Meta-analysis and publication bias

Robbie C.M. van Aert

R.C.M.vanAert@tilburguniversity.edu www.robbievanaert.com

Tilburg University Department of Methodology and Statistics



Session 11 Meta-Analysis



www.bitss.org @ucbitss

RT2 Roadmap

Motivating Issues

Researchers degrees of freedom Scientific misconduct Publication bias Failure to replicate

To achieve

Open materials, data, code, & access Transparent reporting & disclosure Reproducible & replicable results Cumulative meta-analyses

Organized Workflow and File Management (OSF, Github)

Design	Conduct	Dissemination	Archiving
Pre-Registration	Data Management	Transparent Reporting & Disclosure	Data Repositories
Pre-Analysis Plans	Version Control	Preprints	Dynamic Documents
Power Planning	Open Notebooks (Jupyter/Docker)	Open Access	



www.bitss.org @ucbitss

Robbie van Aert, Tilburg University: I am lucky, because I work primarily with simulated data mainly for applying methods in the social sciences.



Something about myself...

- Recently finished my PhD thesis on meta-analysis and publication bias methods
- Now: Postdoc at Tilburg University working on meta-analysis and publication bias methods
- Meta Research Center: <u>www.metaresearch.nl</u>



Overview

1. Introduction to meta-analysis

Lunch break!

- 2. Introduction to publication bias
- 3. Publication bias methods
- 4. Practical part
- 5. Wrap-up/Conclusions



1. Meta-analysis

- Information explosion: more and more studies get published
- It becomes more and more difficult to keep up with reading all the relevant literature
- Methods are needed to summarize research findings, and to give an objective overview
- But how to do this?!



1. Meta-analysis: Some history

- <u>Prior to 1990s</u>: Narrative literature review where a expert reads the literature and answers a research question
- Drawbacks of narrative literature reviews:
 - Subjective
 - Lack of transparency
 - Hard to update if new information becomes available

• Vote counting: # significant results vs. # nonsignificant results



1. Meta-analysis: Some history

- Now: Systematic review and meta-analysis
- Systematic review: clear set of rules that are specified in advance with respect to inclusion or exclusion of studies
- Meta-analysis: "the statistical synthesis of the data from separate but similar studies leading to a quantitative summary" (Last, 2001)
- Goals of meta-analysis:
 - Estimating average effect size (and between-study variance)
 - Examine whether differences in effect sizes are caused by study characteristics



1. Meta-analysis

Number of published meta-analyses increases:





1. Meta-analysis: Stages

- I. Formulating a problem/research question
- II. Literature search
- III. Extracting information from literature
- IV. Data preparation (converting effect sizes)
- V. Combining effect sizes (meta-analysis)

Books on how to do a systematic review:

- Cooper et al., (2009). The handbook of research synthesis and metaanalysis
- Cooper (2010). Research synthesis and meta-analysis: A step-by-step approach

- VI. Interpretation and sensitivity analysis
- VII. Presentation of results



1. Meta-analysis: Models

- Meta-analysis is a weighted average of studies' effect sizes
- Two types of meta-analysis models: fixed-effect (or common-effect) and random-effects
- Fixed-effect: inference on the studies included in the meta-analysis
- Random-effects: studies are sample of a population of studies and we want to generalize results to this population
- Theoretical arguments should motivate model selection!









- All studies estimate the same population effect size θ
- Model: $y_i = \theta + \varepsilon_i \text{ with } \varepsilon_i \sim N(0, \sigma_i^2)$
- Parameter estimate: $\hat{\theta} = \frac{\sum w_i y_i}{\sum w_i}$ with $w_i = \frac{1}{\sigma_i^2}$ and $Var[\hat{\theta}] = \frac{1}{\sum w_i}$

• Inference: $z = \frac{\hat{\theta}}{\sqrt{Var[\hat{\theta}]}}$ and $\hat{\theta} \pm 1.96\sqrt{Var[\hat{\theta}]}$











- Studies' effect sizes are sampled form a population of effects with mean μ and variance τ^2

• Model:
$$y_i = \overline{\mu + \mu_i} + \varepsilon_i$$
 with $\varepsilon_i \sim N(0, \sigma_i^2)$ and $u_i \sim N(0, \tau^2)$

- Parameter estimate: $\hat{\mu} = \frac{\sum w_i y_i}{\sum w_i}$ with $w_i = \frac{1}{\sigma_i^2 + \hat{\tau}^2}$ and $\text{Var}[\hat{\mu}] = \frac{1}{\sum w_i}$
- Inference: $z = \frac{\hat{\mu}}{\sqrt{Var[\hat{\mu}]}}$ and $\hat{\mu} \pm 1.96\sqrt{Var[\hat{\mu}]}$



- Meta-analysis on psi a.k.a. extrasensory perception
- Psi denotes "anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms" (Bem, 2011)
- Paper by Bem (2011) contains 9 experiments with 8 of them yielding significant results in favor of psi



• Example of an experiment by Bem (2011):





• Example of an experiment by Bem (2011):



• Future position of erotic picture was more frequently correctly identified: 53.1%, *t*(99) = 2.51, *p* = .01, *d* = 0.25



- Multiple studies were conducted and both the existence and absence of psi was found
- Random-effects meta-analysis based on 90 studies: $\hat{\mu} = 0.09$, z=6.40, p < .001
- <u>Conclusion</u>: Psi does really exist, and we can really look into the future
- Or... is this meta-analysis biased because of, for instance, publication bias and questionable research practices?



1. Meta-analysis: Meta-regression

- Heterogeneity or between-study variance in true effect size implies that the primary studies' true effect size differ (so $\tau^2 > 0$)
- This heterogeneity can be attributed to random or systematic differences between the true effect sizes
- Systematic differences:
 - Methodological differences between primary studies
 - Differences in the studied population
 - Differences in the length of a treatment
- Characteristics of primary studies can be included in the model to explain this between-study variance



1. Meta-analysis: Meta-regression

• Fixed-effects with moderators model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$$

• Mixed-effects model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \mu_i + \varepsilon_i$$

 τ² is also estimated in mixed-effects model now referring to the amount of residual between-study variance after including the moderators in the model



1. Meta-analysis: Meta-regression

- Meta-regression may reveal interesting relationships among the variables
- However, one cannot make causal statements about these relationships → observational study instead of experiment
- Meta-regression used for *hypothesis generating* → relationships among variables should be studied in a new experiment or RCT



1. Meta-analysis: Quantifying heterogeneity

- Many estimators exist for estimating τ^2 :
 - DerSimonian and Laird is most often used
 - Restricted maximum likelihood and Paule-Mandel are nowadays recommended
- Estimates of τ^2 are imprecise if the meta-analysis contains a small number of effect sizes
- Q-profile and generalized Q-statistic method can be used for computing confidence interval around $\hat{\tau}^2$
- Drawback of
 ² → cannot be used for comparing the amount of heterogeneity across meta-analyses



1. Meta-analysis: Quantifying heterogeneity

• For that reason, the *P*-statistic was proposed:

$$l^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + s^2}$$

where s² is an estimate of the "typical within-study variance"

- The *P*-statistic computes the proportion of total variance that can be attributed to between-study variance
- The *P*-statistic ranges from 0 to 1 (0.25 low, 0.5 medium, 0.75 large)
- Q-profile and generalized Q-statistic method can also be used for constructing a confidence interval around the *P*-statistic



1. Meta-analysis: Software

- R (metafor and meta packages)
- STATA: metan() command
- SPSS: not included, but macros can be used
- SAS: SAS PROC MEANS program
- Comprehensive Meta-analysis Software (CMA)
- Excel (add in MetaEasy)
- RevMan from Cochrane Collaboration
- MetaWin
- Multilevel software



1. Meta-analysis: Other models

- Meta-Analytic Structural Equation Modelling (MASEM)
- Multivariate meta-analysis
- Network meta-analysis
- Multilevel meta-analysis
- Individual patient/participant data (IPD) analysis
- Bayesian statistics



1. Meta-analysis: Criticism

- Meta-analysis is an exercise of mega-silliness (Eysenck, 1978)
- Meta-analysis is statistical alchemy for the 21st century (Feinstein, 1995)

Main criticisms:

Pub

۲

Mixing apples and oranges

Ź

• Garbage in. garbag



Concluding remarks

Take-home message 1:

- Meta-analysis is a powerful tool to aggregate findings from different studies
- Quality of the data determines the quality of the meta-analysis
- Theoretical arguments should motivate model selection (FE or RE)
- Explaining heterogeneity/between-study variance → no causal statements



Lunch break!





2. Publication bias

- A video: https://www.youtube.com/watch?v=iC_1WpZOLE8
- This was Slade Manning playing with ping pong balls
 - A 3 minutes video based on 3 (!) years playing
 - Some tricks needed 5,000 attempts
- Slade Manning about the video:

"I didn't really have any skill or control, so it was just a matter of hitting balls over and over until one finally happened to go the right distance and direction."

 <u>Conclusion</u>: What you see is not all what happened → this also holds for science, but it will not be as bad as in the video


2. Publication bias

- Publication bias is "the selective publication of studies with a statistically significant outcome"
- Longer history in dealing with publication bias in medical research than social sciences
- Nowadays, increased attention for publication bias in various fields



- Evidence for publication bias is overwhelming
- 95% of published articles contain significant results in psychology
- But this is not in line with average statistical power is (about 20-50%)
- Assuming power is 50% → only 1 out of 40 nonsignificant results get published

		Physical sc
Space Science (SP, N=104)		Biological sc.
Geosciences (GE, N=127)		Social sc.
Environment/Ecology (EE, N=149)	1	I = life
Plant and Animal Sciences (PA, N=193)	1	s = soft
Computer Science (CS, N=63)	a +	
Agricultural Sciences (AG, N=109)	la 🛏	
Physics (PH, N=71)		
Neuroscience & Behaviour (NB, N=143)	I ⊢	
Microbiology (MI, 140)	I	\vdash
Chemistry (CH, N=95)	•	-
Social Sciences, General (SO, N=144)	ls ⊢	\rightarrow
Immunology (IM, N=145)	I ⊢	-
Engineering (EN, N=77)	a	
Molecular Biology & Genetics (MB, N=126)	I ⊢	
Economics & Business (EB, N=117)	as ⊢	7-1
Biology & Biochemistry (BB, N=113)	1 ⊢	
Clinical Medicine (CM, N=130)	la ⊢	-
Pharmacology & Toxicology (PT, N=142)	la ⊦	
Materials Science (MS, N=105)	a ⊦	
Psychiatry/Psychology (PP, N=141)	l s	,
50 Pa	% 60% 70% 80%	90% 100 r the tested H



Adapted from Fanelli (2010)

- Fanelli (2012) studied percentage of significant results in literature between 1990-2007 across disciplines
- Increase in significant results from 70.2% (1990) to 85.9% in (2007)



Coursol and Wagner (1986) surveyed researchers on the effects of positive findings

Table 1

Relation Between Outcome (Positive vs. Neutral or Negative) and Decision to Submit Research for Publication

Direction of outcome	Submission decision		
	Yes	No	Total
Positive (Client improved)	106	23	129
Neutral or negative	28	37	65
(Client did not improve)			
Total	134	60	194



Coursol and Wagner (1986) surveyed researchers on the effects of positive findings

Table 2

Relation Between Outcome (Positive vs. Neutral or Negative) and Acceptance of Research Submitted for Publication

Direction of outcome	Accepted	Not accepted	Total	
Positive (Client improved)	85	21	106	
Neutral or negative	14	14	28	
(Client did not improve)				
Total	99	35	134	



Coursol and Wagner (1986) surveyed researchers on the effects of positive findings

Table 3

Relation Between Outcome (Positive vs. Neutral or Negative) and Final Disposition of Study (Published vs. Unpublished)

Direction of outcome	Published	Not published	Total
Positive (Client improved)	85	44	129
Neutral or negative	14	51	65
(Client did not improve)			
Total	99	95	194



- Open Science Collaboration initiated Reproducibility Project which was a large-scale replication attempt of psychological research
- 100 studies were replicated from three flagship journals: JPSP, Psychological Science, and Journal of Experimental Psychology
- Results shocked many people inside and outside academia:
 - 97% of original studies were significant and only 36% of replications
 - Effect size estimates decreased from *r*=0.4 to 0.2





p-value in Original study

- Experimental economics: 89% of original studies were significant and 69% of replications
- Hematology and oncology: 11% of studies were deemed to be successfully replicated
- Substantial amount of critique on these projects
- Two possible causes of this low replicability:
 - Publication bias
 - Questionable research practices



2. Publication bias: Consequences

• What do you think are consequences of publication bias? Why is publication bias detrimental for science?



- Three consequences:
 - Type-I errors \rightarrow False impression that an effect exists
 - Overestimation of effect size
 - Questionable research practices



3. Publication bias methods

- Multiple methods have been developed to examine publication bias
- Methods to assess publication bias:
 - Failsafe N
 - Funnel plot
 - Egger's test
 - Rank-correlation test
 - *p*-uniform's publication bias test
- Methods to correct effect size estimates:
 - Trim-and-fill method
 - Selection models
 - *p*-uniform and *p*-curve
 - PET-PEESE



3. Publication bias methods: Example

- Meta-analysis by Rabelo et al. (2015) on the effect of weight on judgments of importance
- <u>Theory:</u> the physical experience of weight influences how much importance people assigns to things, issues, and people
- Meta-analysis based on 25 studies: $\hat{\mu} = 0.571$, $\hat{\tau}^2 = 0$, 95% CI (0.468; 0.673), z = 10.904, p < .001



3. Failsafe N

- Unpublished studies are hidden in the *file drawers* of researchers
- Failsafe *N* computes number of effect sizes with $\theta = 0$ that need to be retrieved before the meta-analytic estimate is no longer significantly different from zero
- Well-known and popular method, but discouraged to be used
- Drawbacks of Failsafe *N*
 - Focus on statistical rather than substantive significance
 - Effect size of hidden studies is assumed to be zero
- 1098 (!) effect sizes with $\theta = 0$ are needed in example



3. Funnel plot

- Funnel plot shows relationship between effect size and its precision
- An asymmetric funnel suggests the presence of *small-study effects*
- Eyeballing a funnel plot is unreliable, so tests were developed



Effect size



3. Funnel plot

- Funnel plot shows relationship between effect size and its precision
- An asymmetric funnel suggests the presence of *small-study effects*
- Eyeballing a funnel plot is unreliable, so tests were developed



Effect size



3. Funnel plot asymmetry tests

- Two most often used tests for funnel plot asymmetry are rankcorrelation test and Egger's test
- Rank-correlation test ranks the effect size and standard error and then computes the correlation between these ranks (τ=0.6, p<.0001)



3. Funnel plot asymmetry tests

- Egger's test fits a regression line through the points in a funnel plot
- Vertical line suggests a symmetric funnel
- If slope is significantly different from zero → funnel plot asymmetry
- *z* = 1.629, *p* = .103





Effect size

3. Funnel plot asymmetry tests

- Two most often used tests for funnel plot asymmetry are rankcorrelation test and Egger's test
- Rank-correlation test ranks the effect size and standard error and then computes the correlation between these ranks (τ=0.6, p<.0001)
- Drawbacks of these tests:
 - Low statistical power and are recommended not to be used with only 10 effect sizes
 - Test small-study effects and not publication bias
- Low power, so is it not better to correct estimates for publication bias?!



3. Trim-and-fill method

- Popular method to correct effect size estimate
- Missing effect sizes from ۲ one side of funnel plot are "trimmed" and "filled" in other side
- Method is discouraged to be ۲ used due to misleading results (Terrin et al., 2003)
- $\hat{\mu} = 0.571$ and after imputing





Effect size



3. Selection models

- Selection model approaches combine effect size and selection model
 - Effect size model: Distribution of effect size
 - Selection model: Mechanism that determines which studies are observed
- Very many different selection model approaches exist
- Some selection models estimate selection model whereas others assume that selection model is known
- Not often used in practice, because sophisticated assumptions have to be made and convergence problems may arise



3. Selection models

- Hardly any user-friendly software exist for applying selection model approaches
- R package "weightr" exists to apply the Vevea & Hedges weightfunction model
- Applying weight-function model to example: $\hat{\mu} = 0.571$ vs. 0.266 (*p*=.0002)
- Promising method → good statistical properties in recent simulation studies (Carter et al., 2017; McShane et al., 2016)



3. PET-PEESE

- Estimate equals the effect size where standard error is zero (infinite sample size)
- Performance of PET-PEESE is topic of further study
- <u>Limitation</u>: Studies' sample size should be different from each other
- Estimate is $\hat{\mu} = 0.571$ vs. 0.066 (*p*=.472)



Effect size



- [Robbie adds disclaimer]
- Both methods are based on the same methodology, but slightly differ in implementation
- Methods use the probability of observing a particular effect size conditional on the effect size being statistically significant



How are one-tailed *p*-values, $P(y \ge y_i; \theta = 0)$, distributed computed from a random sample of *N*(**0.2**, 0.04)?





How are one-tailed *p*-values, $P(y \ge y_i; \theta = 0)$, distributed computed from a random sample of $N(\mathbf{0}, 0.04)$?







How are one-tailed *p*-values <u>at the true effect size</u> $\theta = 0.2$, $P(y \ge y_i; \theta =$



TILBURG

- Both methods are based on the same methodology, but slightly differ in implementation
- Methods use the probability of observing a particular effect size conditional on the effect size being statistically significant
- Statistical principle: *p*-values are not only uniformly distributed under the null hypothesis, but also at the true effect size
- Methods discard nonsignificant effect sizes



• Conditional *p*-values are computed with:

 $\frac{P(y \ge y_i; \theta)}{P(y \ge y_{cv}; \theta)}$

where y_{cv} denotes the critical value (effect size)

- Effect size estimate is obtained when these conditional *p*-values are uniformly distributed
- Assumptions of the methods:
 - Significant effect sizes have equal probability of getting published
 - Effect sizes are statistically independent
- Note: Both methods take sampling variance in primary studies into account and are not solely based on the (conditional) *p*-values

• Example with three observed effect sizes (δ =0.5):

t(48)=3.133, p=.0029 t(48)=2.302, p=.011 t(48)=2.646, p=.025





Example with three observed effect sizes (δ=0.5):
t(48)=3.133, p=.0029
t(48)=2.302, p=.011
t(48)=2.646, p=.025



Example with three observed effect sizes (δ=0.5):
t(48)=3.133, p=.0029
t(48)=2.302, p=.011
t(48)=2.646, p=.025



- Effect size estimate is...
 - < 0 if p > .025
 - -0 if p = .025
 - > 0 if p < .025
- *p*-uniform has some advantages over *p*-curve (van Aert et al., 2016):
 - Effect size can always be estimated
 - Estimation of a confidence interval
 - Publication bias test
- Limitations:
 - Overestimation caused by moderate to large between-study heterogeneity
 - Unpredictable bias in effect size estimates caused by *p*-hacking/QRPs



3. *p*-uniform (and *p*-curve): Heterogeneity

- Simonsohn et al. (2014) state that *p*-curve (and *p*-uniform) yield an accurate estimate if heterogeneity is present
- Simulation study with two-independent groups design and δ =0.397



3. *p*-uniform (and *p*-curve): Heterogeneity

	No	Moderate	Large	Larger	Very large
<i>p</i> -curve	.393	.530	.703	.856	1.094
<i>p</i> -uniform	.387	.522	.679	.776	.903
FE	.553	.616	.738	.875	1.104
RE	.553	.616	.743	.897	1.185

- Recommendation:
 - At most moderate: interpret as average *true* effect size
 - More than moderate: interpret as estimate of only the significant studies
 - If possible, create homogeneous subgroups of studies

3. *p*-uniform (and *p*-curve): Heterogeneity

- Simonsohn et al. (2014) state that *p*-curve (and *p*-uniform) yield an accurate estimate if heterogeneity is present
- Simulation study with two-independent groups design and δ =0.397
- We are now working on *p*-uniform* which also includes nonsignificant effect sizes to deal with heterogeneity
- *P*-uniform* estimates both the average effect size and the betweenstudy variance



3. *p*-uniform (and *p*-curve): *p*-hacking

- *P*-hacking (or QRPs) is a term for all behaviors that researchers can use to obtain desirable results
- If *p*-hacking would always result in *p*-values just below the α-level the methods will underestimate the true effect size
- Simulation study with *p*-hacking:
 - Optional stopping
 - Only reporting the first significant dependent variable
 - Only reporting the most significant dependent variable


3. *p*-uniform (and *p*-curve): *p*-hacking



- Recommendation:
 - Be reluctant with interpreting the methods' results in case of indications of *p*-hacking

3. *p*-uniform (and *p*-curve): Conclusion

- *P*-uniform and *p*-curve are promising tools, but also have their limitations
- P-uniform^{*}, hopefully, accurately estimates effect size and between-study variance if heterogeneity is present → results look promising!

Software:

- *P*-curve:
- R Code available in Simonsohn et al. (2014)
- Web application: <u>http://p-curve.com/</u>
- *P*-uniform:
 - R package "puniform": <u>https://github.com/RobbievanAert/puniform</u>
 - Web application: <u>https://rvanaert.shinyapps.io/p-uniform</u>



Web application p-uniform

Manual on how to use this application	Effect size estimate p-uniform:								
Paper about p-uniform	estima	te ci.lb	ci.ub	1.0	nval	ksia			
Author: Robbie C.M. van Aert		0.1792	2 -0.2379	0.3545	-1.1814	0.1187	11		
Enter the characteristics of your meta-analysis below:	ics of your met	a-analysis belo	w: iorma	l distributio	n				
Select effect size measure	Select effect size measure Select effect size mea								
One-sample mean	One-sample mean			orm:					
 Two-independent means 	 Two-independent n 	neans							
 One correlation 	 One correlation 								
Alpha level in primary studies (default .05)	Alpha level in prima	ry studies (d	efault .05)	rmal	rmal distribution				
Select direction of effect in primary studies	Select direction of e	ffect in prima	ry studies						
 Right (positive) 	 Right (positive) 		al	ci.lb	ci.ub	pval	Qstat	Qpval	
 Left (negative) 	 Left (negative) 		591	0.3156	0.509	<.001	6.7409	0.8197	
Select estimation method for p-uniform P (Irwin-Hall) LNP LN1MINP KS AD	Select estimation me P (Irwin-Hall) LNP LN1MINP KS AD	ethod for p-u	hiform	form	ed p-value	• • • • • • • •			
Data entry		igi		1					
Select how you will enter data (see manual)		- 4 conc		1.	•				
 Ma CSV file 		0 0	- /	• •					
 Manually in table 		2 Obse							
Enter data via CSV file Choose File data.mccall93.csv Upload complete		0.0	0.0 0.2 Expec	0.4 sted conditio	0.6 0. nal p-values	8 1.0			
Analyze		🛓 Dowr	load Output as	s.pdf					

Web application p-uniform



Web application p-uniform

											_	
Manual on how to use this application				Effect siz	e estim	ate p-uni	iform:					
Paper about p-uniform				ostima	to	cill	ci ub	1.0	nyal	keig		
Author: Robbie C.M. van Aert				estina	ite	CI.ID	ci.ub	L.0	pvai	Kaiy		
				0.179	2 -(0.2379	0.3545	-1.1814	0.1187	11		
Enter the characteristics of your meta-analysis below:				Notes: - p-value a	approxir	mated with	h normal	distributio	n			
Select effect size measure												
 One-sample mean 				Publication bias test p-uniform:								
 Two-independent means 												
 One correlation 	Effect size estimate p-uniform:											
Alpha level in primary studies (default .05)	estimate	ci.lb	ci.ub	L.0	pval	ksig	mal c	listribution				
0.05	0.1792	-0.2379	0.3545	-1.1814	0.1187	11	hare	astribution				
Select direction of effect in primary studies	Notes:	avinated wi	ith norma	Lelietribution								
Right (positive)	- p-value appro	oximated w	itri norma	astribution			1	ci.lb	ci.ub	pval	Qstat	Qpval
Left (negative)	Dublication bios test a uniform						91	0.3156	0.509	< 001	6 7409	0.8197
0 2011 (1.0gali.10)	Fublication b	las test p-	unnom.				~	0.5150	0.505	0.001	0.7405	0.0137
Select estimation method for p-uniform	L.pb	pval										
 P (Irwin-Hall) 	2.6154	0.0045					orme	ed p-value	S			
○ LNP												
LN1MINP	p-value approx	imated with	h normal	distribution				•				
⊖ KS								•••				
AD				value (
				al p-1				1				
Data entry				dition			1	/				
Select how you will enter data (see manual)				1 con			, í •	•				
Ma CSV file				erve		• /*	•					
 Manually in table 				_ 5 Obs								
				Ŭ	1	•						
Enter data via CSV file				0:0 -	<i></i>]			
Choose File data.mccall93.csv					0.0	0.2	0.4	0.6 0.	8 1.0			
Upload complete						Expecte	d conditio	nal <i>p</i> -values				
Analyze				🕹 Dowr	nload O	utput as .	pdf					

4. Practical part

- Assignment:
 - 1. Inspect the data of meta-analysis by Rabelo et al. (2015) with a spreadsheet program (<u>https://goo.gl/E67gMr</u>):
 - » How many significant effect sizes?
 - » What is the (unweighted) mean of the effect sizes?
 - » What estimate of *p*-uniform do you expect based on the *p*-values?
 - Download CSV file of data (<u>https://goo.gl/X8A3Hh</u>), go to <u>https://rvanaert.shinyapps.io/p-uniform</u>, and analyze data with *p*-uniform
 - 3. Interpret the results, was your expected effect size close to *p*-uniform's estimate?



4. Practical part

	А	В	С	D	Е	F	G	Н	Ι	J
1	m1i	m2i	n1i	n2i	sd1i	sd2i	tobs	pval	yi	vi
2	5.802	5.376	26	28	0.76	0.79	2.016	0.048939	0.541201	0.07709
3	4.010476	3.25	21	22	0.725324	1.725659	1.867	0.06901	0.559201	0.097057
4	7.264	6.3	30	30	1.578	1.334	2.554	0.01329	0.651201	0.070436
5	0	-0.4225	50	50	1	1	2.113	0.037185	0.419257	0.040913
6	6.97	<mark>6.09</mark>	50	50	2.03	1.63	2.39	0.018754	0.474359	0.041169

- m1i and m2i: Sample means group 1 and 2
- n1i and n2i: Sample size group 1 and 2
- sd1i and sd2i: Standard deviation group 1 and 2
- tobs: Observed *t*-value
- pval: Two-tailed *p*-value
- yi: Observed standardized effect size (Hedges' g)
- vi: Sampling variance of yi



Concluding remarks

Take-home message 2:

- Publication bias is a major threat to the validity of meta-analyses that causes overestimation in effect size
- Each publication bias method has its own advantages and disadvantages, so use and report multiple methods (triangulation)
- Keep an eye on the development of PET-PEESE, selection model approaches, and *p*-uniform* (and *p*-curve)



Psi meta-analysis

- Does psi really exist?!; Publication bias in the psi meta-analysis?
- Multitude of publication bias methods was applied → no convincing evidence for the presence of publication bias
- Or...
- Characteristics of the data do not suit publication bias methods
- QRPs/p-hacking may be used in the primary studies

Large scale preregistered replication is conducted





5. Wrap-up/final conclusions

Take-home message:

- Meta-analysis is a powerful tool to aggregate findings from different studies
- Quality of the data determines the quality of the meta-analysis
- Publication bias is a major threat to the validity of meta-analyses that causes overestimation in effect size
- Each publication bias method has its own advantages and disadvantages, so use and report multiple methods (triangulation)



Further reading

- General books on systematic reviews and meta-analysis:
 - Cooper et al., (2009). The handbook of research synthesis and meta-analysis
 - Cooper (2010). Research synthesis and meta-analysis: A step-by-step approach
 - Borenstein et al. (2009). Introduction to meta-analysis
- Difference between fixed-effect and random-effects models:
 - Borenstein et al. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis
- Overview of publication bias methods:
 - Jin et al. (2014). Statistical methods for dealing with publication bias in meta-analysis
 - Rothstein et al. (2005). Publication bias in meta-analysis: Prevention, assessment and adjustments
- *P*-uniform and *p*-curve:
 - van Assen et al. (2015). Meta-analysis using effect size distributions of only statistically significant studies
 - van Aert et al. (2016). Conducting meta-analyses based on *p* values: Reservations and recommendations for applying *p*-uniform and *p*-curve
 - Simonsohn et al. (2014). P-curve: A key to the file drawer
 - Simonsohn et al. (2015). *p*-curve and effect size: Correcting for publication bias using only significant results
- PET-PEESE:
- Stanley & Doucouliagos (2014). Meta-regression approximations to reduce publication selection bias
- Selection model approaches:
 - Chapter in Rothstein et al. (2005). Publication bias in meta-analysis: Prevention, assessment and adjustments

Thank you for your attention

R.C.M.vanAert@tilburguniversity.edu

www.robbievanaert.com

