Comparison of the performance of methods to assess publication bias in real data

Helen Niemeyer, Robbie C. M. van Aert, Sebastian Schmid, Dominik Uelsmann, Christine Knaevelsrud & Olaf Schulte-Herbrueggen

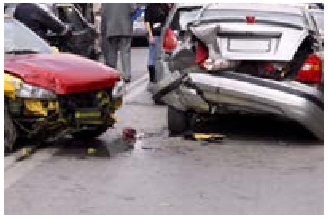
Open Science Conference

Leibniz Institute, Trier, 14.03.2019

Publication bias assessment in real data

- **Why posttraumatic stress disorder (PTSD)?**
    - Severe mental disorder
    - 9,8% (life-time) prevalence





Diagnostic Criteria (DSM-5)

- Exposure to actual or threatened death, serious injury, or sexual violence
- Intrusion symptoms
- Persistent avoidance of stimuli associated with the traumatic event(s)
- Negative cognitions and mood
- Alterations in arousal and reactivity

(American Psychiatric Association, 2013; Kilpatrick et al., 2013)

Publication bias assessment in real data

- **Why posttraumatic stress disorder (PTSD)?**
  - Evidence-based psychotherapy
  - Scientist-practitioner model
  - Treatment studies since 1980
  - Publication bias not assessed
  - Biased effect estimates can overestimate the efficacy

  - Treatments include exposure to the traumatic event
  - Potential detrimental effects

(Bisson et al., 2013; Driessen et al., 2017; Ioanndis, 2008; Shapiro & Forrest, 2001)

Publication bias assessment in real data

Assessment in psychotherapy research

- Assessment in published meta-analyses
- Extent in whole fields of research and with multiple methods
- Assessment in 1 of 46 (2.17%) meta-analyses for treatment of schizophrenia and 12 of 85 (14.12%) for depression
- 15 % of meta-analyses for schizophrenia and depression affected by bias
- Rank-correlation test, Egger´s regression test, trim and fill
- Simulation studies show that methods differ in statistical properties

(Banks et al., 2012; Begg & Mazumdar, 1994; Carter et al., 2017; Coburn & Vevea, 2015; Duval & Tweedie, 2000a, 2000b; Egger et al., 1997; Ioannidis, 2009; Ioannidis & Trikalinos, 2007; Kepes et al., 2012; McShane et al., 2016; Niemeyer et al., 2012, 2013; Simonsohn et al., 2014; van Assen et al., 2015)

Study design

Aims

1. Investigate the degree and impact of publication bias in meta-analyses for the treatment of PTSD
2. Compare the performance of publication bias methods

Inclusion criteria

Meta-analyses

- Any form of psychotherapy
- Clinical or subclinical PTSD
- Adults
- 1980 - 09/2015
- English or German language
- Published and unpublished

Data sets

> 6 primary studies

Homogeneity (*Q*-test, I^2)

Summary effect size reported

(Rothstein, Sutton & Borenstein, 2005; Sterne, Becker, & Egger, 2005)

Study design

Literature search

- Data bases — *PsycINFO, Psyndex, PubMed and Cochrane Database*
- Search terms — *"metaana*" OR "meta-ana*" OR "review" OR "Übersichtsarbeit" AND "stress disorders, post traumatic" (MeSH) OR "post-trauma*" OR "posttrauma*" OR "posttraumatic stress disorder" OR "trauma*" OR "PTSD" OR "PTBS"*

- Snowball search system
- Contacted authors and retrieved primary studies if data not reported

- Three independent raters for
  - inclusion of studies
  - data extraction and
  - coding of treatment (standardized coding form)

(Lipsey & Wilson, 2001)

Study design

Six methods to assess publication bias

Examining the presence of publication bias

Egger's regression test

Rank correlation test

Test of excess significance (TES)

P-uniform

Testing the null hypothesis of no effect and correcting effect estimates for publication bias

Trim and fill

PET-PEESE

P-uniform

(Begg & Mazumdar, 1994; Duval & Tweedie, 2000a, 2000b; Egger, Davey Smith, Schneider & Minder, 1997; Ioannidis & Trikalinos, 2007; Stanley & Doucouliagos, 2012; van Assen et al., 2015)

Study design

Publication bias test

- Two-tailed hypothesis tests for Egger´s regression test and rank-correlation test
- One-tailed hypothesis tests for TES and *p*-uniform

Test of no effect and estimating corrected effect sizes

- Fixed (FE) or random-effects (RE) model depending on original meta-analysis
- Null hypothesis of no effect: two-tailed test
- Estimates transformed to Cohen´s *d* for effect size corrections
- Difference scores for each methods´ effect size estimate to estimate of traditional meta-analysis

- Alpha level of .05 for all analyses

Study design

Congruence of methods

Loevinger's H for dichotomous decision (statistically significant or not) of publication bias tests and test of no effect

		Method II		
		Not sig.	Sig.	
Method I	Not sig.	A	B	P1
	Sig.	C	D	Q1
	Total	P2	Q2	

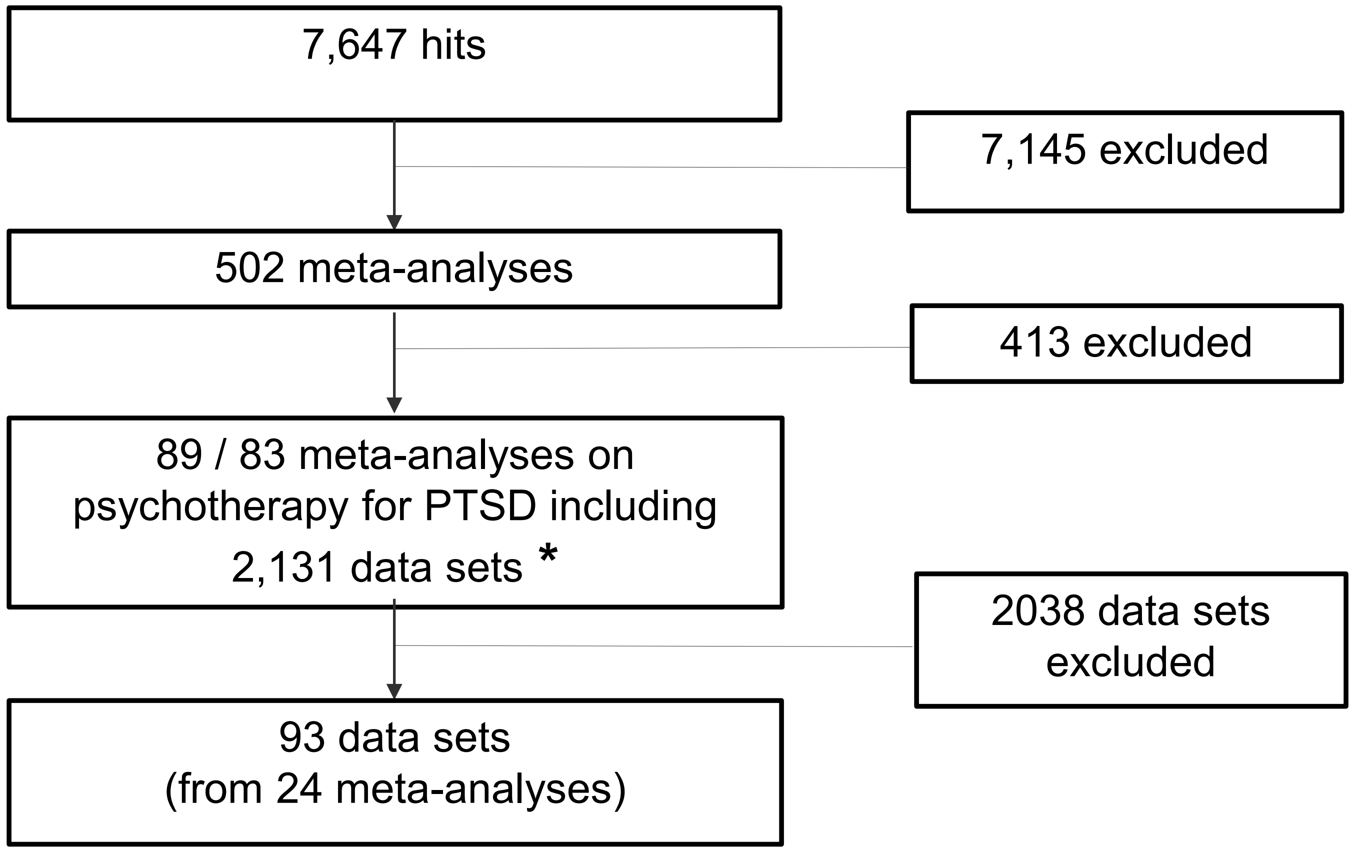
$$H = \frac{(A \times D - B \times C)}{\min(P1 \times Q2, P2 \times Q1)}.$$

R 3.3.0., metafor and puniform package

See OSF for more information, data and R code (https://osf.io/9b4df/, https://osf.io/m34wc/, https://osf.io/frm8h/)

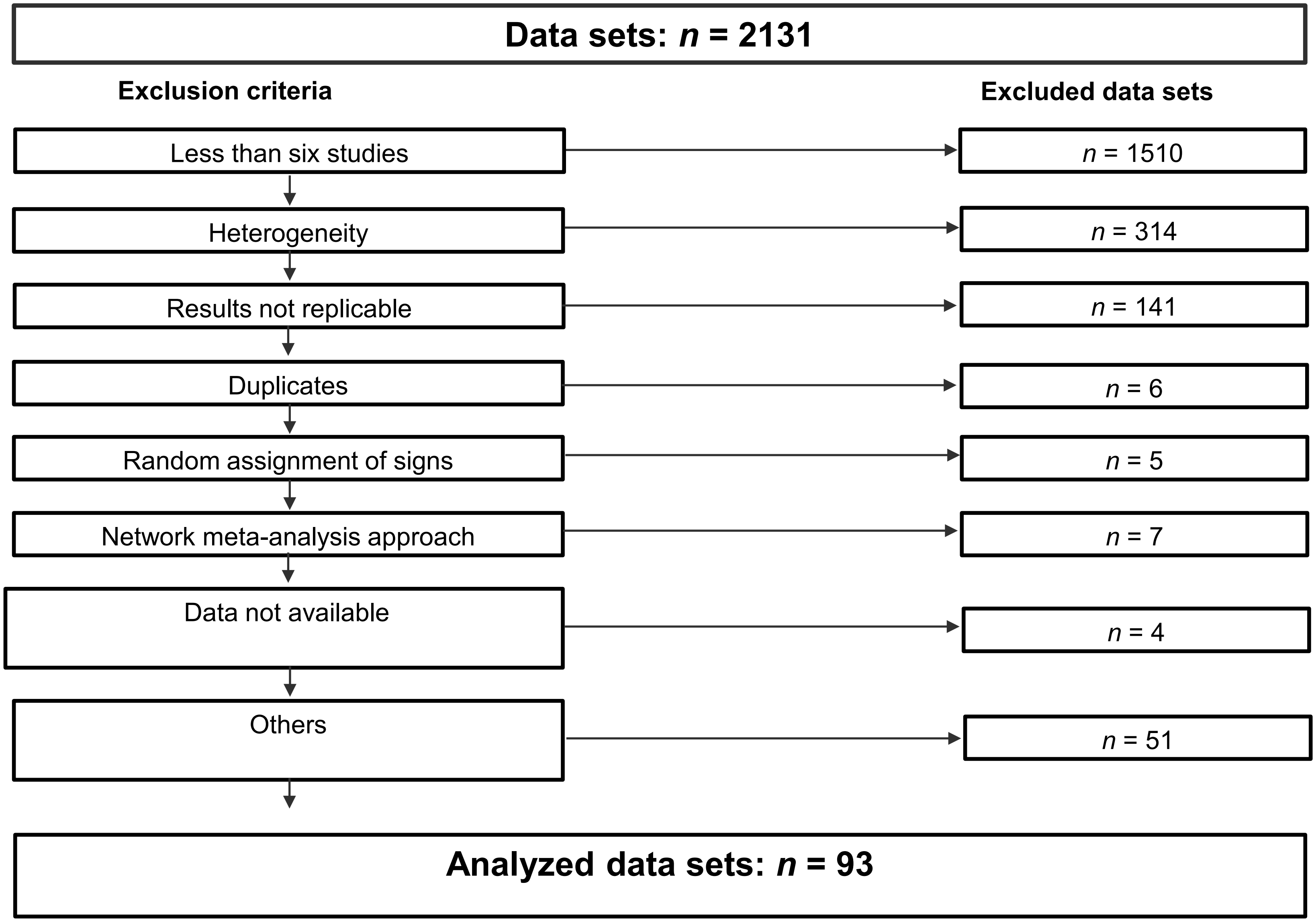
(Loevinger, 1948; Mokken, 1971; R Core Team, 2016; Viechtbauer, 2010)

Results of the literature search



- Primary studies n = 447
  See https://osf.io/pkzx8/

Selection of data sets



Results - descriptives

- Interrater reliability: Fleiss' Kappa $\kappa = .973$, $p < .001$
- 20 meta-analyses (24.14%) included unpublished studies
- Median number of effect sizes in a data set = 7 (first quartile 7, third quartile 10)
- Median number of statistically significant effect sizes = 3 (33.3%; first quartile 1 (10%), third quartile 6 (83.3%))
- 72 data sets (77.4%) included at least one significant effect size
- Median I^2-statistic = 12.5% (first quartile 0%, third quartile 36.3%)

(Fleiss, 1971)

Publication bias tests

Number of data sets and percentages where the tests were statistically significant

a)

		Rank correlation		
		Not sig.	Sig.	
Egger	Not sig.	74	2	76 (81.7%)
	Sig.	14	3	17 (18.3%)
	Total	88 (94.6%)	5 (5.4%)	H = .511

b)

		Rank correlation		
		Not sig.	Sig.	
TES	Not sig.	88	4	92 (98.9%)
	Sig.	0	1	1 (1.1%)
	Total	88 (94.6%)	5 (5.4%)	H = 1

c)

		Rank correlation		
		Not Sig.	Sig.	
p-uniform	Not sig.	64	5	69 (95.8%)
	Sig.	3	0	3 (4.2%)
	Total	67 (93.1%)	5 (6.9%)	H = -.075

d)

		Egger		
		Not Sig.	Sig.	
TES	Not sig.	75	17	92 (98.9%)
	Sig.	1	0	1 (1.1%)
	Total	76 (81.7%)	17 (18.3%)	H = -.224

Publication bias tests

e)		Egger		
		Not sig.	Sig.	
	Not sig.	52	17	69 (95.8%)
p-uniform	Sig.	3	0	3 (4.2%)
	Total	55 (76.4%)	17 (23.6%)	H = -.309

f)		TES		
		Not sig.	Sig.	
	Not sig.	68	1	69 (95.8%)
p-uniform	Sig.	3	0	3 (4.2%)
	Total	71 (98.6%)	1 (1.4%)	H = -.043

- Of all methods Egger's regression test detected publication bias the most, i.e. in 17 data sets (18.3%)
- At most two methods detected publication bias test in the same data set, which occurred in 4 data sets (4.3%)

Tests of no effect

Number of data sets where the null-hypothesis of no effect was rejected

a)

		Meta-analysis		
		Not sig.	Sig.	
T&F	Not sig.	24	4	28 (30.1%)
	Sig.	4	61	65 (69.9%)
	Total	28 (30.1%)	65 (69.9%)	H = .796

b)

		Meta-analysis		
		Not sig.	Sig.	
PET-PEESE	Not sig.	27	46	73 (78.5%)
	Sig.	1	19	20 (21.5%)
	Total	28 (30.1%)	65 (69.9%)	H = .831

c)

		Meta-analysis		
		Not sig.	Sig.	
p-uniform	Not sig.	8	23	31 (43.1%)
	Sig.	0	41	41 (56.9%)
	Total	8 (11.1%)	64 (88.9%)	H = 1

d)

		T&F		
		Not sig.	Sig.	
PET-PEESE	Not sig.	26	47	73 (78.5%)
	Sig.	2	18	20 (21.5%)
	Total	28 (30.1%)	65 (69.9%)	H = .668

Tests of no effect

e)		T & F		
		Not sig.	Sig.	
	Not sig.	10	21	31 (43.1%)
p-uniform	Sig.	0	41	41 (56.9%)
	Total	10 (13.9%)	62 (86.1%)	H = 1

f)		PET-PEESE		
		Not sig.	Sig.	
	Not sig.	29	2	31 (43.1)
p-uniform	Sig.	24	17	41 (56.9%)
	Total	53 (73.6%)	19 (26.4%)	H = .756

Estimating corrected effect sizes

Mean of difference in effect size estimates

		Meta-analysis
Mean of difference, median, (SD)	Trim and fill	-0.018, -0.018 (0.104)
	PET-PEESE (72 data sets)	-0.108, -0.009 (0.886). -0.158, -0.049 (0.971)
	p-uniform (72 data sets)	0.002, 0.043 (0.355)

		Positive meta-analytic estimates	Negative meta-analytic estimates
Mean, median, [min.; max.], (SD) of estimates	Meta-analysis	.574, 0.405, [0.015;1.683], (0.509)	-0.614, -0.573, [-1.288;-0.092], (0.362)
	Trim and fill	0.533, 0.405, [-0.047;1.426], (0.446)	-0.6, -0.557, [-1.288;-0.095], (0.351)
	PET-PEESE	0.112, 0.033, [-1.656;3.075], (0.829)	-0.413, -0.358, [-1.965;1.074], (0.614)
	p-uniform	0.339, 0.791, [-6.681;1.884], (2.069)	-0.641, -0.673, [-2.158;2.938], (0.721)

Discussion

- Hardly any evidence for publication bias
- 20 (24.14%) meta-analyses included unpublished studies
- 22.6% of the data sets did not include a significant result
- Low power due to small number of primary studies
- Tests did not often result in the same conclusion for the same data sets
- No large differences between estimates of traditional meta-analysis and those corrected for publication bias
- Larger imprecision in effect size estimates of PET-PEESE and *p*-uniform
- Unrealistic if small number of (significant) studies and small variation in standard errors of studies

Discussion

Limitations

- Heterogeneous data sets excluded based on *Q*-test
- Low power when applied to a small number of effect sizes
- Questionable research practices may have further biased the results

Planned extensions

- Inclusion of data sets based on $I^2 < 50\%$
- Simulation study
- Selection model

(Hardy & Thompson, 1998; Simmons, Nelson, & Simonsohn, 2011; Thompson & Pocock, 1991)

Thank you very much for your attention!

Any questions?

Literature

American Psychological Association. (2010). *Publication manual of the American Psychological Association.* Washington, DC: Author.

American Psychological Association (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author.

Bakker, van Dijk, A. & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science, 7,* 543–554.

Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: the antagonist of meta-analytic reviews and effective policymaking*. Educational Evaluation and Policy Analysis, 34*, 259-277. doi:10.3102/0162373712446144

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088–1101. http://dx.doi.org/10.2307/2533446.

Bisson, J. I., Roberts, N. P., Andrew, M., Cooper, R., & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database of Systematic Reviews, 12*. doi:10.1002/14651858.CD003388.pub4

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley. doi:10. 1002/9780470743386

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017). *Correcting for bias in psychology: A comparison of meta-analytic methods*. Retrieved from osf.io/preprints/psyarxiv/9h3nu

Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological Methods, 20*, 310-330. doi:10.1037/met0000047

Duval, S. & Tweedie, R. (2000a). A Nonparametric “Trim and Fill“ Method of Accounting for Publication Bias in Meta-Anlaysis. *Journal of the American Statistical Association, 95,* 89-98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis.) *Biometrics, 56*, 455-463. doi:10.1111/j.0006-341X.2000.00455.x

Literature

Driessen, E., Hollon, S. Treatment D., Bockting, C. L. H., & Cuijpers, P. (2017). Does Publication Bias Inflate the Apparent Efficacy of Psychological for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials. *PLOS ONE, 10*, e0137864. doi:doi:10.1371/journal.pone.0137864

Egger, M., Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. British *Medical Journal, 315*, 629-634.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in medicine, 17,* 841-856.

Ioannidis, J. P. (2009). Integration of evidence from multiple meta-analyses: A primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Canadian Medical Association Journal, 181*, 488-493. doi:10.1503/cmaj.081086

Ioannidis, J. P. A & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. Clin Trials, 4, 245 – 255. DOI: 10.1177/1740774507079441.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15*, 624-662.

Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National Estimates of Exposure to Traumatic Events and PTSD Prevalence Using DSM-IV and DSM-5 Criteria. *Journal of Traumatic Stress, 26*, 537-547. doi:10.1002/jts.21848

Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin, 45*, 507-529. doi:10.1037/h0055827

McShane, B. B., Boeckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives On Psychological Science, 11*, 730-749. doi:doi:10.1177/1745691616662243

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*, 264-269.

Mokken, R. J. (1971). A theory and procedure of scale analysis: With applications in political research (Vol. 1): Walter de Gruyter.

Literature

Niemeyer, H., Musch, J., & Pietrowsky, R. (2012). Publication bias in meta-analyses of the efficacy of psychotherapeutic interventions for schizophrenia. *Schizophrenia Research, 138*, 103–112.

Niemeyer, H., Musch, J., & Pietrowsky, R. (2013). Publication Bias in Meta-Analyses of the Efficacy of Psychotherapeutic Interventions for Depression. *Journal of Conculting and Clinical Psychology, 81*, 58 – 74.

R Core Team (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication Bias in Meta-Analysis*. Wiley, West Sussex.

Shapiro, F., & Forrest, M. S. (2001). *Eye movement desensitization and reprocessing: Basic principles, protocols and procedures*. New York, NY: Guilford Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359-1366. doi:10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and effect size correcting for publication bias using only significant results. *Perspectives On Psychological Science, 9*, 666-681. doi:10.1177/1745691614553988

Stanley, T. D. & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Res. Syn. Meth., 5,* 60–78.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*, 108-112. doi:10.2307/2684823

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111-125). Chichester, England: Wiley.

Thompson, S. G., & Pocock, S. J. (1991). Can meta-analyses be trusted? *The Lancet, 338*, 1127-1130

Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Literature

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-Analysis Using Effect Size Distributions of Only Statistically Significant Studies. *Psychological Methods, 20,* 293–309.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36*, 1-48.