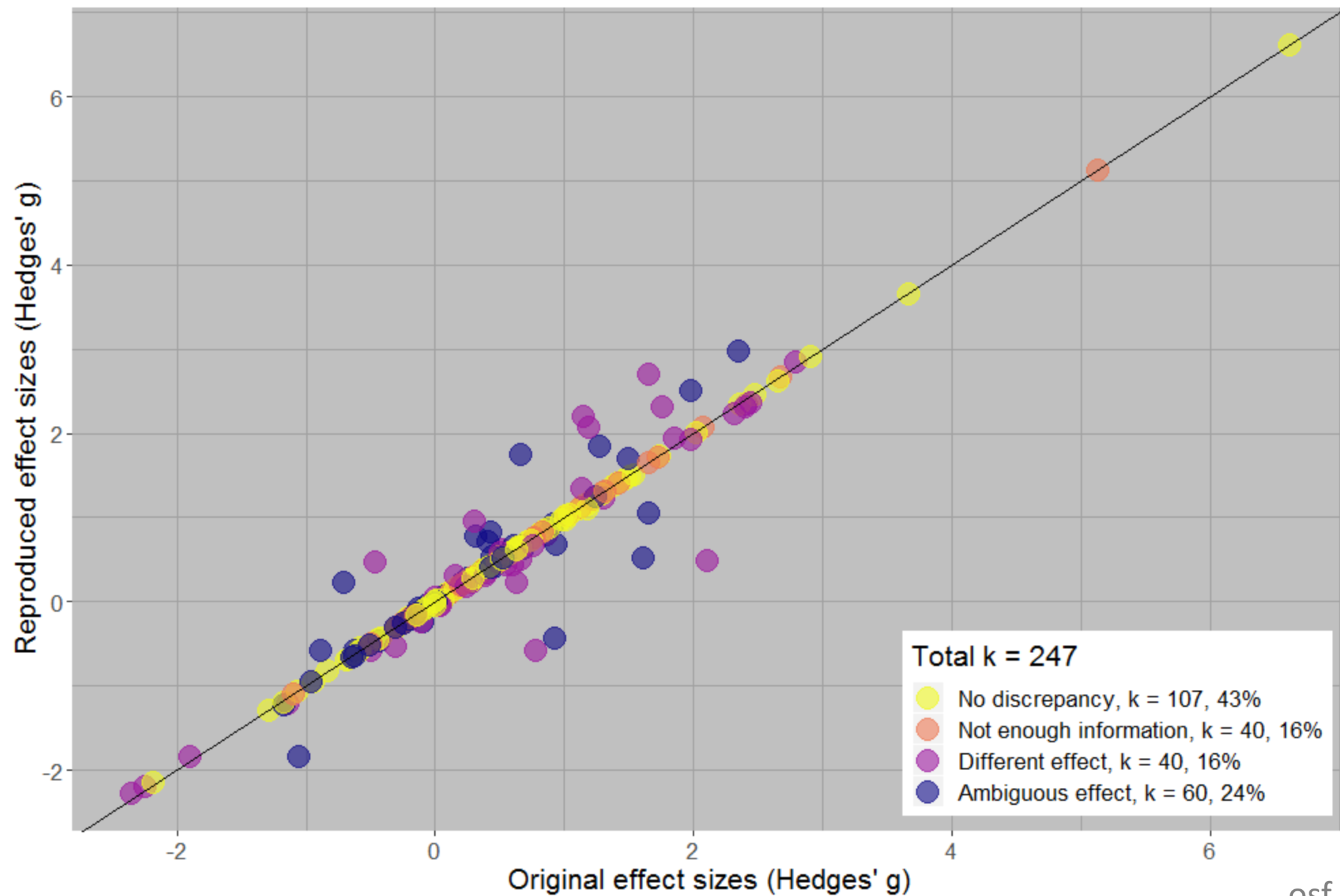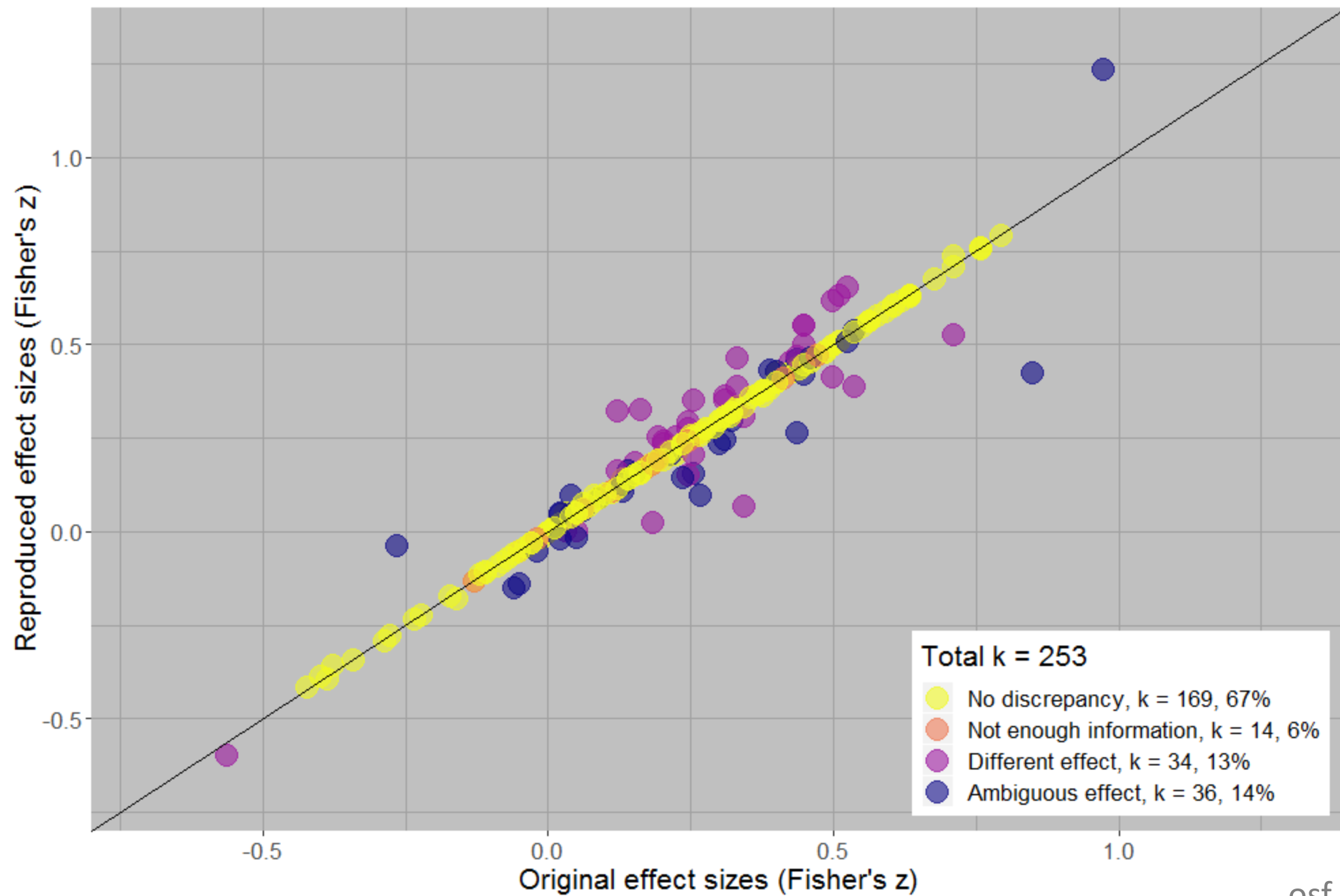
$\eta_p^2 = .15$

Part 1: Outcomes

- No discrepancy
- Not enough information
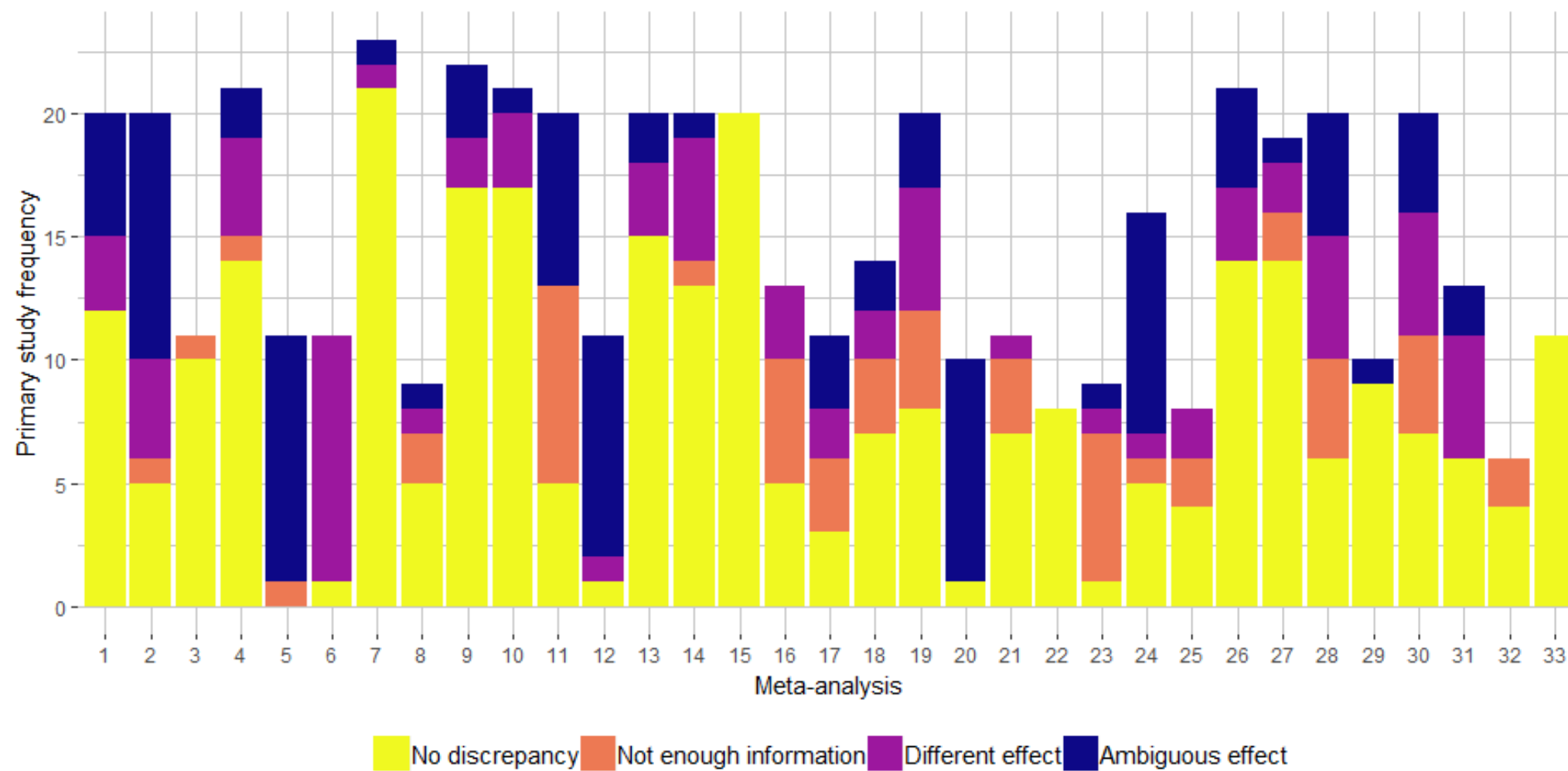- Different effect
- Ambiguous effect

osf.io/7nsmd/

Part 1: Outcomes

Table 2. Group Means and Standard Deviations for Social Problem Solving Indices

| | | | Standard SPS | | | | Directed Learning | | | | Discovery Learning | |
Variable	(n)	Pre	Post	Follow	(n)	Pre	Post	Follow	(n)	Pre	Post	Follow
Primary	(26)	19.25	26.62	23.38	(24)	20.64	24.72	24.17	(29)	19.86	24.31	21.52
SPS		5.36	6.83	7.48		6.81	6.81	9.83		5.56	7.69	8.39
Obstacle	(26)	13.67	18.00	15.85	(24)	13.28	17.96	17.42	(29)	15.86	17.35	14.90
SPS		7.92	4.69	7.38		7.00	3.56	5.09		6.41	4.78	6.34
Primary	(26)	7.17	8.97	8.37	(24)	7.12	8.55	8.00	(29)	7.13	8.36	8.21
Expectancies		1.78	1.20	1.57		1.27	1.29	2.21		1.52	1.14	1.54
Obstacle	(26)	6.53	8.69	8.00	(24)	6.80	9.46	8.67	(29)	7.15	8.52	7.24
Expectancies		3.66	2.02	3.14		3.44	1.70	2.87		2.43	2.72	3.09
Integration	(26)	5.76	6.92	7.26	(24)	5.40	6.96	7.00	(29)	5.97	7.45	6.97
of Story		1.52	1.41	1.77		1.53	1.48	2.13		1.68	1.38	1.66
Parts												
Role	(26)	6.67	7.65	7.48	(24)	7.16	7.31	7.17	(29)	6.86	7.66	7.59
Taking		2.16	3.05	2.39		2.38	2.53	2.94		2.39	2.87	2.69

osf.io/7nsmd/

Total k = 247

- No discrepancy, k = 107, 43%
- Not enough information, k = 40, 16%
- Different effect, k = 40, 16%
- Ambiguous effect, k = 60, 24%

osf.io/7nsmd/

Total k = 253

- No discrepancy, k = 169, 67%
- Not enough information, k = 14, 6%
- Different effect, k = 34, 13%
- Ambiguous effect, k = 36, 14%

osf.io/7nsmd/

osf.io/7nsmd/

Part 1: Conclusion

45% of individual study effect sizes irreproducible

- Not enough information (k = 54, 11%)

- Different effect calculated (k = 74, 15%)

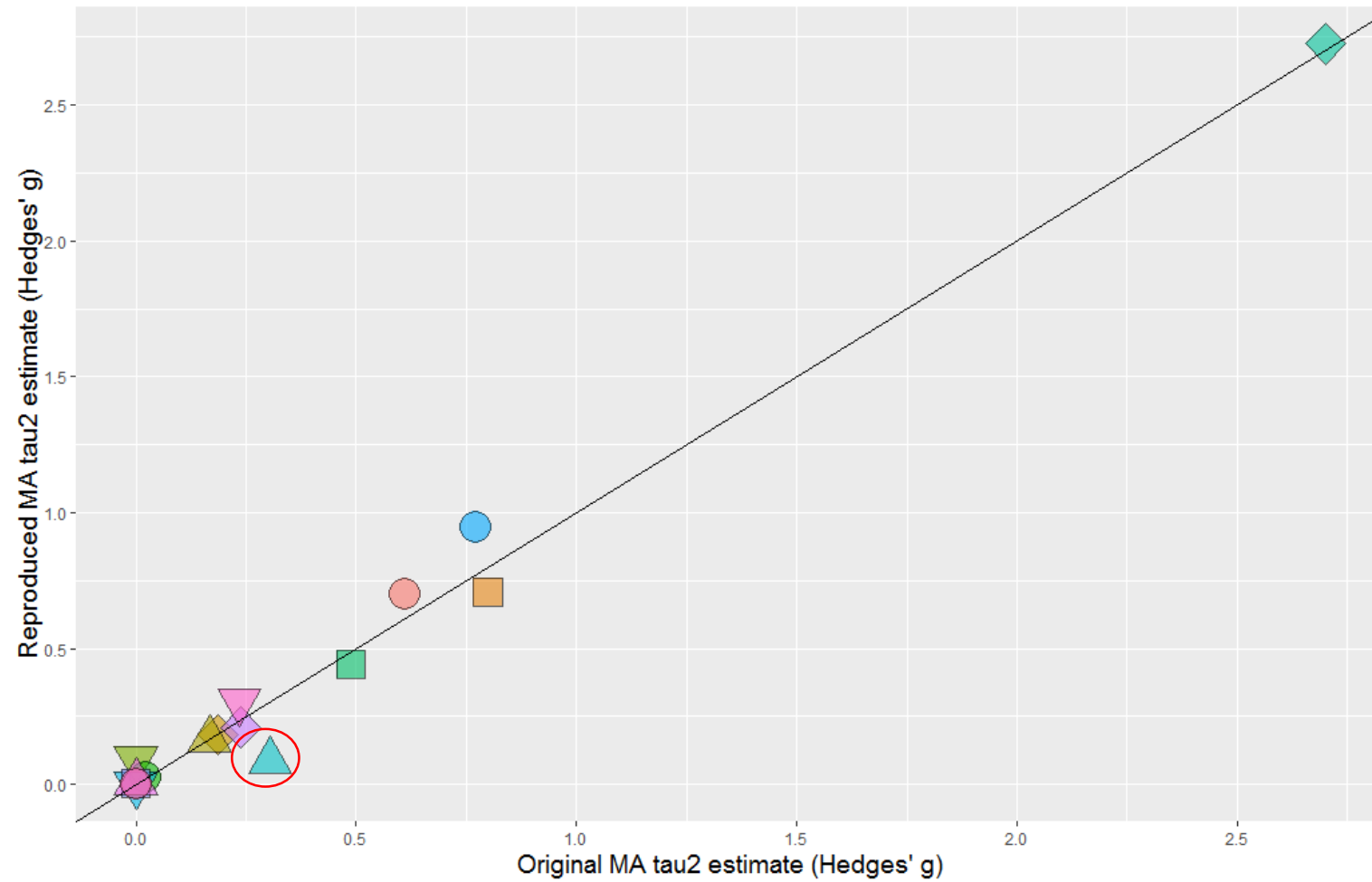- Ambiguous effect (k = 96, 19%)

osf.io/7nsmd/

Part 2: Outcomes

Original vs. reproduced meta-analysis:

- Pooled effect size estimate

- Pooled effect size confidence interval

- Between-study variance estimate τ^2
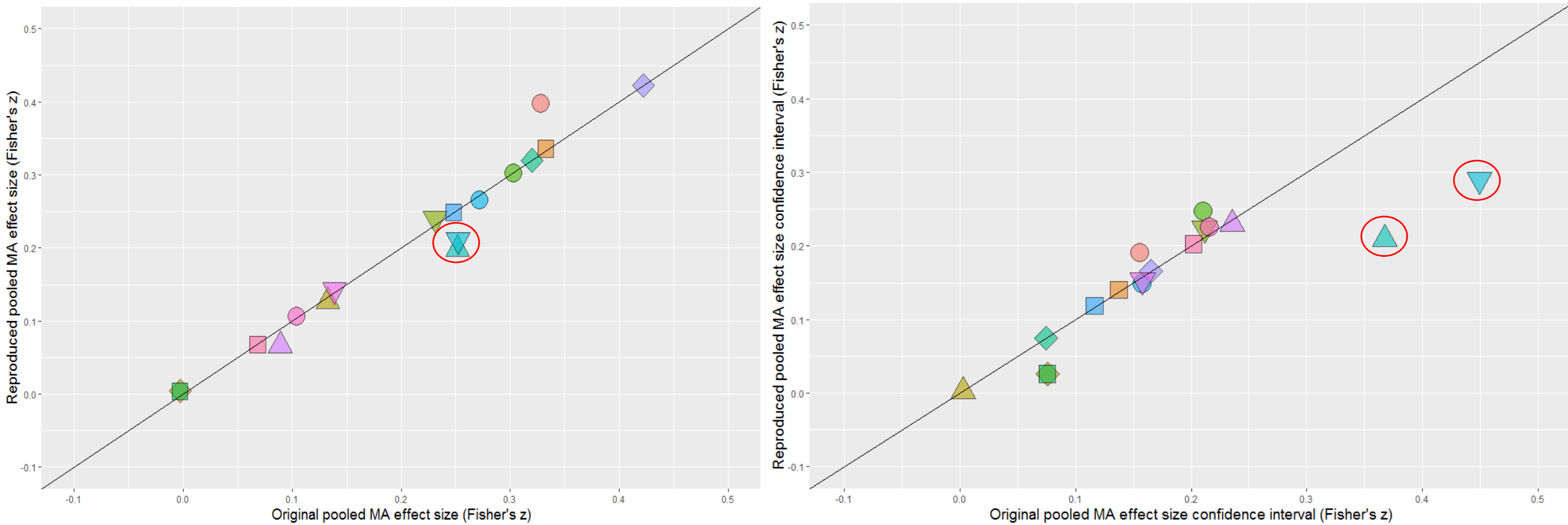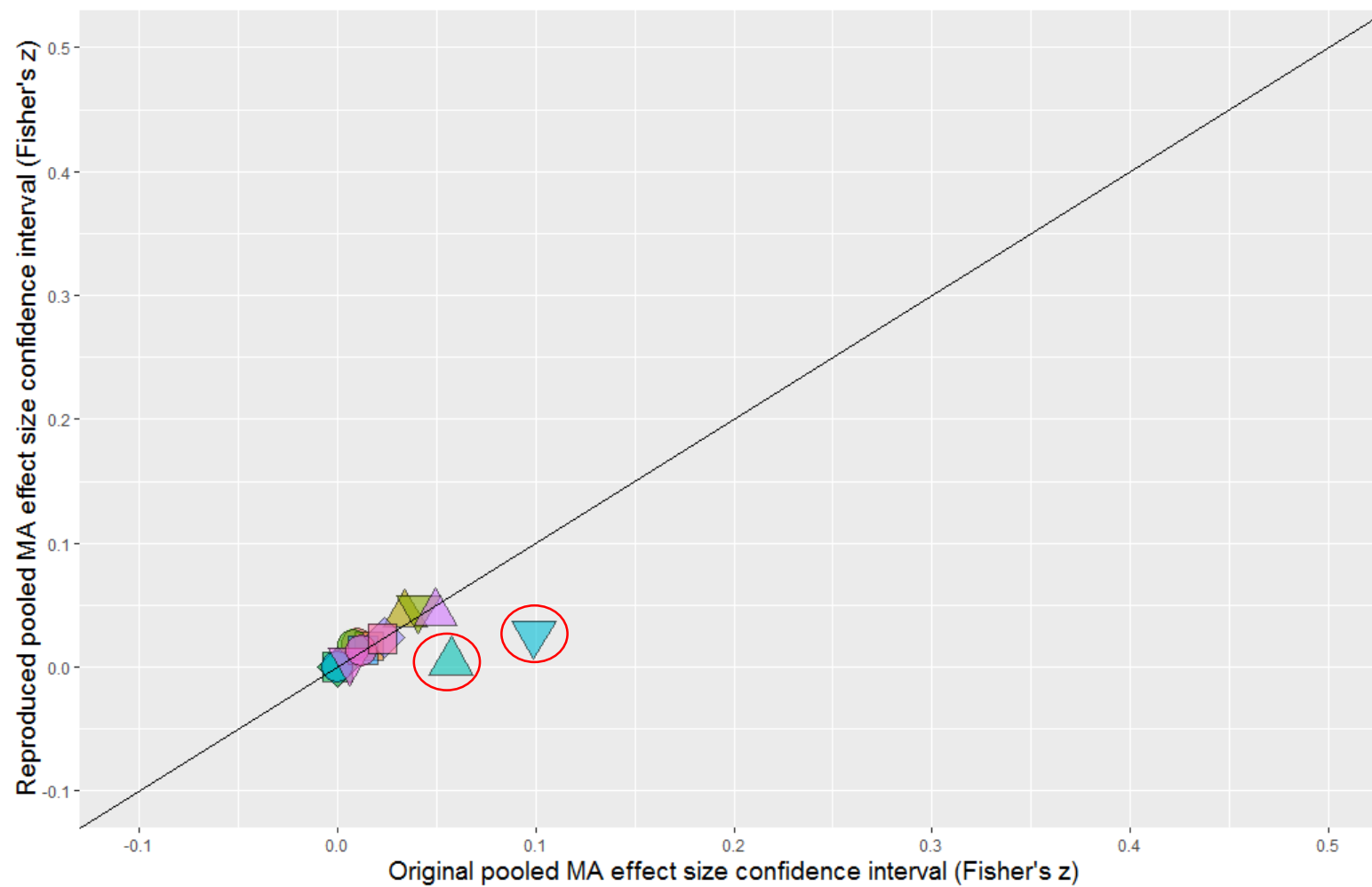
SMDs: Pooled MA effect size (L) and CI (R)

osf.io/7nsmd/
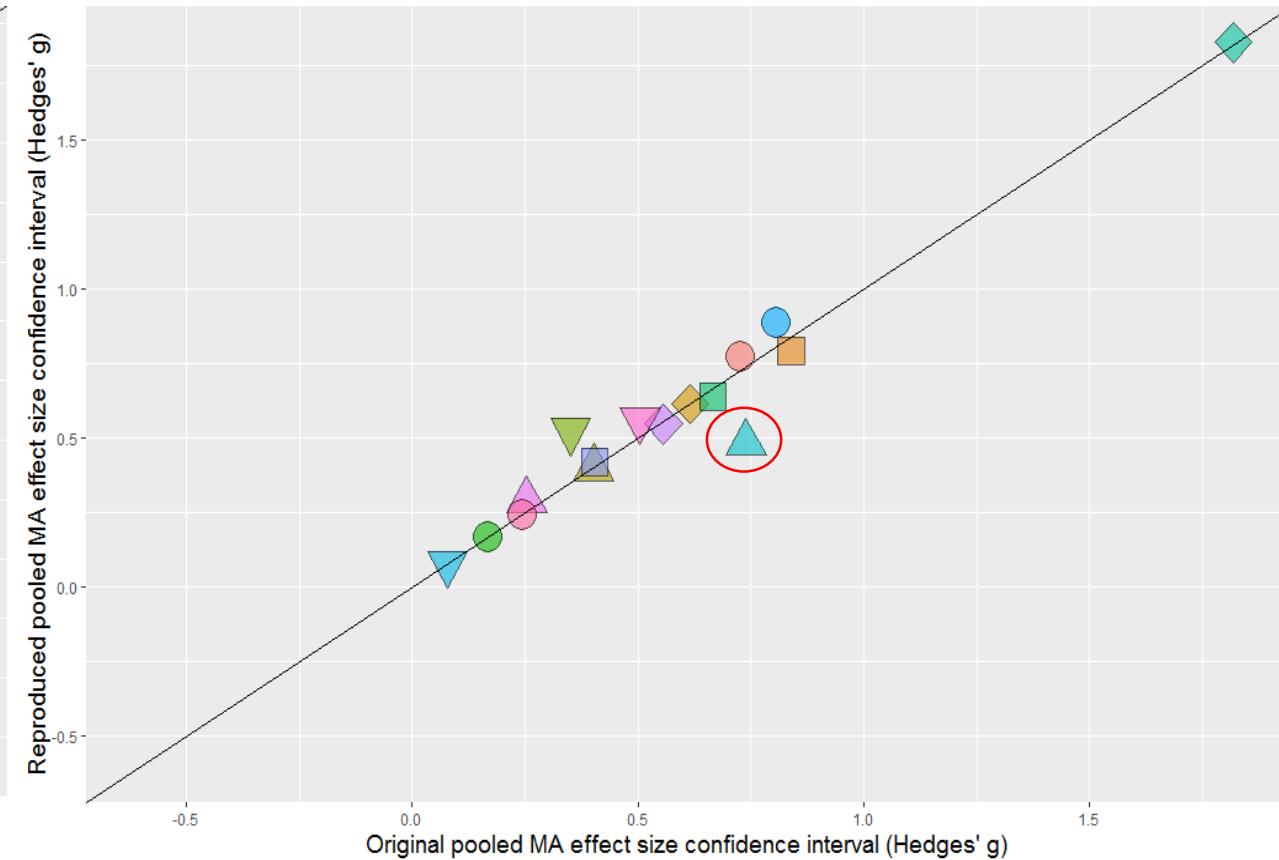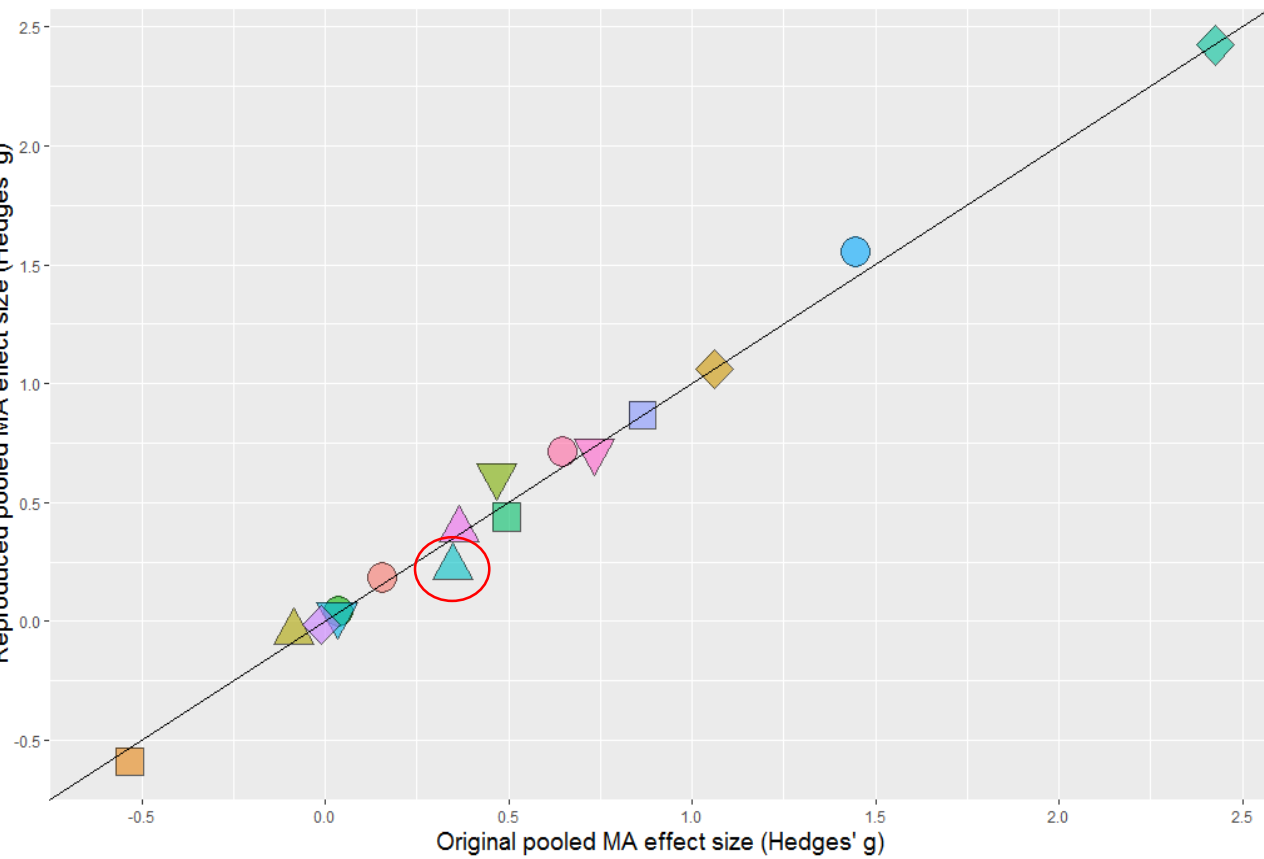
SMDs: τ^2

Correlations: Pooled MA effect size (L) and CI (R)

Correlations: τ^2

Conclusion

45% of individual study effect sizes not reproducible

Small to moderate effects on meta-analytic estimates

No changes in statistical significance

But... substandard reporting practices

osf.io/7nsmd/

What can we do?

- Preregistration
- Detailed coding information
  - Refer to effects as mentioned in the individual study paper, or page no.
  - Combining effect sizes
  - Effect size calculation
- Dynamic documenting
- Meta-Analysis Reporting Standards (MARS)
- Sharing data, materials, and workflow
- → Do a pilot study!

Thank you!

✉ e.maassen@uvt.nl

🐦 @esthermaassen

osf.io/7nsmd/

metaresearch.nl

TILBURG UNIVERSITY