Dealing with publication bias in a meta-analysis

Robbie C.M. van Aert

Tilburg University, Meta-Reserach Center

SIPS July 9, 2019, Rotterdam

Publication bias

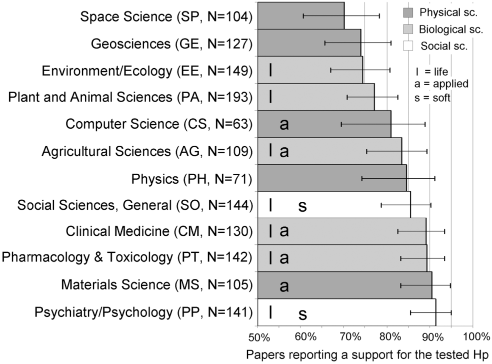
- Publication bias is "the selective publication of studies with a statistically significant outcome"

- Consequences of publication bias:
  - Type-I errors $\rightarrow$ false impression that an effect exists
  - Overestimation of effect size
  - Questionable research practices/*p*-hacking

- Meta-analysis actually enables us to assess publication bias by using meta-information

Publication bias

- Overwhelming evidence for bias in the existing literature
- ≈90% of main hypotheses are significant in psychology
- But this is not in line with average statistical power (about 20-50%)



Adapted from Fanelli (2010)

Publication bias methods

Question: How would you study publication bias in the ideal world where all published and unpublished studies are available?

Publication bias methods

Question: How would you study publication bias in the ideal world where all published and unpublished studies are available?

- 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
  - PET-PEESE
  - Selection model approaches
  - *p*-uniform and *p*-curve

Example

- Example meta-analysis by Rabelo et al. (2015) on the effect of weight on judgments of importance
- **Theory:** the physical experience of weight influences how much importance people assign to things, issues, and people
- Meta-analysis contains $k = 25$ standardized mean differences (i.e., Hedges' g)

Example

- Data are from Table 4 of Rabelo et al. (2015):

```
##                             study   m1i   m2i n1i n2i  sd1i  sd2i   yi   vi
## 1 Ackerman et al. (2010), Exp. 1  5.80  5.38  26  28  0.76  0.79 0.54 0.08
## 2 Ackerman et al. (2010), Exp. 2  4.01  3.25  21  22  0.73  1.73 0.56 0.10
## 3 Chandler et al. (2012), Exp. 2  7.26  6.30  30  30  1.58  1.33 0.65 0.07
## 4 Chandler et al. (2012), Exp. 1  0.00 -0.42  50  50  1.00  1.00 0.42 0.04
## 5 Chandler et al. (2012), Exp. 3  6.97  6.09  50  50  2.03  1.63 0.47 0.04
## 6          Hafner (2013), Exp. 1 91.94 78.45  30  30 25.86 25.30 0.52 0.07
```

- A positive standardized mean difference ($y_i > 0$) indicates that people assigned more importance to judgments if they held a heavy object

- R is used for applying all publication bias methods

- Packages `metafor`, `puniform`, and `weightr` will be used

Example: Random-effects model

```r
rma(yi = yi, vi = vi, data = dat) # Random-effects meta-analysis
```

```
##
## Random-Effects Model (k = 25; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.019)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  1.00
##
## Test for Heterogeneity:
## Q(df = 24) = 4.695, p-val = 1.000
##
## Model Results:
##
## estimate     se    zval   pval  ci.lb  ci.ub
##    0.569  0.052  10.893  <.001  0.467  0.672  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Interpretation:** Average effect of medium size ($\hat{\mu} = 0.569$)

Assessing publication bias

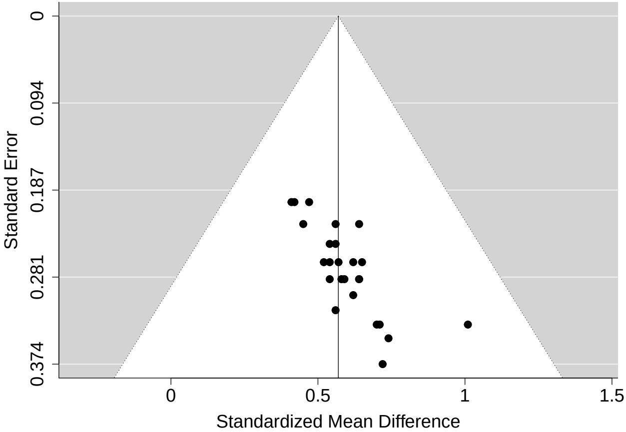
Failsafe *N*

- Unpublished studies are hidden in the *file drawers* of researchers
- Computes the number of effect sizes equal to zero that are needed to make the meta-analytic effect size nonsignificant
- 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

Funnel plot

- Shows relationship between effect size and its precision

```r
res <- rma(yi = yi, vi = vi, data = dat) # Random-effects meta-analysis
funnel(res) # Create funnel plot
```

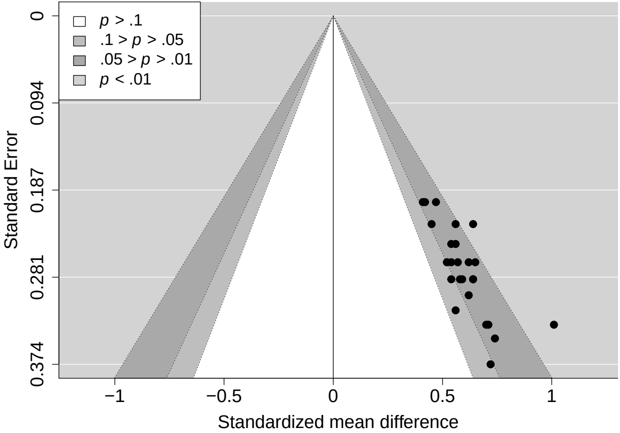


Funnel plot

- An asymmetric funnel is often interpreted as evidence of publication bias
- However, it actually suggests the presence of so-called *small-study effects*
- Causes of small-study effects (see Sterne et al., 2000):
  - Publication bias
  - Heterogeneity in true effect size
  - Different designs in small vs. large studies
  - Power analysis to determine the required sample size in combination with heterogeneity
  - Chance
  - Etc.

Contour-enhanced funnel plot

- Eyeballing a funnel plot for asymmetry is difficult → contour-enhanced funnel plot



Funnel plot asymmetry tests: Rank-correlation test

- Funnel plot asymmetry tests test for small-study effects
- Rank-correlation test computes a rank-order correlation between the effect sizes and their precision (Begg & Mazumdar, 1994)
- Applied to example:

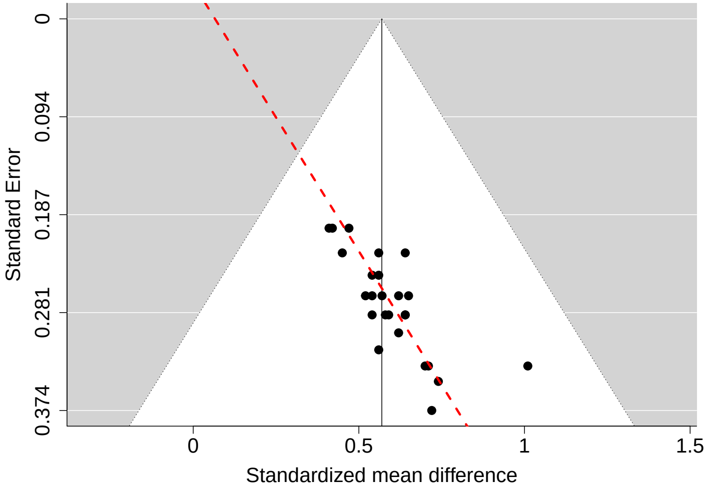
```r
ranktest(res) # Apply rank-correlation test
```

```
##
## Rank Correlation Test for Funnel Plot Asymmetry
##
## Kendall's tau = 0.6124, p < .0001
```

- **Interpretation:** null-hypothesis of no small-study effects is rejected

Funnel plot asymmetry tests: Egger's test

- Egger's test fits a regression line through the points in a funnel plot



Funnel plot asymmetry tests: Egger's test

- Vertical line suggests a symmetric funnel
- If slope is significantly different from zero → funnel plot asymmetry
- Applied to second-hand smoke example:

```r
regtest(res) # Apply Egger's test
# rma(yi = yi, vi = vi, mods = ~ sqrt(vi), data = dat) # Is equivalent
```

```
##
## Regression Test for Funnel Plot Asymmetry
##
## model:     mixed-effects meta-regression model
## predictor: standard error
##
## test for funnel plot asymmetry: z = 1.6757, p = 0.0938
```

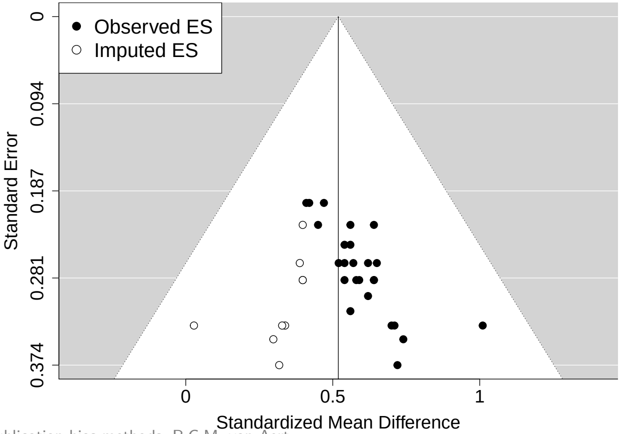
Funnel plot asymmetry tests

- Drawbacks of these tests:
  - Low statistical power in case of less than 10 effect sizes
  - Test for absence of small-study effects instead of publication bias
- Low power, so why not correcting estimates for publication bias?!
- Corrected estimates are probably also more interesting for applied researchers

Correcting for publication bias

Trim-and-fill method

- Intuitive method to correct effect size estimate
- Missing effect sizes from one side of funnel plot are *trimmed* and *filled* in other side



Trim-and-fill method

- Applied to example ($\hat{\mu} = 0.569$):

```r
trimfill(res) # Apply trim-and-fill method
```

```
##
## Estimated number of missing studies on the left side: 9 (SE = 3.1336)
##
## Random-Effects Model (k = 34; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0169)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  1.00
##
## Test for Heterogeneity:
## Q(df = 33) = 10.0480, p-val = 1.0000
##
## Model Results:
##
## estimate      se     zval    pval   ci.lb   ci.ub
##   0.5189  0.0462  11.2278  <.0001  0.4283  0.6095  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

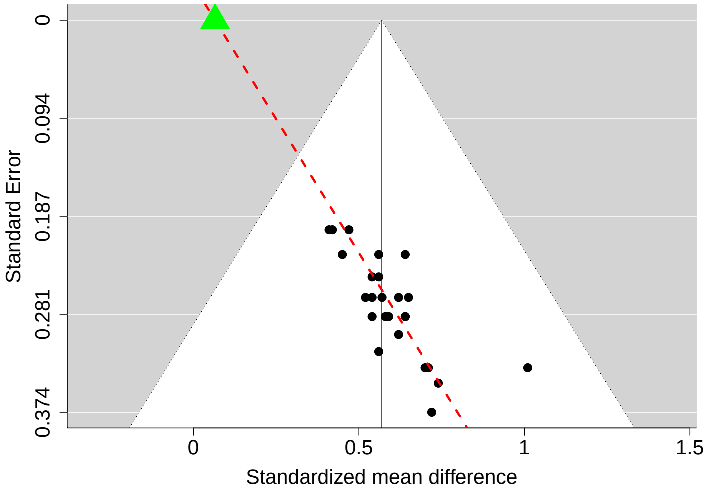
Trim-and-fill method

- Most often used method but nowadays discouraged to be used

- Drawbacks:
  - Method based on funnel plot, so it corrects for small-study effects rather than publication bias
  - Method should not be applied when true effect size is heterogeneous
  - Simulation studies revealed that the method imputes studies when none are missing (e.g., Terrin et al., 2003)

PET-PEESE

- Estimate equals the effect size where standard error is zero (infinite sample size)



PET-PEESE

- Conditional estimator based on two meta-regression analyses:
  - PET $\rightarrow$ *standard error* as moderator
  - PEESE $\rightarrow$ *sampling variance* as moderator

- Selection of PET/PEESE depends on whether $H_0 : \mu = 0$ is rejected in PET-analysis

- In R with the `metafor` package:

```
rma(yi = yi, vi = vi, mods = ~ sqrt(vi), data = dat) # PET
rma(yi = yi, vi = vi, mods = ~ vi, data = dat) # PEESE
```

PET-PEESE

- Conditional estimator based on two meta-regression analyses:
  - PET $\rightarrow$ *standard error* as moderator
  - PEESE $\rightarrow$ *sampling variance* as moderator

- Selection of PET/PEESE depends on whether $H_0 : \mu = 0$ is rejected in PET-analysis

- In R with the `metafor` package:

```r
rma(yi = yi, vi = vi, mods = ~ sqrt(vi), data = dat) # PET
rma(yi = yi, vi = vi, mods = ~ vi, data = dat) # PEESE
```

- $H_0 : \mu = 0$ is not rejected in the second-hand smoke example ($z = 0.215$, $p = .8298$), so PET:

```
##            estimate      se    zval    pval    ci.lb   ci.ub
## intrcpt      0.0656  0.3051  0.2150  0.8298  -0.5324  0.6636
## sqrt(vi)     1.9566  1.1676  1.6757  0.0938  -0.3319  4.2450  .
```

Selection model approaches

- Generic term for methods combining effect size model with selection model
- *Effect size model:* distribution of effect sizes in the absence of publication bias
- *Selection model:* mechanism by which effect size estimates are selected to be observed
- Selection model is often based on *p*-values
- Drawbacks:
  - Complicated
  - Require substantial number of studies in a meta-analysis

Selection model approaches: Vevea and Hedges (1995)

- Selection model approach implemented in `weightr` package

- Applied to example with:
  - Effect size model → random-effects model
  - Selection model → a single cut-off at 0.025

```r
install.packages("weightr") # Install "weightr" package
library(weightr) # Load weightr package
weightfunct(effect = dat$yi, v = dat$vi) # Apply sel. model approach
```

```
##
##  Adjusted Model (k = 25):
##
##  tau^2 (estimated amount of total heterogeneity): 0.0000 (SE =   NaN)
##  tau (square root of estimated tau^2 value):  0.0000
##
##  Model Results:
##
##               estimate std.error z-stat   p-val     ci.lb   ci.ub
##  Intercept     0.18065       NaN    NaN      NA       NaN     NaN
##  0.025 < p < 1  0.01488   0.01268  1.173 0.24073 -0.009981 0.03974
```

P-uniform and *p*-curve

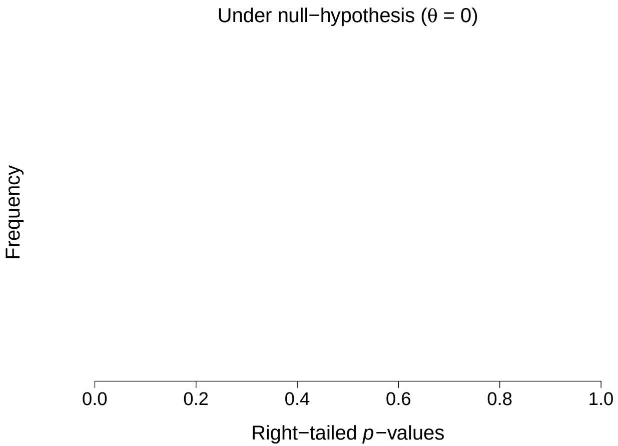
- [Robbie adds disclaimer]
- *P*-uniform and *p*-curve can be seen as a selection model approach
- Both methods are based on same methodology but slightly differ in implementation

P-uniform and *p*-curve

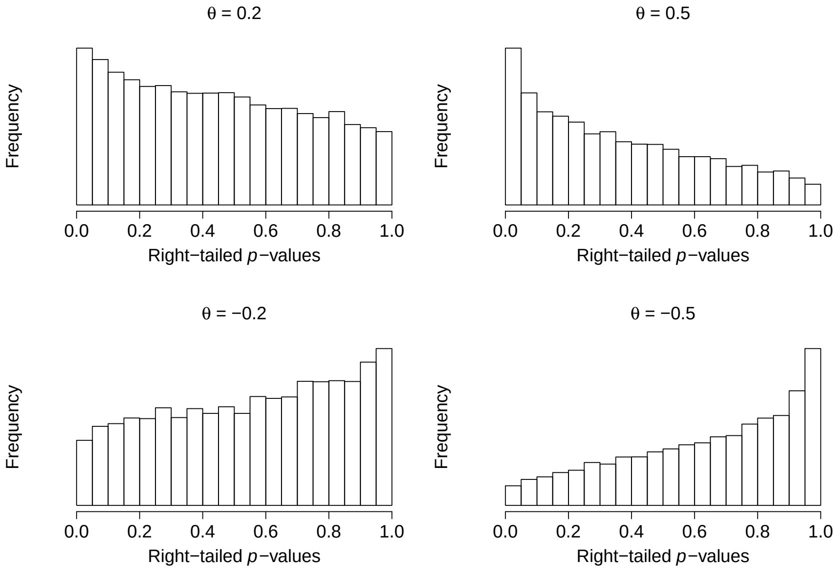
- [Robbie adds disclaimer]
- *P*-uniform and *p*-curve can be seen as a selection model approach
- Both methods are based on same methodology but slightly differ in implementation
- Web application for users who are unfamiliar with R https://rvanaert.shinyapps.io/p-uniform/

P-uniform and *p*-curve

- **Main idea:** *p*-values are uniformly distributed under the null-hypothesis



P-uniform and p-curve

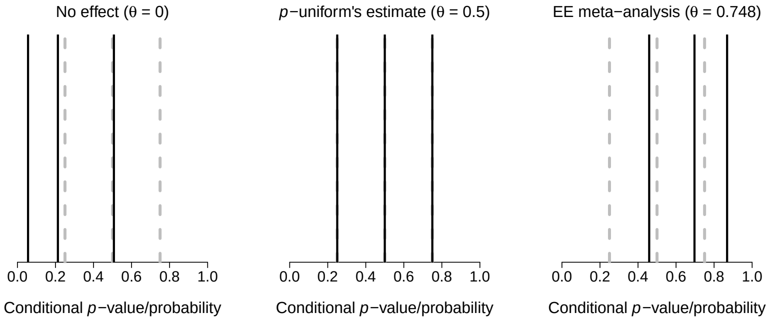


P-uniform and *p*-curve

- *P*-values are not only uniformly distributed under the null-hypothesis *but also at the true effect size*
- Methods discard nonsignificant effect sizes $\rightarrow$ conditional *p*-values/probabilities
- Methods use probability of observing an effect size conditional on effect size being statistically significant
- Conditional *p*-values/probabilities are *p*-values conditional on these being significant

P-uniform (and p-curve)

- Example with three observed effect sizes ($\theta = 0.5$)
  - $t(48) = 3.133$, two-tailed $p = .0029$
  - $t(48) = 2.302$, two-tailed $p = .011$
  - $t(48) = 2.646$, two-tailed $p = .025$



P-uniform

- Applied to example ($\hat{\mu} = 0.569$):

```r
install.packages("puniform") # Install "puniform" package
library(puniform) # Load "puniform" package
puniform(yi = dat$yi, vi = dat$vi, side = "right")
```

```
## Method: P
##
## Effect size estimation p-uniform
##
##        est     ci.lb    ci.ub     L.0     pval    ksig
##    -0.2133   -0.7203   0.1377   1.1173   0.8681     22
##
## ===
##
## Publication bias test p-uniform
##
##      L.pb    pval
##     5.075   <.001
```

P-uniform*

- Drawbacks of *p*-uniform:
  - Overestimation in case of heterogeneity in true effect size
  - Not all available information is used (i.e., not efficient method)

- *P*-uniform* is an improvement over *p*-uniform because:
  1. It enables estimating and testing of heterogeneity in true effect size (τ^2)
  2. Takes into account significant and nonsignificant effect sizes

P-uniform*

- Drawbacks of *p*-uniform:
  - Overestimation in case of heterogeneity in true effect size
  - Not all available information is used (i.e., not efficient method)

- *P*-uniform* is an improvement over *p*-uniform because:
  1. It enables estimating and testing of heterogeneity in true effect size (τ^2)
  2. Takes into account significant and nonsignificant effect sizes

- A function to apply *p*-uniform* is in the `puniform` package:

```
puni_star(yi = dat$yi, vi = dat$vi, side = "right") # Apply p-uniform*
```

- Web application for users who are unfamiliar with R
  https://rvanaert.shinyapps.io/p-uniformstar/

P-uniform*

```
##
## Method: ML (k = 25; ksig = 22)
##
## Estimating effect size p-uniform*
##
##        est     ci.lb     ci.ub       L.0      pval
##     0.1487   -0.0842      0.36    1.6129    0.2041
##
## ===
##
## Estimating between-study variance p-uniform*
##
##       tau2   tau2.lb   tau2.ub     L.het      pval
##          0         0    0.0191         0         1
##
## ===
##
## Publication bias test p-uniform*
##
##       L.pb    pval
##    17.3923   <.001
```

Hands-on part

- Open file “practical_pubbias.pdf” on the OSF page (link)
- “practical_pubbias.pdf” helps you with step-by-step applying the methods and contains some questions that you can answer
- You can also follow the steps and answer the questions using your own data
- Please ask questions if something is unclear!

Recommendations and take-home messages

- Recommendations:
  1. Try to also include unpublished primary studies
  2. Judge which methods have good properties when applied to your data
  3. Triangulation → apply and report multiple publication bias methods if these pass 2.

Recommendations and take-home messages

- Recommendations:
  1. Try to also include unpublished primary studies
  2. Judge which methods have good properties when applied to your data
  3. Triangulation $\rightarrow$ apply and report multiple publication bias methods if these pass 2.

- Take-home messages
  1. Publication bias is a threat to the validity of meta-analyses
  2. Publication bias is **not** the only cause of small-study effects
  3. Do **not** use failsafe N and trim-and-fill method

Further reading on publication bias methods

- A book on publication bias methods:

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), Publication bias in meta-analysis: Prevention, assessment and adjustments. Chichester, UK: Wiley.

- Funnel plot asymmetry tests:

Sterne, J. A. C., Harbord, R. M., Sutton, A. J., Jones, D. R., Ioannidis, J. P., Terrin, N., . . . Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, 343(7818), 1-8. doi:10.1136/bmj.d4002

- Trim-and-fill method:

Duval, S., & Tweedie, R. L. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98. doi:10.1080/01621459.2000.10473905

- PET-PEESE:

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5(1), 60-78.

Further reading on publication bias methods

- Selection model approach:

Coburn, K. M., & Vevea, J. L. (2015). Publication bias as a function of study characteristics. *Psychological Methods*, 20(3), 310-330. doi:10.1037/met0000046

- *P*-uniform and *p*-curve:

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, 11(5), 713-729. doi:10.1177/1745691616650874

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(6), 666-681. doi:10.1177/1745691614553988

- *P*-uniform*

van Aert, R. C. M., & van Assen, M. A. L. M. (2019). Correcting for publication bias in a meta-analysis with the p-uniform* method. Manuscript submitted for publication. Retrieved from: https://osf.io/preprints/bitss/zqjr9.

Thank you for your attention

www.robbievanaert.com

www.metaresearch.nl