P-uniform*: A new meta-analytic method to correct for publication bias

Robbie C.M. van Aert & Marcel A.L.M. van Assen

October 12, 2019

European Research Council

Established by the European Commission

Message

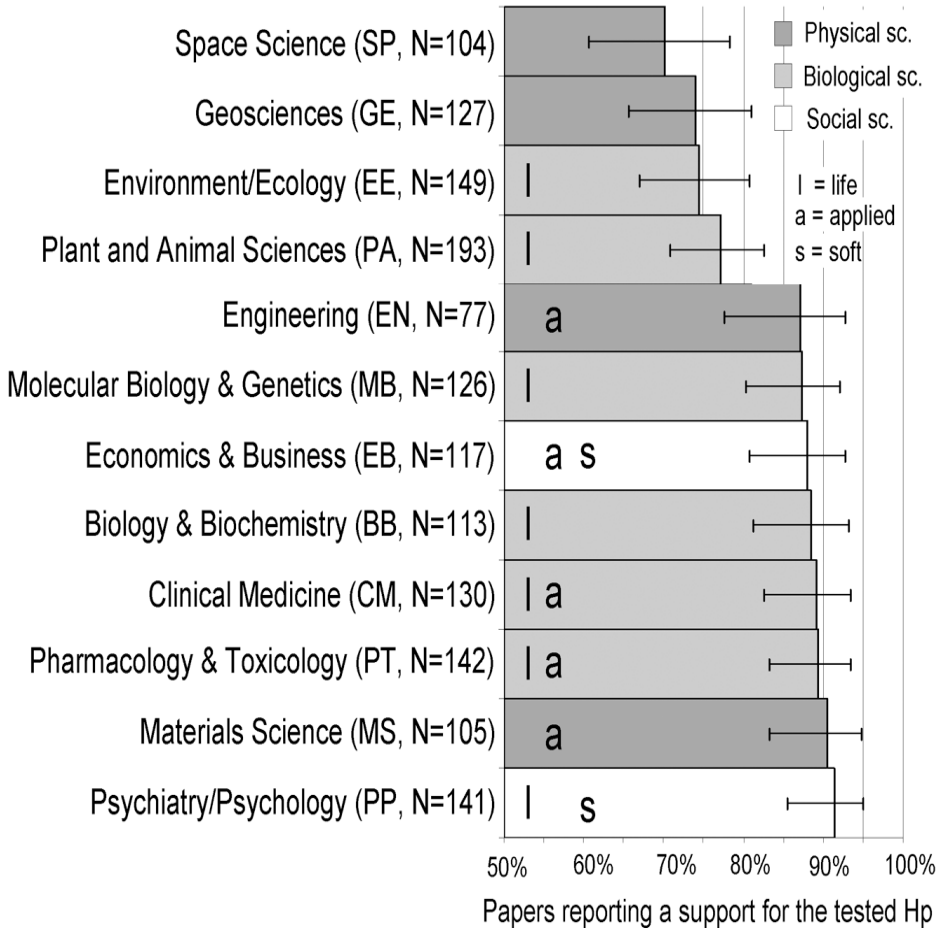
- Publication bias is omnipresent in science
- Publication bias $\rightarrow$ overestimation of effect size in meta-analysis
- The publication bias method *p*-uniform overestimates effect size in case of heterogeneity in true effect size
- The improved and extended method *p*-uniform*:
  1. eliminates overestimation due to heterogeneity
  2. uses information of significant *and* nonsignificant effect sizes
  3. enables estimating and testing of the extent of heterogeneity

Overview

1. Publication bias
2. From *p*-uniform to *p*-uniform*
3. Illustration based on two examples
4. Software
5. Conclusion and discussion

Publication bias

- Publication bias is "the selective publication of studies with a significant outcome"
- Close to 90% of main hypotheses are significant in economics
- But this is not in line with median statistical power (18%; Ioannidis, et al. 2017)



Adapted from Fanelli (2010)

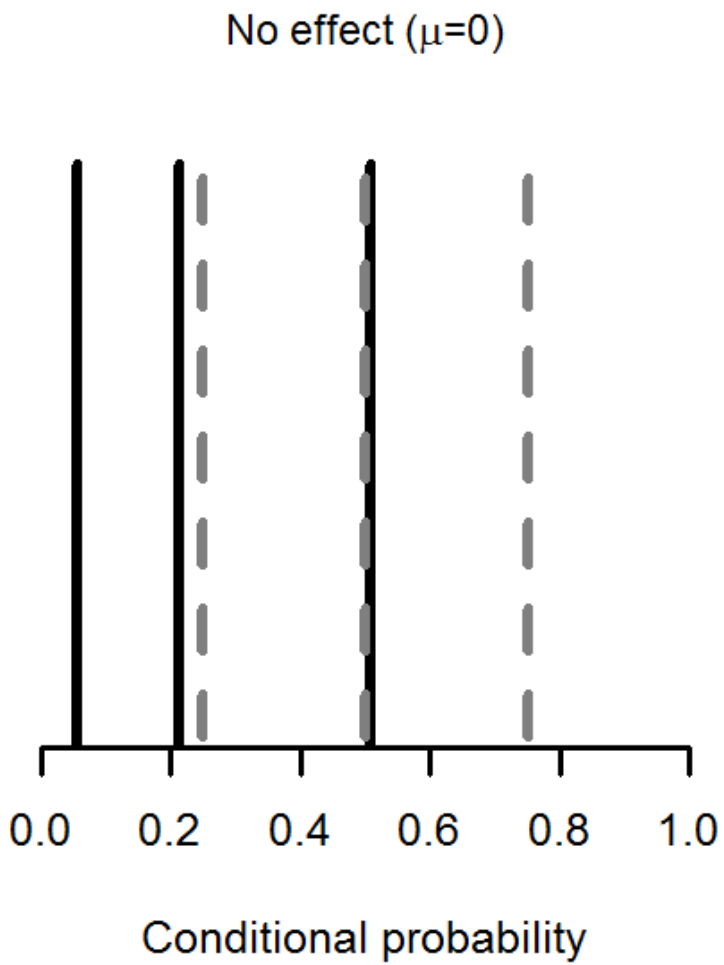
From *p*-uniform to *p*-uniform*: *p*-uniform

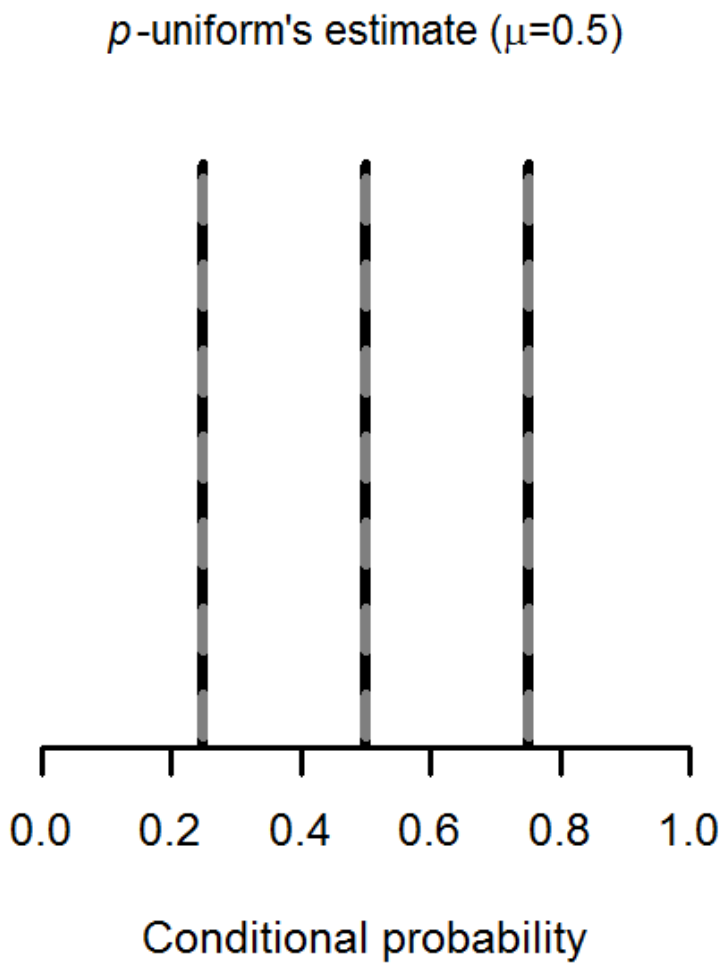
- Only considers significant effect sizes and discards others
- **Statistical principle:** Distribution of *p*-values *at the true effect size* is uniform
- Only significant effect sizes $\rightarrow$ *p*-values/probabilities conditional on significance are needed
- Important assumptions:
  - Homogeneous true effect size
  - All significant effect sizes have an equal probability of getting included in a meta-analysis

From *p*-uniform to *p*-uniform*: *p*-uniform

- Example with three observed effect sizes ($\mu = 0.5$):

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





From *p*-uniform to *p*-uniform*: *p*-uniform*

- Drawbacks of *p*-uniform:
  1. overestimation due to heterogeneity
  2. uses only information of significant effect sizes $\rightarrow$ suboptimal
  3. no estimating and testing of the extent of heterogeneity

From *p*-uniform to *p*-uniform*: *p*-uniform*

- Drawbacks of *p*-uniform:
  1. overestimation due to heterogeneity
  2. uses only information of significant effect sizes $\rightarrow$ suboptimal
  3. no estimating and testing of the extent of heterogeneity

- *P*-uniform* considers the significant **and** nonsignificant effect sizes

- Now effect sizes not only conditional on significance but also on nonsignificance

- Important assumption:
  - Probability of including a significant and nonsignificant effect size in a meta-analysis is assumed to be constant (but may differ from each other)

Illustration: Examples

- **Example 1: Alfieri et al. (2011)**
  - Meta-analysis studying whether discovery learning has greater learning benefits
  - 24 Hedges' g effect sizes
  - Sample sizes vary from 12 to 531

- **Example 2: McCall and Carriger (1993)**
  - Meta-analysis on the relationship between infants' habituation and recognition memory to a give stimulus and their later IQ
  - 12 Pearson correlation coefficients
  - Sample sizes vary from 11 to 96

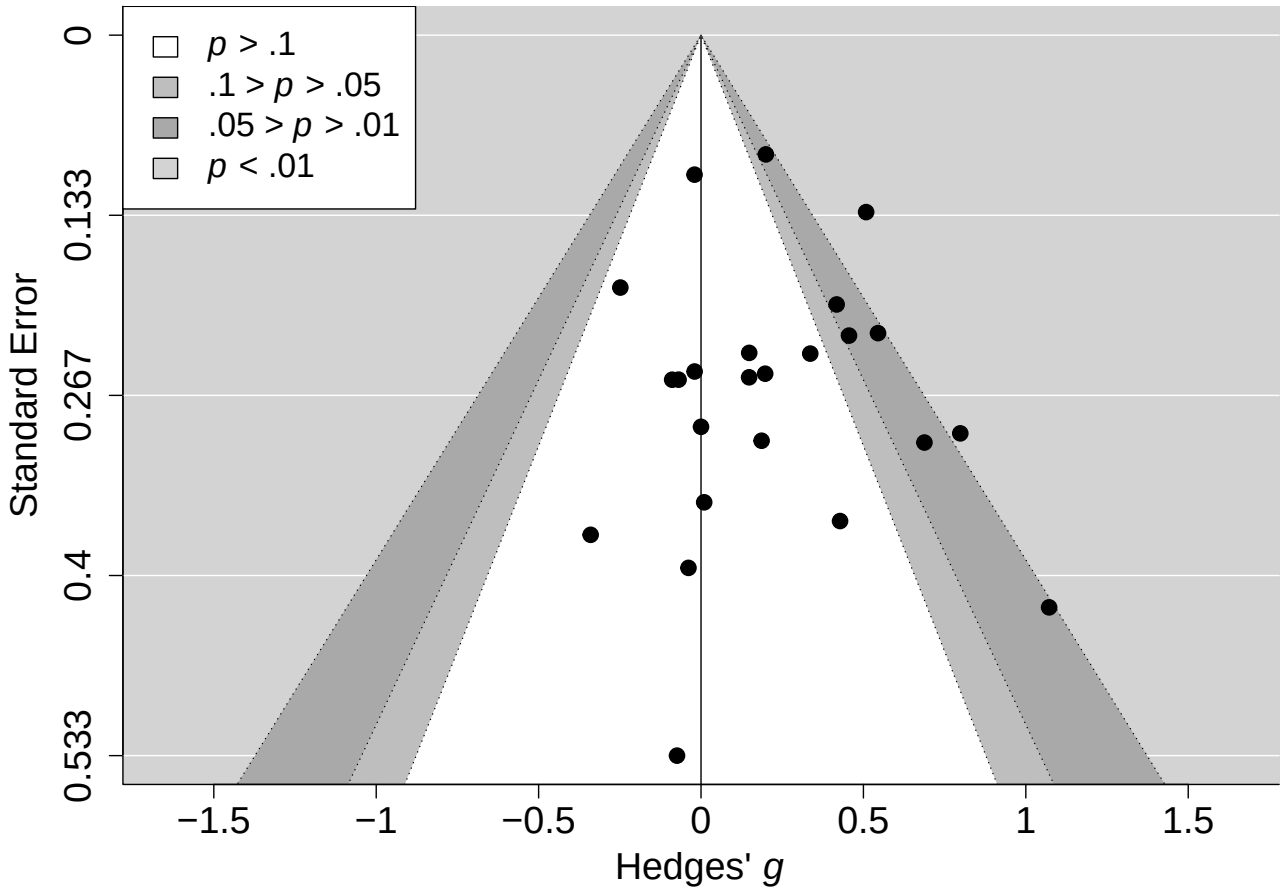
Illustration: Examples

- **Example 1: Alfieri et al. (2011)**
  - Meta-analysis studying whether discovery learning has greater learning benefits
  - 24 Hedges' g effect sizes
  - Sample sizes vary from 12 to 531

- **Example 2: McCall and Carriger (1993)**
  - Meta-analysis on the relationship between infants' habituation and recognition memory to a give stimulus and their later IQ
  - 12 Pearson correlation coefficients
  - Sample sizes vary from 11 to 96

- Included methods:
  - p-uniform*
  - random-effects meta-analysis
  - selection model approach by Hedges (1992)
  - PET and PEESE
  - WAAP

Illustration: Example 1



- No clear indication of small-study effects
- Egger's test for small-study effects not significant ($p = 0.799$)

Illustration: Example 1

	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	**0.213**	**0.032**	**(0.091; 0.334)**
p-uniform*			
Sel. model			
PET			
PEESE			
WAAP			

Illustration: Example 1

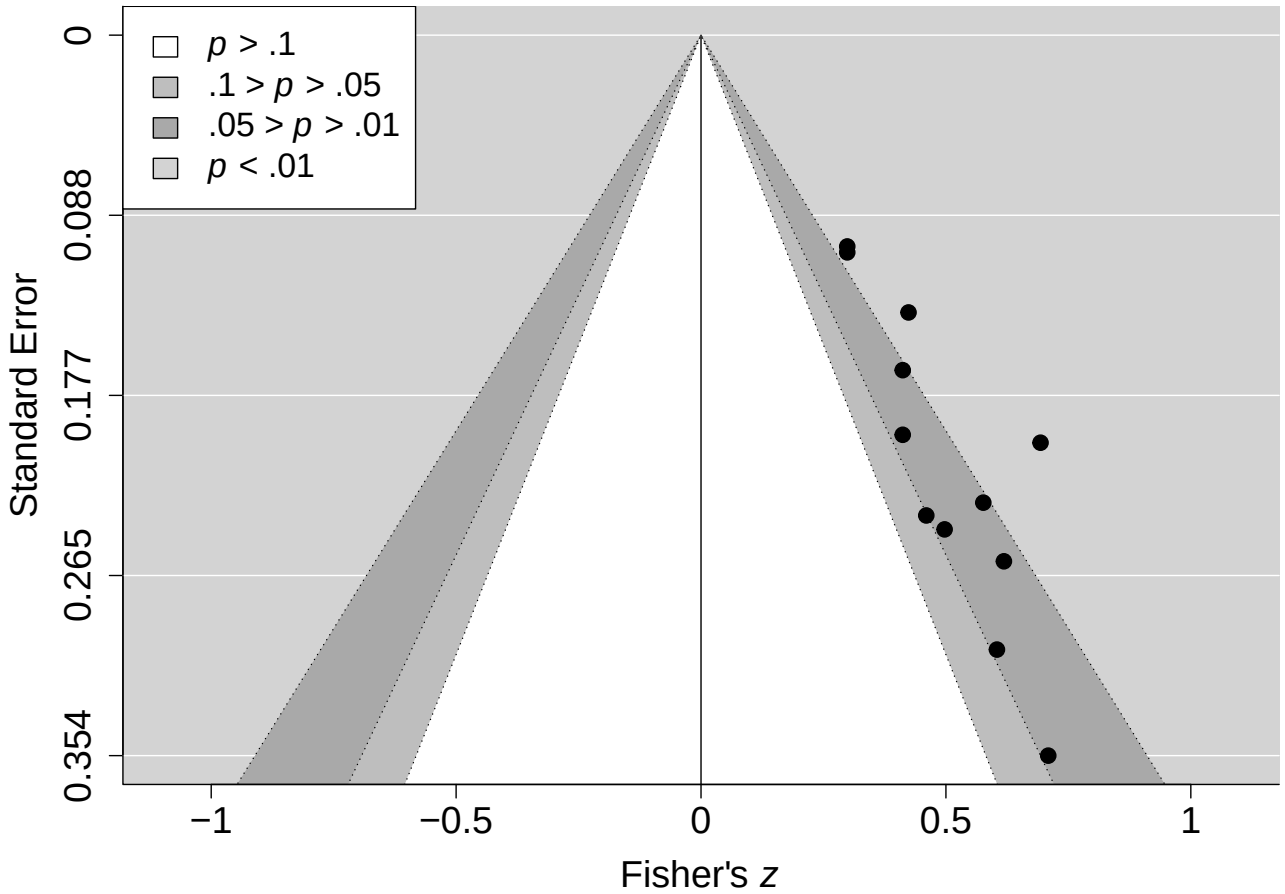
	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	0.213	0.032	(0.091; 0.334)
p*-uniform	**0.069**	**0.01**	**(-0.068; 0.21)**
Sel. model	**0.075**	**0.011**	**(-0.059; 0.209)**
PET			
PEESE			
WAAP			

Illustration: Example 1

	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	0.213	0.032	(0.091; 0.334)
p-uniform*	0.069	0.01	(-0.068; 0.21)
Sel. model	0.075	0.011	(-0.059; 0.209)
PET	**0.147**	-	**(-0.111; 0.404)**
PEESE	**0.178**	-	**(0.014; 0.342)**
WAAP	**0.508**	-	-

- Notes:
  - WAAP is here based on one effect size
  - Null-hypothesis of no effect *not* rejected in PET-analysis $\rightarrow$ interpret PET's estimate
  - Null-hypothesis of no effect rejected by RE and PEESE

Illustration: Example 2



- Small-study effects seem to be present
- Egger's test for small-study effects significant ($p = 0.025$)

Illustration: Example 2

	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	**0.39**	**0**	**(0.306; 0.469)**
p-uniform*			
Sel. model			
PET			
PEESE			
WAAP			

Illustration: Example 2

	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	0.39	0	(0.306; 0.469)
p*-uniform	**0.269**	**0**	**(0.097; 0.41)**
Sel. model	**0.263**	**0**	**(0.224; 0.301)**
PET			
PEESE			
WAAP			

Illustration: Example 2

	ES	$\hat{\tau}^2$	95% CI ES
RE (REML)	0.39	0	(0.306; 0.469)
p-uniform*	0.269	0	(0.097; 0.41)
Sel. model	0.263	0	(0.224; 0.301)
PET	**0.132**	-	**(0.009; 0.251)**
PEESE	**0.279**	-	**(0.203; 0.352)**
WAAP	**0.314**	-	**(0.125; 0.481)**

- Notes:
  - WAAP is here based on three effect sizes
  - Null-hypothesis of no effect rejected in PET-analysis $\rightarrow$ interpret PEESE's estimate
  - Null-hypothesis of no effect rejected by all methods

Software

- *p*-uniform*: R package `puniform` and web application https://rvanaert.shinyapps.io/p-uniformstar

Web application p-uniform*

Manual on how to use this application

Author: Robbie C.M. van Aert

Enter the characteristics of your meta-analysis below:

Select effect size measure

- ○ One-sample mean
- ◉ Two-independent means
- ○ One correlation

Alpha level in primary studies (default .05)

0.05

Select direction of effect in primary studies

- ◉ Right (positive)
- ○ Left (negative)

Select estimation method for p-uniform

- ◉ ML
- ○ P
- ○ LNP

Data entry

Select how you will enter data (see manual)

- ◉ Via CSV file
- ○ Manually in table

Enter data via CSV file

Browse... rabelo.csv

Upload complete

Analyze

p-uniform* (k = 25; ksig = 23)

Estimating effect size p-uniform*:

estimate	ci.lb	ci.ub	L.0	pval
0.0749	-0.1876	0.3067	0.3395	0.5601

Estimating between-study variance p-uniform*:

estimate	tau2.lb	tau2.ub	L.het	pval
0	0	0.0224	0	1

Publication bias test p-uniform*:

L.pb	pval
21.2298	<.001

Random-effects meta-analysis (tau^2 estimator PM):

Estimating effect size random-effects meta-analysis:

estimate	se	ci.lb	ci.ub	zval	pval
0.5706	0.0523	0.468	0.6731	10.9038	<.001

Estimating between-study variance random-effects meta-analysis:

estimate	se	tau2.lb	tau2.ub	Q	pval
0	0.0198	0	0	4.5523	1

Software

- *p*-uniform*: R package `puniform` and web application https://rvanaert.shinyapps.io/p-uniformstar
- Hedges' (1992) selection model approach: R package `weightr` and web application https://vevealab.shinyapps.io/WeightFunctionModel
- PET, PEESE, and WAAP: Any software that enables fitting weighted least squares regression models

Conclusion and discussion

- *P*-uniform* is an improvement over *p*-uniform, because
  1. eliminates overestimation due to heterogeneity
  2. is a more efficient estimator than *p*-uniform's estimator
  3. enables estimating and testing of the extent of heterogeneity

- Signs of overestimation in two examples, but methods disagreed on the extent

Conclusion and discussion

- *P*-uniform* is an improvement over *p*-uniform, because
  1. eliminates overestimation due to heterogeneity
  2. is a more efficient estimator than *p*-uniform's estimator
  3. enables estimating and testing of the extent of heterogeneity

- Signs of overestimation in two examples, but methods disagreed on the extent

- Recommendations:
  - Report results of publication bias methods in any meta-analysis
  - Include publication bias methods that are suitable for the characteristics of your meta-analysis

Conclusion and discussion

- Future research:
  - Violations of the assumption of equal probabilities of significant and nonsignificant effect sizes for getting published
  - Consequences of *p*-hacking
  - Simulation study including all publication bias methods that are now seen as state-of-the-art

Thank you for your attention

www.robbievanaert.com

www.metaresearch.nl

Preprint of paper about *p*-uniform*:
https://osf.io/preprints/bitss/zqjr9/