Meta-analysis and publication bias

Robbie C.M. van Aert

R.C.M.vanAert@tilburguniversity.edu

Tilburg University
Department of Methodology and Statistics

Tilburg University

NWO Nederlandse Organisatie voor Wetenschappelijk Onderzoek Maatschappij- en Gedragswetenschappen

Session 11
Meta-Analysis

BITSS CEGA www.bitss.org @ucbitss

RT2 Roadmap

Motivating Issues	To achieve
Researchers degrees of freedom	Open materials, data, code, & access
Scientific misconduct	Transparent reporting & disclosure
Publication bias	Reproducible & replicable results
Failure to replicate	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	

BITSS CEGA 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.

BITSS CEGA www.bitss.org @ucbitss

Something about myself…

- Studied: Research Master at Tilburg University
- Now: PhD student working on meta-analysis and publication bias methods
- Meta Research Center: www.metaresearch.nl

Overview

1. Introduction to meta-analysis
2. Introduction to publication bias

   Short break?!

3. Publication bias methods
4. Practical part
5. Wrap-up/Conclusions

Tilburg University 6

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?!

TILBURG UNIVERSITY 7

1. Meta-analysis: Some history

- Prior to 1990s: 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

TILBURG UNIVERSITY 8

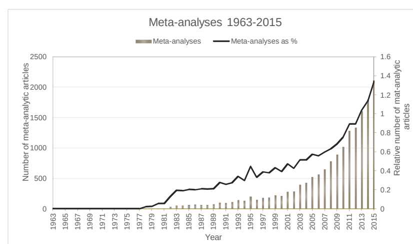
1. Meta-analysis: Some history

- 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

TILBURG UNIVERSITY 9

1. Meta-analysis

- Number of published meta-analyses increases:



TILBURG UNIVERSITY 10

Source: PsycINFO

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

Books on how to do a systematic review:

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

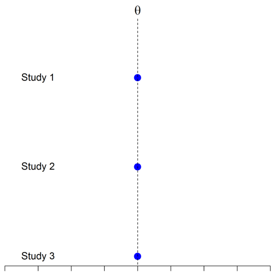
TILBURG UNIVERSITY 11

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!

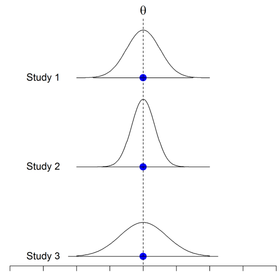
TILBURG UNIVERSITY 12

1. Meta-analysis: Fixed-effect



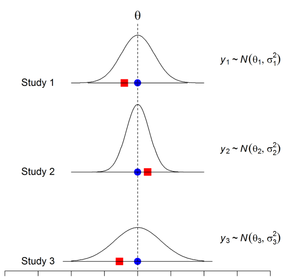
13

1. Meta-analysis: Fixed-effect



14

1. Meta-analysis: Fixed-effect



15

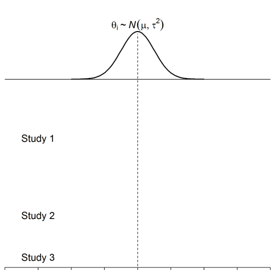
1. Meta-analysis: Fixed-effect

- All studies estimate the same population effect size θ
- Model: $y_i = \theta + \varepsilon_i$ 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 $\hat{\theta} \pm 1.96\sqrt{Var[\hat{\theta}]}$

TILBURG UNIVERSITY

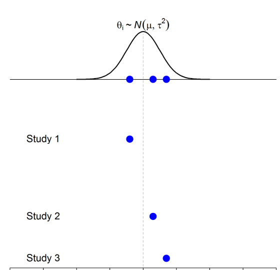
16

1. Meta-analysis: Random-effects



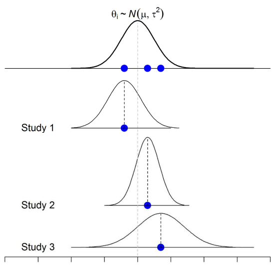
17

1. Meta-analysis: Random-effects



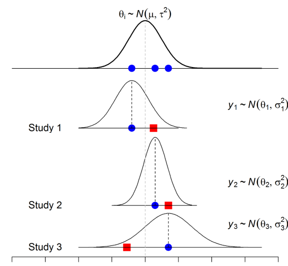
18

1. Meta-analysis: Random-effects



19

1. Meta-analysis: Random-effects



20

1. Meta-analysis: Random-effects

- Studies' effect sizes are sampled form a population of effects with mean μ and variance τ^2
- Model: $y_i = \overbrace{\mu + \mu_i}^{\theta_i} + \varepsilon_i$ 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 Var[$\hat{\mu}$] = $\frac{1}{\sum w_i}$
- Inference: $z = \frac{\hat{\mu}}{\sqrt{Var[\hat{\mu}]}}$ and $\hat{\mu} \pm 1.96\sqrt{Var[\hat{\mu}]}$

TILBURG UNIVERSITY 21

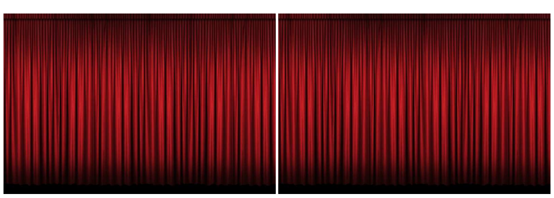
1. Meta-analysis: Example

- 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

TILBURG UNIVERSITY 22

1. Meta-analysis: Example

- Example of an experiment by Bem (2011):



TILBURG UNIVERSITY 23

1. Meta-analysis: Example

- Example of an experiment by Bem (2011):

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

TILBURG UNIVERSITY 24

1. Meta-analysis: Example

- 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
- Conclusion: 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?

TILBURG UNIVERSITY 25

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

TILBURG UNIVERSITY 26

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:

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

- τ^2 is also estimated in mixed-effects model now referring to the amount of residual between-study variance after including the moderators in the model

TILBURG UNIVERSITY 27

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

TILBURG UNIVERSITY 28

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 $\hat{\tau}^2$ → cannot be used for comparing the amount of heterogeneity across meta-analyses

TILBURG UNIVERSITY 29

1. Meta-analysis: Quantifying heterogeneity

- For that reason, the I^2-statistic was proposed:

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

where s^2 is an estimate of the "typical within-study variance"

- The I^2-statistic computes the proportion of total variance that can be attributed to between-study variance
- The I^2-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 I^2-statistic

TILBURG UNIVERSITY 30

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
- …

TILBURG UNIVERSITY 31

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

TILBURG UNIVERSITY 32

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:

- Mixing apples and oranges
- Garbage in, garbag
- Pub

TILBURG UNIVERSITY 33

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

TILBURG UNIVERSITY 34

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."

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

TILBURG UNIVERSITY 35

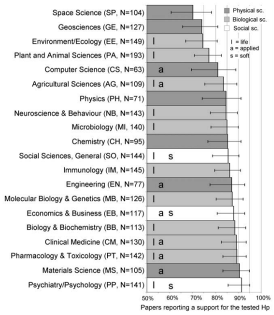
2. Publication bias

- 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

TILBURG UNIVERSITY 36

2. Publication bias: Evidence

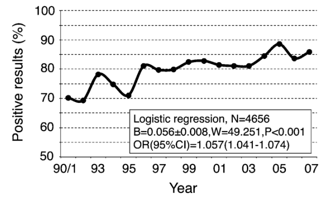
- Evidence for publication bias is overwhelming
- 95% of published articles contain significant results in psychology
- But this is not in line with average statistical power is (about 20-50%)
- Assuming power is 50% → only 1 out of 40 nonsignificant results get published



Adapted from Fanelli (2010)

2. Publication bias: Evidence

- 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)



38

2. Publication bias: Evidence

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

39

2. Publication bias: Evidence

- 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 (Client did not improve)	14	14	28
Total	99	35	134

40

2. Publication bias: Evidence

- 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 (Client did not improve)	14	51	65
Total	99	95	194

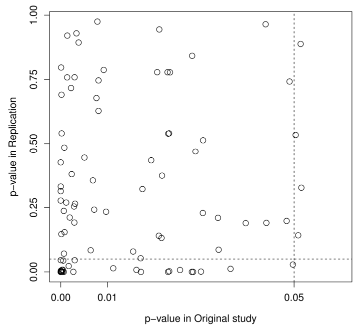
41

2. Publication bias: Evidence

- 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

42

2. Publication bias: Evidence



43

2. Publication bias: Evidence

- 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

TILBURG UNIVERSITY 44

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 → False impression that an effect exists
  - Overestimation of effect size
  - Questionable research practices

TILBURG UNIVERSITY 45

Short break?!

TILBURG UNIVERSITY 46

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

TILBURG UNIVERSITY 47

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

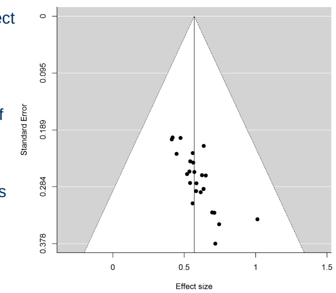
TILBURG UNIVERSITY 48

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

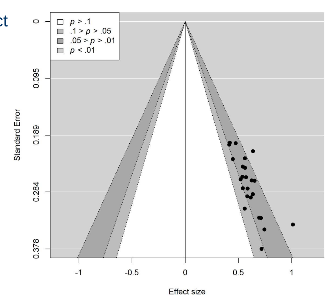
3. Funnel plot

- 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



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

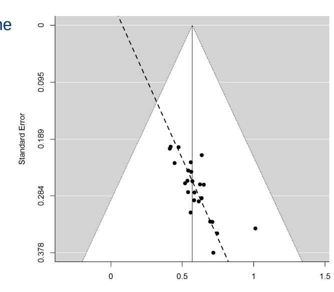


3. Funnel plot asymmetry tests

- Two most often used tests for funnel plot asymmetry are rank-correlation 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

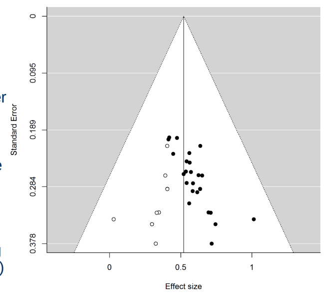


3. Funnel plot asymmetry tests

- Two most often used tests for funnel plot asymmetry are rank-correlation 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 one side of funnel plot are “trimmed” and “filled” in other side
- Method is discouraged to be used due to misleading results (Terrin et al., 2003)
- $\hat{\mu} = 0.571$ and after imputing nine studies 0.521 (p<.0001)



Tilburg University 55

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

Tilburg University 56

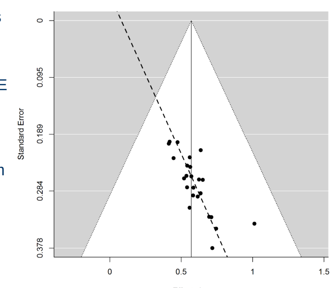
3. Selection models

- Hardly any user-friendly software exist for applying selection model approaches
- R package “`weightr`” exists to apply the Vevea & Hedges weight-function 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)

Tilburg University 57

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
- Limitation: Studies’ sample size should be different from each other
- Estimate is $\hat{\mu} = 0.571$ vs. 0.066 (p=.472)



Tilburg University 58

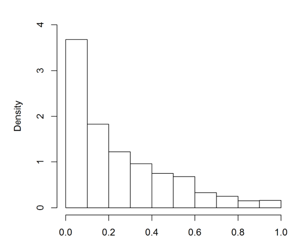
3. *p*-uniform (and *p*-curve)

- [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

Tilburg University 59

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

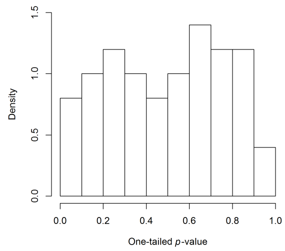
How are one-tailed p-values, $P(y \geq y_i;\ \theta = 0)$, distributed computed from a random sample of N(**0.2**, 0.04)?



Tilburg University 60

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

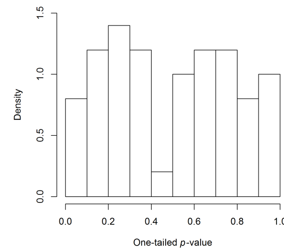
How are one-tailed *p*-values, $P(y \geq y_i;\ \theta = 0)$, distributed computed from a random sample of *N*(**0**, 0.04)?



TILBURG UNIVERSITY

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

How are one-tailed *p*-values <u>*at the true effect size*</u> $\theta = 0.2$, $P(y \geq y_i;\ \theta = 0.2)$, distributed computed from a random sample of *N*(**0.2**, 0.04)?



TILBURG UNIVERSITY

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

- 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

TILBURG UNIVERSITY

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

- Conditional *p*-values are computed with:

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

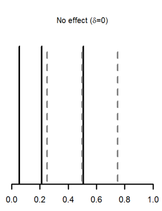
where y_{cv} denotes the critical value (effect size)

- 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

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

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

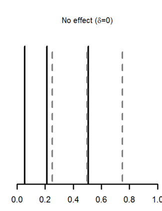
t(48)=3.133, *p*=.0029 *t*(48)=2.302, *p*=.011 *t*(48)=2.646, *p*=.025

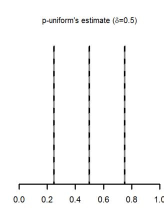


3. *p*-uniform (and *p*-curve)

- 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

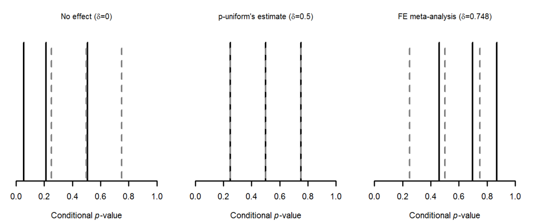




3. *p*-uniform (and *p*-curve)

- 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



67

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

- 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

TILBURG UNIVERSITY 68

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

TILBURG UNIVERSITY 69

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

	No	Moderate	Large	Larger	Very large
p-curve	.393	.530	.703	.856	1.094
p-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

70

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 between-study variance

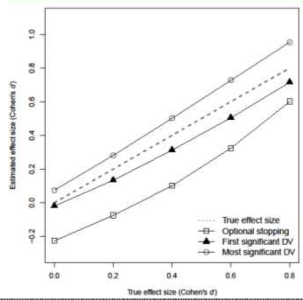
TILBURG UNIVERSITY 71

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

TILBURG UNIVERSITY 72

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



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

73

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

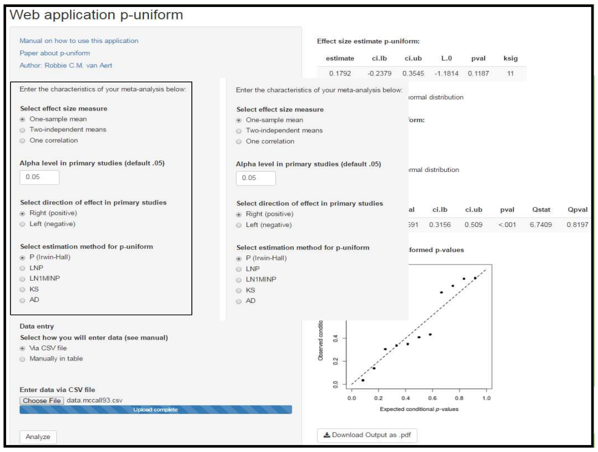
- *P*-uniform and *p*-curve are promising tools, but also have their limitations
- *P*-uniform*, hopefully, accurately estimates effect size and between-study variance if heterogeneity is present → results look promising!

Software:

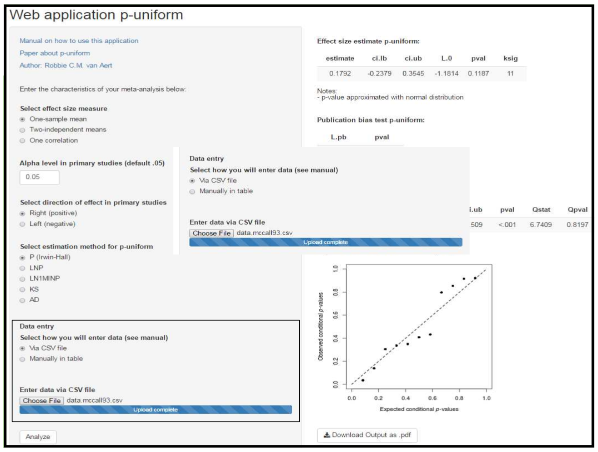
- *P*-curve:
  - R Code available in Simonsohn et al. (2014)
  - Web application: http://p-curve.com/
- *P*-uniform:
  - R package "`puniform`": https://github.com/RobbievanAert/puniform
  - Web application: https://rvanaert.shinyapps.io/p-uniform

TILBURG UNIVERSITY 74

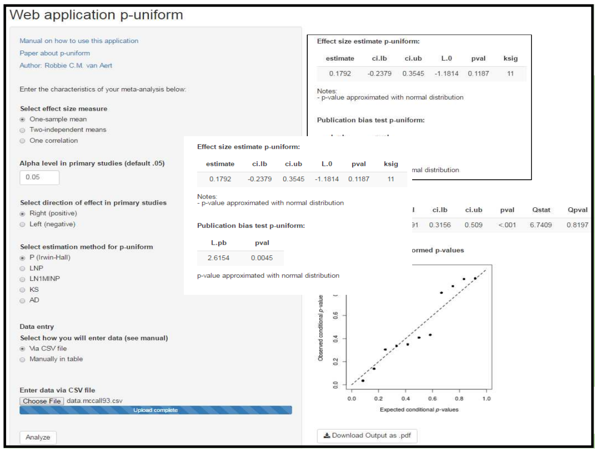
Web application p-uniform



Web application p-uniform



Web application p-uniform



4. Practical part

- Assignment:
  1. Download data of meta-analysis by Rabelo et al. (2015):
     CSV: https://goo.gl/X8A3Hh
     Excel: https://goo.gl/E67gMr
  2. Inspect the data:
     » 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?
  3. Go to https://rvanaert.shinyapps.io/p-uniform and analyze data (CSV file!) with *p*-uniform
  4. Interpret the results, was your expected effect size close to *p*-uniform's estimate?

TILBURG UNIVERSITY 78

4. Practical part

	A	B	C	D	E	F	G	H	I	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	6.09	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

TILBURG UNIVERSITY 79

Concluding remarks

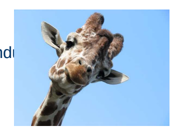
Take-home message 2:

- 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)

TILBURG UNIVERSITY 80

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 condu



TILBURG UNIVERSITY 81

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)

TILBURG UNIVERSITY 82

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

83

Thank you for your attention

R.C.M.vanAert@tilburguniversity.edu

TILBURG UNIVERSITY 84