

Conducting meta-analyses based on p -values: Reservations and recommendations for applying p -uniform and p -curve

Robbie van Aert, Jelte Wicherts, and Marcel van Assen
R.C.M.vanAert@tilburguniversity.edu

Tilburg University
Faculty TSB
Department of Methodology and Statistics

The message

- Publication bias is widespread in psychology
- Overestimation in effect size (meta-analysis) in case of publication bias
- p -uniform and p -curve were developed for accurate effect size estimation
- p -uniform and p -curve do not always perform well
 - Heterogeneity
 - Sensitive to p -values close to .05
 - Unpredictable bias caused by p -hacking

Overview

1. Basic idea underlying p -uniform and p -curve
2. Three reservations
3. Software for applying p -uniform and p -curve
4. Conclusion and discussion

1. Basic idea

- Publication bias is ‘the selective publication of studies with a significant outcome’
- Distribution of p -values conditional on the true effect size is uniform
- Effect size estimate is obtained when (statistically significant) conditional p -values are uniformly distributed
- Assumptions:
 - Significant effect sizes have equal probability of getting published
 - Effect sizes are statistically independent

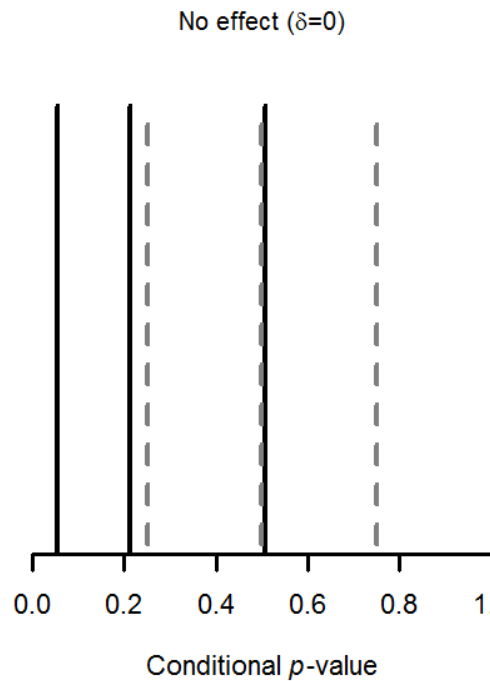
1. Basic idea

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

$$t(48)=3.133, p=.0029$$

$$t(48)=2.302, p=.011$$

$$t(48)=2.646, p=.025$$



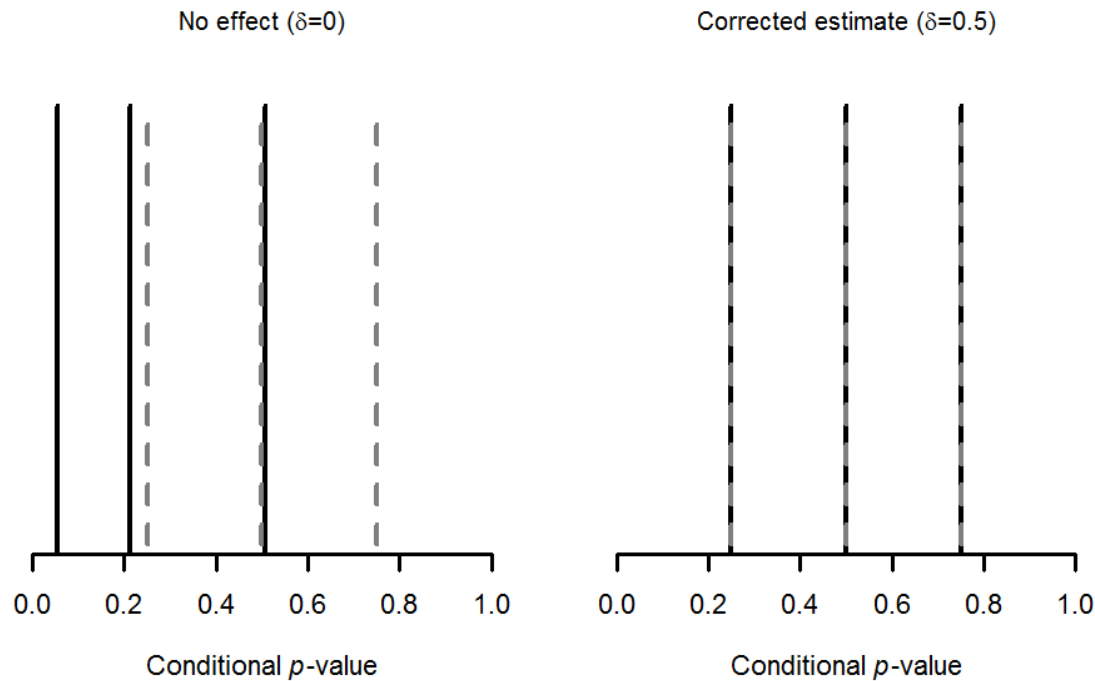
1. Basic idea

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

$$t(48)=3.133, p=.0029$$

$$t(48)=2.302, p=.011$$

$$t(48)=2.646, p=.025$$



2. Three reservations

- p -uniform and p -curve have great promise estimating effect size
- However, we will add three reservations to the current methodology:
 - a. Overestimation caused by moderate to large heterogeneity
 - b. Sensitivity to p -values close to .05
 - c. Unpredictable bias in effect size estimates caused by p -hacking

2a. Heterogeneity

- Heterogeneity means that different true effect sizes are underlying the observed effect sizes in a meta-analysis
- Simonsohn et al. (2014) state that p -curve (and p -uniform) yield an accurate estimate
- Simulation study with two-independent groups design and $\delta=0.397$

2a. 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

2b. Sensitive to p -values close to α

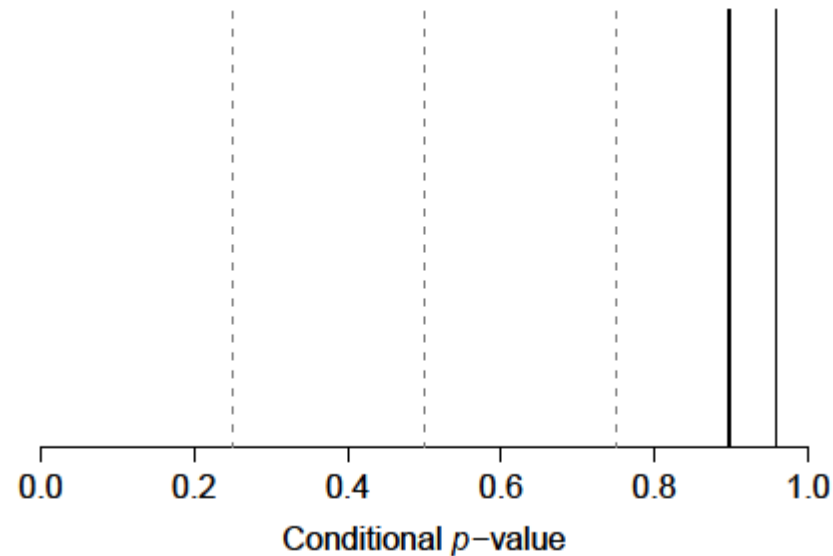
- Many p -values close to .05 are in line with a *negative* true effect size

- Example with three observed effect sizes:

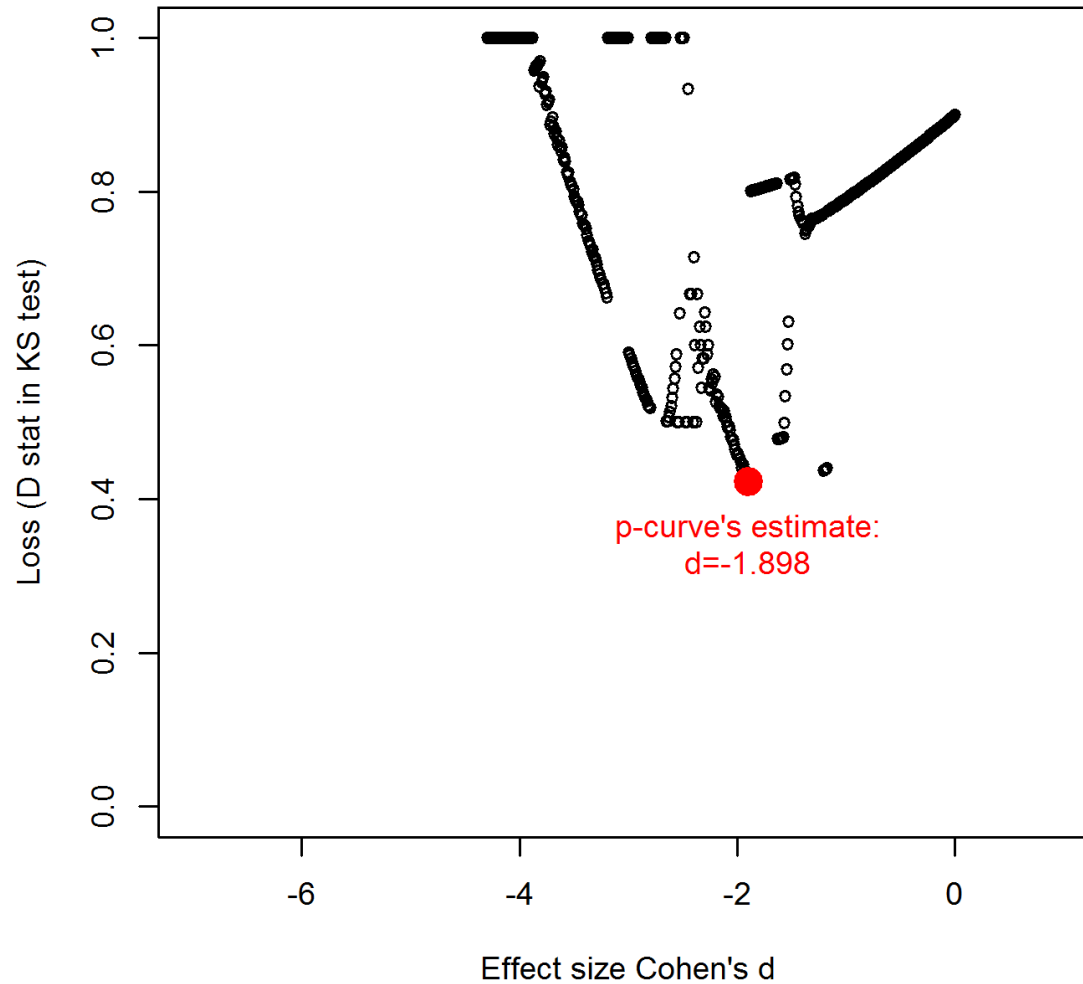
$$t(18)=2.154, p=.045$$

$$t(48)=2.058, p=.045$$

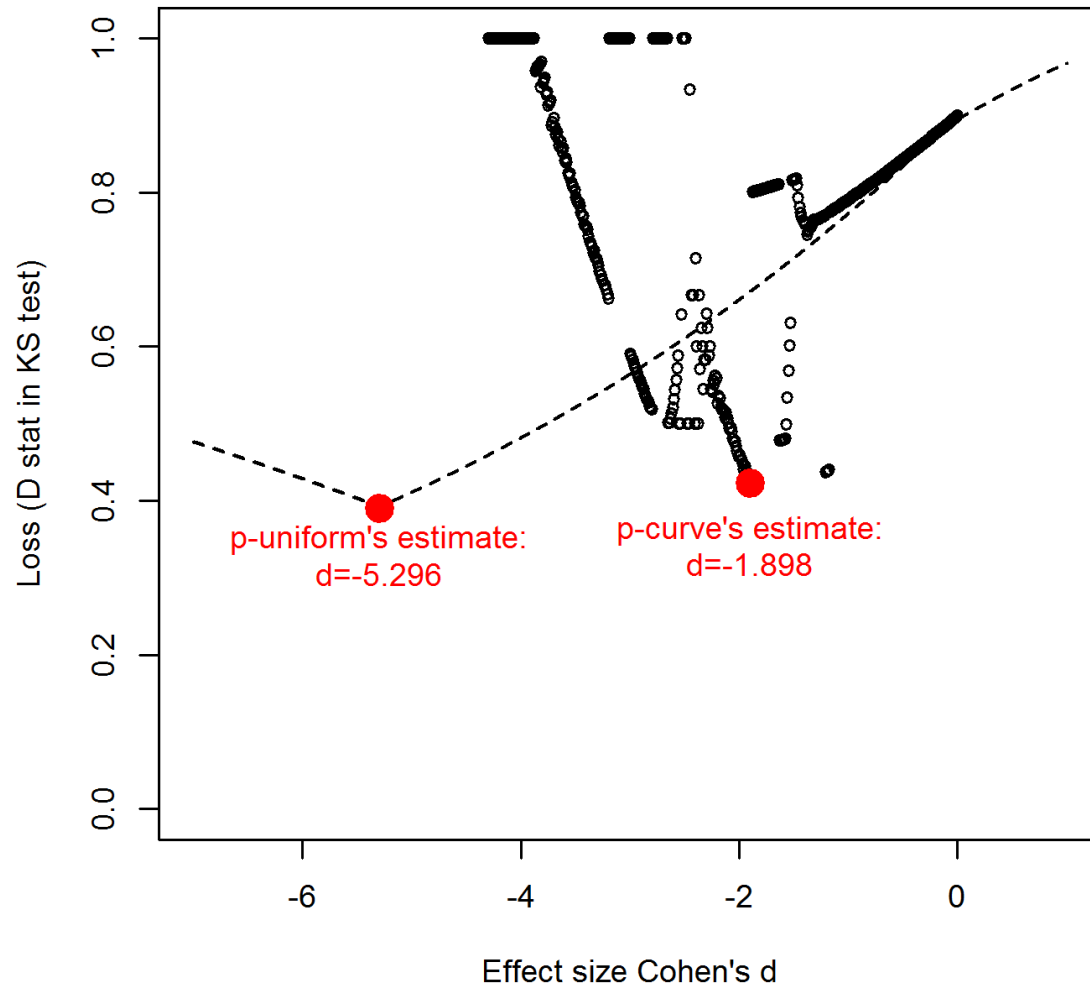
$$t(98)=2.002, p=.048$$



2b. Sensitive to p -values close to α



2b. Sensitive to p -values close to α



2b. Sensitive to p -values close to α

- p -uniform's and p -curve's estimate is negative if the average significant p -value is larger than $\alpha/2$ ($\alpha/4$ for two-tailed test)

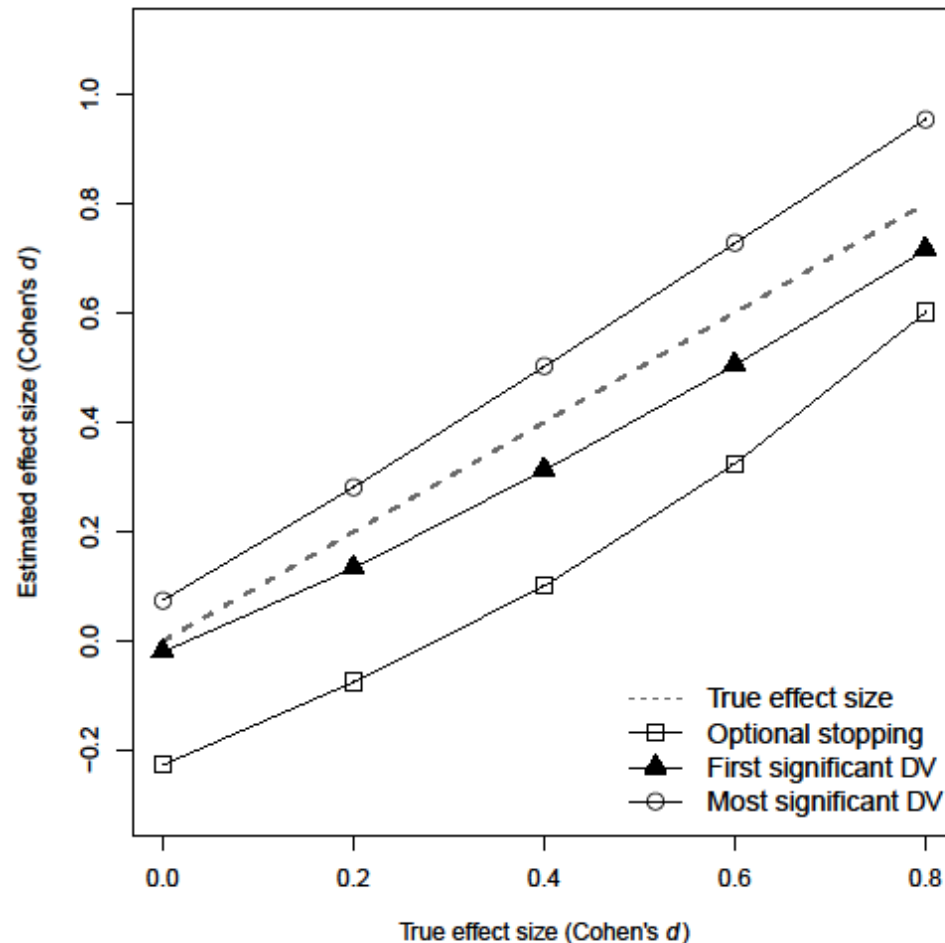
- Recommendation:
 - Set estimate to zero if estimate will be negative

- In line with one-tailed hypothesis testing with the observed effect size in the opposite direction

2c. Bias caused by p -hacking

- p -hacking is a term for all behaviors that researchers can use to obtain desirable results
- Simonsohn et al. (2014) concluded that p -curve (and p -uniform) underestimate effect sizes in case of p -hacking
- This holds if p -hacking results in p -values just below .05 → not always the case
- Simulation study with p -hacking:
 - Optional stopping
 - Only reporting the first significant dependent variable
 - Only reporting the most significant dependent variable

2c. Bias caused by p -hacking



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

3. Software

- R package “puniform” on GitHub and R code for p -curve available
- Web application for p -uniform:
<https://rvanaert.shinyapps.io/p-uniform/>

Web application p-uniform

[Manual on how to use this application](#)

[Paper about p-uniform](#)

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)

Select direction of effect in primary studies

- Right (positive)
- Left (negative)

Select estimation method for p-uniform

- P (Irwin-Hall)
- LNP
- LN1MINP
- KS
- AD

Data entry

Select how you will enter data (see manual)

- Via CSV file
- Manually in table

Enter data via CSV file

data.mccall93.csv

Upload complete

Analyze

Effect size estimate p-uniform:

estimate	ci.lb	ci.ub	L.0	pval	ksig
0.1792	-0.2379	0.3545	-1.1814	0.1187	11

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)

Select direction of effect in primary studies

- Right (positive)
- Left (negative)

Select estimation method for p-uniform

- P (Irwin-Hall)
- LNP
- LN1MINP
- KS
- AD

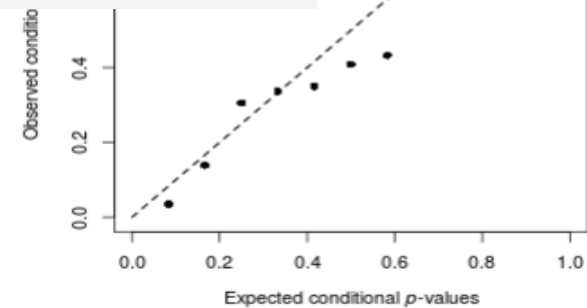
normal distribution

form:

normal distribution

al	ci.lb	ci.ub	pval	Qstat	Qpval
0.05	0.3156	0.509	<.001	6.7409	0.8197

transformed p-values



Web application p-uniform

[Manual on how to use this application](#)

[Paper about p-uniform](#)

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)

Select direction of effect in primary studies

- Right (positive)
- Left (negative)

Select estimation method for p-uniform

- P (Irwin-Hall)
- LNP
- LN1MINP
- KS
- AD

Data entry

Select how you will enter data (see manual)

- Via CSV file
- Manually in table

Enter data via CSV file

 data.mccall93.csv

Upload complete

Data entry

Select how you will enter data (see manual)

- Via CSV file
- Manually in table

Enter data via CSV file

 data.mccall93.csv

Upload complete

Effect size estimate p-uniform:

estimate	ci.lb	ci.ub	L.0	pval	ksig
0.1792	-0.2379	0.3545	-1.1814	0.1187	11

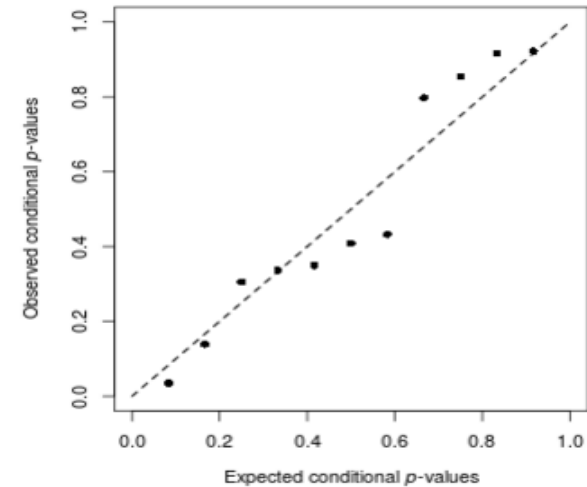
Notes:

- p-value approximated with normal distribution

Publication bias test p-uniform:

L.pb	pval
------	------

.509	<.001	6.7409	0.8197
------	-------	--------	--------



Web application p-uniform

[Manual on how to use this application](#)

[Paper about p-uniform](#)

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)

Select direction of effect in primary studies

- Right (positive)
- Left (negative)

Select estimation method for p-uniform

- P (Irwin-Hall)
- LNP
- LN1MINP
- KS
- AD

Data entry

Select how you will enter data (see manual)

- Via CSV file
- Manually in table

Enter data via CSV file

data.mccall93.csv

Upload complete

Analyze

Effect size estimate p-uniform:

estimate	ci.lb	ci.ub	L.0	pval	ksig
0.1792	-0.2379	0.3545	-1.1814	0.1187	11

Notes:

- p-value approximated with normal distribution

Publication bias test p-uniform:

.....

Effect size estimate p-uniform:

estimate	ci.lb	ci.ub	L.0	pval	ksig
0.1792	-0.2379	0.3545	-1.1814	0.1187	11

Notes:

- p-value approximated with normal distribution

Publication bias test p-uniform:

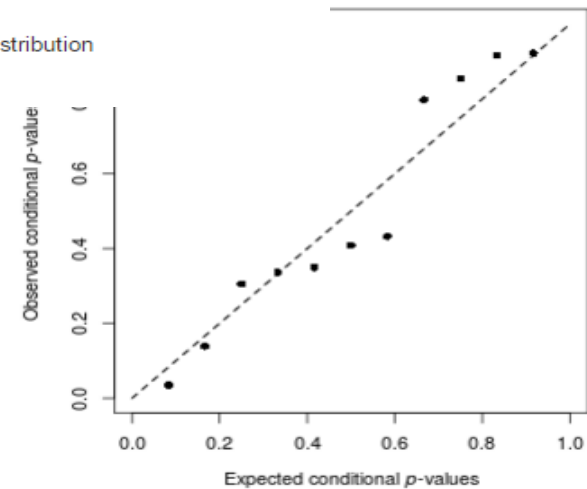
L.pb	pval
2.6154	0.0045

p-value approximated with normal distribution

mal distribution

I	ci.lb	ci.ub	pval	Qstat	Qpval
31	0.3156	0.509	<.001	6.7409	0.8197

Sorted p-values



4. Conclusion and discussion

- p -uniform and p -curve are promising methods to examine publication bias
- However, both p -uniform and p -curve:
 - Overestimate effect size in case of medium or large heterogeneity
 - Sensitive to p -values close to .05
 - May yield unpredictable bias if p -hacking is present
- Future research:
 - Extend p -uniform such that it can deal with heterogeneous true effect sizes
 - Extend p -uniform to allow estimation of other effect size measures (e.g., odds ratios)

! Thank you for your attention !

