## Selection models for publication bias in meta-analysis

Methodology and Statistics Colloquium

Tilburg University

November 3, 2020

Wolfgang Viechtbauer

Maastricht University

2020-11-03



## Meta-Analysis

- collect studies that have examined some phenomenon of interest
- quantify the outcome/effect of each study in terms of an effect size / outcome measure (e.g., standardized mean difference, correlation coefficient, log odds ratio)
- let $y_i$ denote the outcome observed in the $i$th study ($i = 1, \ldots, k$) and $v_i$ the corresponding sampling variance
- random-effects model:

$$y_i = \theta_i + e_i$$

where $\theta_i \sim N(\mu, \tau^2)$ and $e_i \sim N(0, v_i)$
- special case: assume $\tau^2 = 0$, then $\mu \equiv \theta$



## Maximum Likelihood Estimation

- log likelihood:

$$ll(\mu, \tau^2) = -\frac{1}{2}\sum_{i=1}^{k} \ln(\tau^2 + v_i) - \frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - \mu)^2}{\tau^2 + v_i}$$

- can easily obtain MLEs of $\mu$ and $\tau^2$ using (Quasi-)Newton or Nelder-Mead type algorithms
- SE$[\hat{\mu}]$ and SE$[\hat{\tau}^2]$ via Fisher information or inverse Hessian matrix
- profile likelihood methods can also be used (esp. for $\tau^2$ CI)



## Publication Bias in Meta-Analysis

- want to synthesize all of the studies conducted on the phenomenon of interest (that fit certain inclusion criteria)
- finding all studies is difficult (esp. the ‘gray literature’)
- then at least want to obtain a representative sample thereof
- the studies we find (mostly in the published literature) have undergone some selection process that might make them unrepresentative
- if selection is a function of $y_i$, its direction, and/or statistical significance, will get biased estimates of $\mu$ and $\tau^2$



## Sterling (1959) and Smith (1980)

PUBLICATION DECISIONS AND THEIR POSSIBLE EFFECTS ON INFERENCES DRAWN FROM TESTS OF SIGNIFICANCE —OR VICE VERSA*

THEODORE D. STERLING

*University of Cincinnati*

There is some evidence that in fields where statistical tests of significance are commonly used, research which yields nonsignificant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs—an "error of the first kind"—and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

**PUBLICATION BIAS AND META-ANALYSIS**

**Mary Lee Smith**

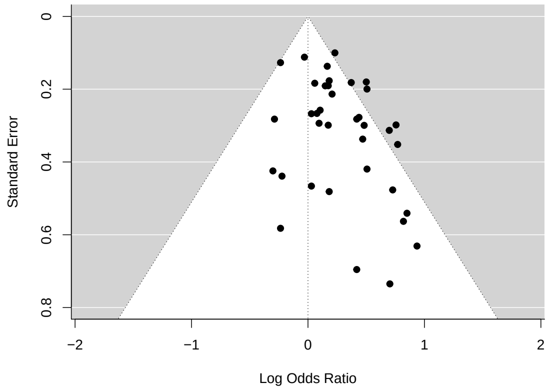
*Laboratory of Educational Research, University of Colorado Boulder, Colorado 80309, USA*



## How to Address Publication Bias

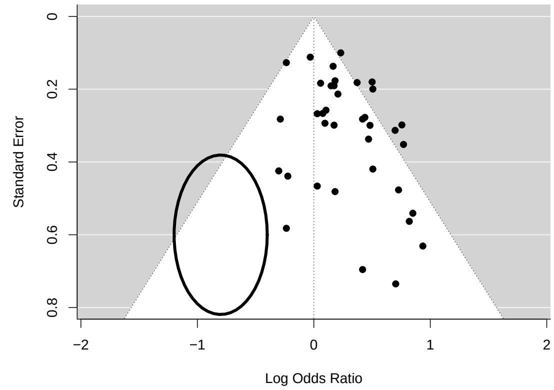
- **ultimately:** need to get rid of it (use an evidence basis that is known to be free of publication bias)
- if not available:
  - examine data for evidence of it
  - consider its potential impact
  - try to correct for it

## Example: Hackshaw (1998)






## Example: Hackshaw (1998)





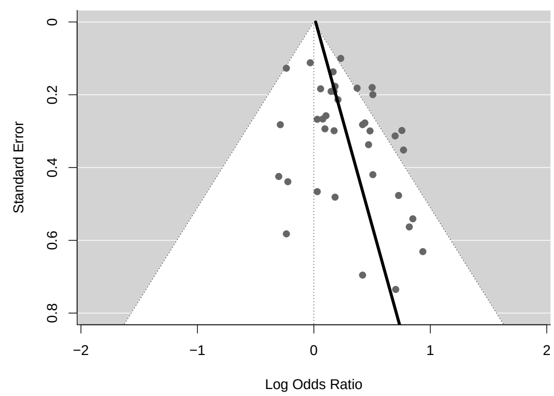

## Example: Hackshaw (1998)

```
## Random-Effects Model (k = 37; tau^2 estimator: DL)
##
## tau^2 (estimated amount of total heterogeneity): 0.0170 (SE = 0.0171)
## tau (square root of estimated tau^2 value):      0.1305
## I^2 (total heterogeneity / total variability):   24.21%
## H^2 (total variability / sampling variability):  1.32
##
## Test for Heterogeneity:
## Q(df = 36) = 47.4979, p-val = 0.0952
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.2139  0.0471  4.5390  <.0001  0.1215  0.3062

##  pred ci.lb ci.ub pi.lb pi.ub
##  1.24  1.13  1.36  0.94  1.63
```

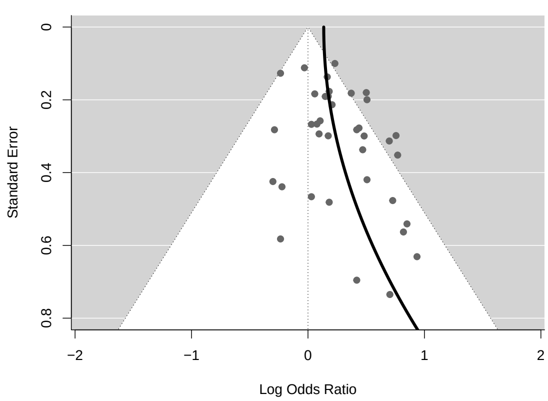


## Regression Test ($SE_i$ as predictor)



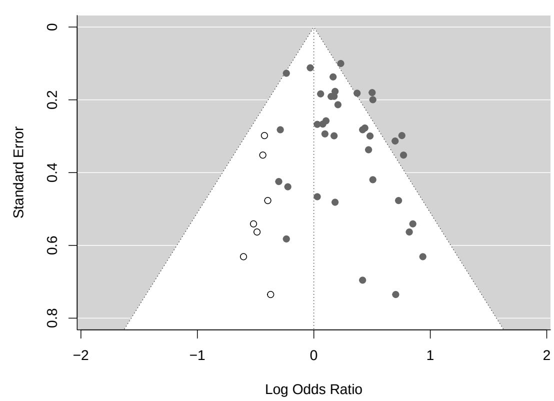



## Regression Test ($v_i$ as predictor)



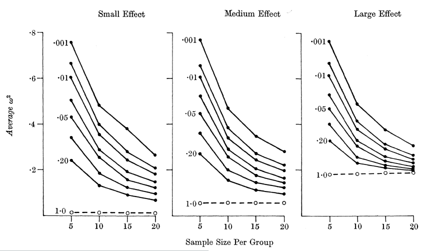



## Trim and Fill Method



## Modeling Selection Effects

- Lane & Dunlap (1978) conducted a simulation study to examine the bias when only significant studies are published[1]
- Hedges (1984) showed how to obtain these results analytically




[1] Also first use of the term ‘publication bias’ according to my searches.



## Modeling Selection Effects

- the idea was further extended by Iyengar & Greenhouse (1988)
- proposed two (slightly more realistic) models:

$$(8) \quad w_1(x;\beta,q) = \begin{cases} \dfrac{|x|^\beta}{t(q,.05)^\beta}, & \text{if } \ |x| \le t(q,.05), \\ 1, & \text{otherwise,} \end{cases}$$

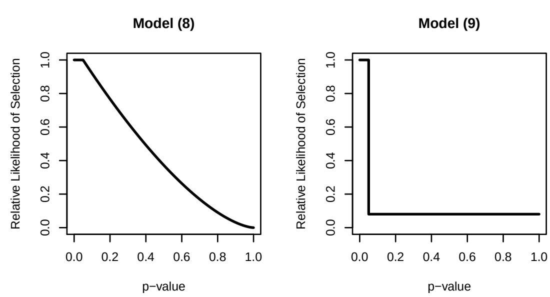
and

$$(9) \quad w_2(x;\gamma,q) = \begin{cases} e^{-\gamma}, & \text{if } \ |x| \le t(q,.05), \\ 1, & \text{otherwise.} \end{cases}$$

- two special cases:
  - $\beta = 0$ and $\gamma = 0$: no selection
  - $\beta \to \infty$ and $\gamma \to \infty$: as in Hedges (1984)



## Modeling Selection Effects






## Modeling Selection Effects: Example

TABLE 4
*Studies of effects of open vs. traditional education on creativity*

| $i$ | $N_i$ | $\hat{\theta}_i$ | $t_i$ | $q_i$ |
|---|---|---|---|---|
| 1 | 90 | −0.583 | −3.91 | 178 |
| 2 | 40 | 0.535 | 2.39 | 78 |
| 3 | 36 | 0.779 | 3.31 | 70 |
| 4 | 20 | 1.052 | 3.33 | 38 |
| 5 | 22 | 0.563 | 1.87 | 42 |
| 6 | 10 | 0.308 | 0.69 | 18 |
| 7 | 10 | 0.081 | 0.18 | 18 |
| 8 | 10 | 0.598 | 1.34 | 18 |
| 9 | 39 | −0.178 | −0.79 | 76 |
| 10 | 50 | −0.234 | −1.17 | 98 |



## Modeling Selection Effects: Example

```
## Fixed-Effects Model (k = 10)
##
## I^2 (total heterogeneity / total variability):   80.90%
## H^2 (total variability / sampling variability):  5.23
##
## Test for Heterogeneity:
## Q(df = 9) = 47.1106, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval    ci.lb   ci.ub
##   0.0566  0.0797  0.7105  0.4774  -0.0995  0.2128
```



## Modeling Selection Effects: Example

TABLE 5
*MLEs of effect size and weight function parameters from* (6) *to* (9)

| | |
|---|---|
| $(\hat{\theta}, \hat{\beta}) = (0.026, 1.33)$ | $(\hat{\theta}, \hat{\gamma}) = (0.022, 2.53)$ |
| $(SD(\hat{\theta}), SD(\hat{\beta})) = (0.052, 0.59)$ | $(SD(\hat{\theta}), SD(\hat{\gamma})) = (0.049, 0.65)$ |
| $V(\hat{\theta}, \hat{\beta}) = \begin{bmatrix} 0.003 & -0.003 \\ -0.003 & 0.348 \end{bmatrix}$ | $V(\hat{\theta}, \hat{\gamma}) = \begin{bmatrix} 0.002 & 0.000 \\ 0.000 & 0.417 \end{bmatrix}$ |

## Modeling Selection Effects: Example

```
## Fixed-Effects Model (k = 10)
##
## Model Results:
##
## estimate     se    zval    pval    ci.lb   ci.ub
##    0.030  0.059   0.513   0.608   -0.085   0.146
##
## Test for Selection Model Parameters:
## LRT(df = 1) = 6.329, p-val = 0.012
##
## Selection Model Results:
##
## estimate     se    zval    pval   ci.lb   ci.ub
##    1.541  0.815   1.890   0.059   0.000   3.139
```



## Modeling Selection Effects: Example

```
## Fixed-Effects Model (k = 10)
##
## Model Results:
##
## estimate     se    zval    pval    ci.lb   ci.ub
##    0.021  0.049   0.431   0.667   -0.076   0.118
##
## Test for Selection Model Parameters:
## LRT(df = 1) = 10.805, p-val = 0.001
##
## Selection Model Results:
##
##                    k  estimate     se      zval    pval   ci.lb   ci.ub
## 0     < p <= 0.05  4     1.000    ---       ---     ---     ---     ---
## 0.05 < p <= 1      6     0.080  0.052   -17.611   <.001   0.000   0.183

## -log(delta[2]) = 2.52
```



## Modeling Selection Effects: Example

- Hedges suggested extension to random-effects models
- based on other comments, Iyengar and Greenhouse also examined a third (asymmetric) weight function:

$$(2)\qquad w_3(x;\,\alpha,\,\beta) = \begin{cases} 1 & \text{for} \quad x > t(q,\,.05), \\ e^{-\alpha} & \text{for} \quad |\,x\,| \le t(q,\,.05), \\ e^{-\beta} & \text{for} \quad x \le -t(q,\,.05), \end{cases}$$

**where $t(q,\,.05)$ is defined as in $w_1$ and $w_2$ of Section 4. The maximum likelihood estimates of the parameters are $(\hat{\alpha},\,\hat{\beta},\,\hat{\theta}) = (3.6,\,2.5,\,-0.075)$, with standard errors (1.08, 2.03, 0.092), respectively.**



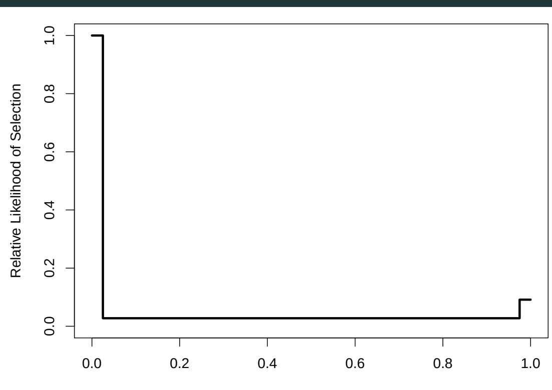
## Modeling Selection Effects: Example

```
## Fixed-Effects Model (k = 10)
##
## Model Results:
##
## estimate     se    zval    pval    ci.lb   ci.ub
##   -0.073  0.093  -0.789   0.430   -0.255   0.109
##
## Test for Selection Model Parameters:
## LRT(df = 2) = 12.288, p-val = 0.002
##
## Selection Model Results:
##
##                      k  estimate     se      zval    pval   ci.lb   ci.ub
## 0      < p <= 0.025  3     1.000    ---       ---     ---     ---     ---
## 0.025 < p <= 0.975   6     0.028  0.030   -32.314   <.001   0.000   0.087
## 0.975 < p <= 1       1     0.091  0.185    -4.912   <.001   0.000   0.454

## -log(delta[2]) = 3.59
## -log(delta[3]) = 2.39
```



## Modeling Selection Effects: Example






## Selection Models

- general class of models that attempt to model and correct for the process by which the studies may have been selected
- various selection models have been proposed:
  - step function model (Hedges, 1992)
  - with estimated thresholds (Dear & Begg, 1992)
  - with moderators (Vevea & Hedges, 1995)
  - with a priori chosen weight functions (Vevea & Woods, 2005)
  - with monotonicity constraints (Rufibach, 2011)
  - Copas selection model (Copas, 1999; Copas & Shi, 2001)
  - exponential decay models (Preston, Ashby, & Smyth, 2004)
  - p-curve (Simonsohn, Nelson, & Simmons, 2014) and p-uniform (Assen, Aert, & Wicherts, 2015) methods
  - beta selection model (Citkowicz & Vevea, 2017)

## General Setup

- let $z_i = y_i / \sqrt{v_i}$ denote the test statistic for H$_0$: $\theta_i = 0$
- let $p_i = 1 - \Phi(z_i)$, $p_i = \Phi(z_i)$, or $p_i = 2(1 - \Phi(|z_i|)$ denote the corresponding (one- or two-sided) p-value
- let $w(p_i, \vec{\delta})$ denote some function that specifies the relative likelihood of selection given the p-value of a study
- log likelihood:

$$ll = \sum_{i=1}^{k} \left\{ \ln\left(w(p_i, \vec{\delta})\right) - \frac{1}{2}\ln(\tau^2 + v_i) - \frac{1}{2}\frac{(y_i - \mu)^2}{\tau^2 + v_i} - \ln(A_i) \right\}$$

where

$$A_i = \int_{-\infty}^{\infty} w(p_i, \vec{\delta}) f(y_i, \mu, \tau^2 + v_i)\, dy_i$$



## Beta Selection Model

- proposed by Citkowicz & Vevea (2017)

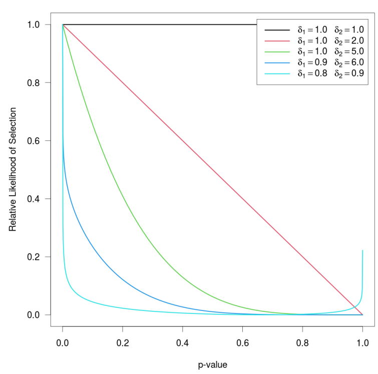
$$w(p_i) = p_i^{\delta_1 - 1} \times (1 - p_i)^{\delta_2 - 1}$$

where $\delta_1 > 0$ and $\delta_2 > 0$

- H$_0$: $\delta_1 = \delta_2 = 1$ represents the case of no selection



## Beta Selection Model



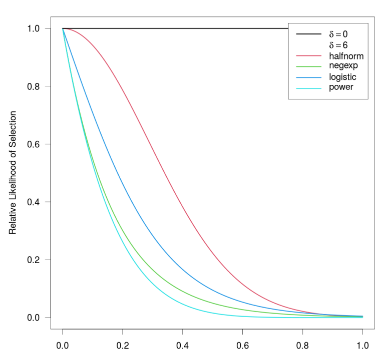



## Exponential Decay Models

- proposed by Preston, Ashby, & Smyth (2004) (except 'power')
- half-normal: $w(p_i) = \exp(-\delta \times p_i^2)$
- negative-exponential: $w(p_i) = \exp(-\delta \times p_i)$
- logistic: $w(p_i) = \dfrac{2 \times \exp(-\delta \times p_i)}{1 + \exp(-\delta \times p_i)}$
- power: $w(p_i) = (1 - p_i)^{\delta}$
- $\delta \geq 0$ and H$_0$: $\delta = 0$ represents no selection
- can extend these, in the spirit of Iyengar & Greenhouse (1988), to set $w(p_i) = 1$ for p-values below some $\alpha$ threshold

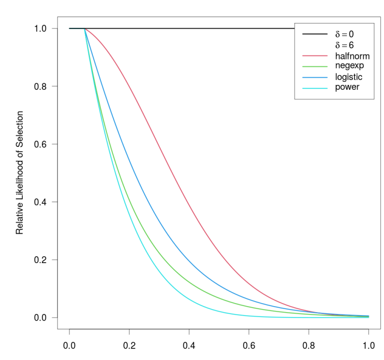


## Exponential Decay Models






## Exponential Decay Models

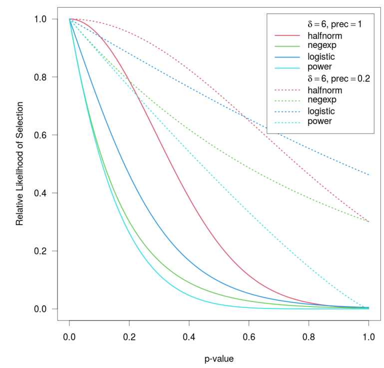


## Exponential Decay Models

- can also extend them based on some measure of (im)precision
- for example, let $s_i = \sqrt{vi}$ be the SE of $y_i$
- if two studies have similar p-values, the larger (more precise) study has a higher selection probability
- half-normal: $w(p_i) = \exp(-\delta \times s_i \times p_i^2)$
- negative-exponential: $w(p_i) = \exp(-\delta \times s_i \times p_i)$
- logistic: $w(p_i) = \dfrac{2 \times \exp(-\delta \times s_i \times p_i)}{1 + \exp(-\delta \times s_i \times p_i)}$
- power: $w(p_i) = (1 - p_i)^{-\delta \times s_i}$



## Exponential Decay Models






## Negative Exponential Power Selection Model

- described by Begg & Mazumdar (1994) for simulating data

$$w(p_i) = \exp(-\delta_1 \times p_i^{1/\delta_2})$$

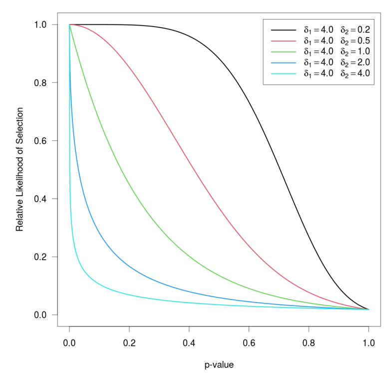
where $\delta_1 \geq 0$ and $\delta_2 \geq 0$

- $\mathrm{H_0}$: $\delta_1 = 0$ (and $\mathrm{H_0}$: $\delta_2 = 0$) represent the case of no selection
- can again be extended based on $w(p_i) = 1$ for $p_i \leq \alpha$ and by adding a measure of precision; e.g.,

$$w(p_i) = \exp(-\delta_1 \times s_i \times p_i^{1/\delta_2})$$



## Negative Exponential Power Selection Model






## Step Function Models

- based on Iyengar & Greenhouse (1988) and then Hedges (1992) and Vevea & Hedges (1995)
- let $\alpha_1 < \alpha_2 < \ldots < \alpha_c$ denote 'cutpoints'
- define $\alpha_0 = 0$ and constrain $\alpha_c = 1$

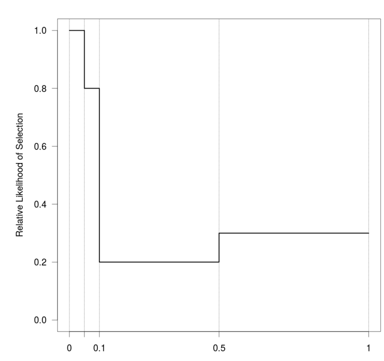
$$w(p_i) = \delta_j \quad \text{if} \quad \alpha_{j-1} < p_i \leq \alpha_j$$

and set $\delta_1 = 1$ for identifiability

- $\mathrm{H_0}$: $\delta_j = 1$ for $j = 1, \ldots, c$ implies no selection
- 'three-parameter selection model' (3PSM) is a special case with a single cutpoint (and parameters $\mu$, $\tau^2$, and $\delta_2$)



## Step Function Models

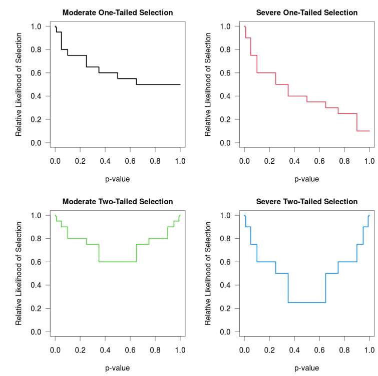


## Step Function Models

- models with many cutpoints are difficult to fit (many parameters)
- Vevea & Woods (2005) proposed to use a priori chosen functions
- can conduct sensitivity analyses by varying the weight function

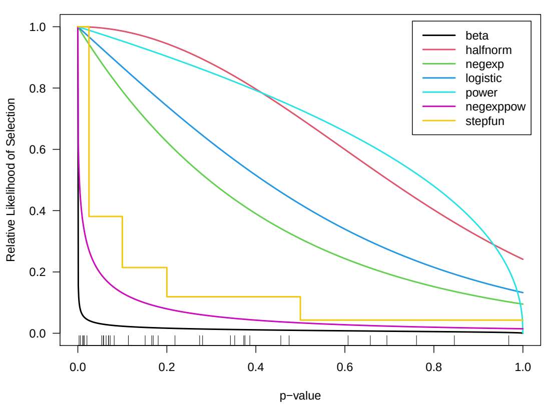


## Step Function Models






## Example: Hackshaw (1998)






## Example: Hackshaw (1998)

| | estimate | se | zval | pval | ci.lb | ci.ub |
|---|---|---|---|---|---|---|
| unadjusted | 0.186 | 0.037 | 4.980 | 0.000 | 0.113 | 0.259 |
| beta | 0.024 | 0.073 | 0.326 | 0.744 | -0.119 | 0.166 |
| halfnorm | 0.140 | 0.054 | 2.607 | 0.009 | 0.035 | 0.246 |
| negexp | 0.077 | 0.072 | 1.068 | 0.286 | -0.064 | 0.218 |
| logistic | 0.100 | 0.066 | 1.532 | 0.126 | -0.028 | 0.229 |
| power | 0.146 | 0.054 | 2.710 | 0.007 | 0.040 | 0.251 |
| negexppow | 0.000 | 0.072 | -0.003 | 0.997 | -0.141 | 0.141 |
| stepfun | 0.000 | 0.065 | 0.001 | 0.999 | -0.127 | 0.127 |



## p-uniform

- p-uniform method (Assen, Aert, & Wicherts, 2015) is in essence identical to:
  - selecting only the significant studies (on one side)
  - assuming only significant studies are published (Hedges, 1984)
  - estimating $\theta$ under this selection model
- can replicate this by also selecting only the significant studies, fitting a step function model with a single cutpoints at $\alpha = .025$, and forcing $\delta_2 \to 0$



## p-uniform

| | estimate | se | zval | pval | ci.lb | ci.ub |
|---|---|---|---|---|---|---|
| unadjusted | 0.186 | 0.037 | 4.980 | 0.000 | 0.113 | 0.259 |
| beta | 0.024 | 0.073 | 0.326 | 0.744 | -0.119 | 0.166 |
| halfnorm | 0.140 | 0.054 | 2.607 | 0.009 | 0.035 | 0.246 |
| negexp | 0.077 | 0.072 | 1.068 | 0.286 | -0.064 | 0.218 |
| logistic | 0.100 | 0.066 | 1.532 | 0.126 | -0.028 | 0.229 |
| power | 0.146 | 0.054 | 2.710 | 0.007 | 0.040 | 0.251 |
| negexppow | 0.000 | 0.072 | -0.003 | 0.997 | -0.141 | 0.141 |
| stepfun | 0.000 | 0.065 | 0.001 | 0.999 | -0.127 | 0.127 |
| puniform | 0.052 | NA | NA | 0.782 | -0.393 | 0.332 |
| Hedges (1984) | 0.052 | 0.180 | 0.287 | 0.774 | -0.302 | 0.405 |

## Software (R Packages)

- metasens: Copas selection model
- weightr: step function model
- puniform: p-uniform method
- dmetar: p-curve method
- metafor: various selection models
  note: currently in 'devel' version on GitHub:
  https://github.com/wviechtb/metafor



## Example: Hackshaw (1998)

```
# load metafor package
library(metafor)

# fit equal-effects model to data from Hackshaw (1998)
res <- rma(yi, vi, data=dat.hackshaw1998, method="FE")
res

# fit negative exponential power selection model
sel1 <- selmodel(res, type="negexppow")
sel1

# fit step function model
sel2 <- selmodel(res, type="stepfun", steps=c(.025, .1, .2, .5, 1))
sel2

# plot selection functions
plot(sel1)
plot(sel2, add=TRUE, col="orange")
```



## Example: Hackshaw (1998)

```
## Fixed-Effects Model (k = 37)
##
## I^2 (total heterogeneity / total variability):   24.21%
## H^2 (total variability / sampling variability):  1.32
##
## Test for Heterogeneity:
## Q(df = 36) = 47.4979, p-val = 0.0952
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.1858  0.0373  4.9797  <.0001  0.1126  0.2589
```



## Example: Hackshaw (1998)

```
## Fixed-Effects Model (k = 37)
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb   ci.ub
##  -0.0002  0.0720  -0.0033  0.9974  -0.1414  0.1409
##
## Test for Selection Model Parameters:
## LRT(df = 2) = 7.8703, p-val = 0.0195
##
## Selection Model Results:
##
##          estimate      se    zval    pval   ci.lb   ci.ub
## delta.1    4.2209  1.6818  2.5097  0.0121  0.9246  7.5172
## delta.2    3.1438  2.0108  1.5635  0.1179  0.0000  7.0849
```

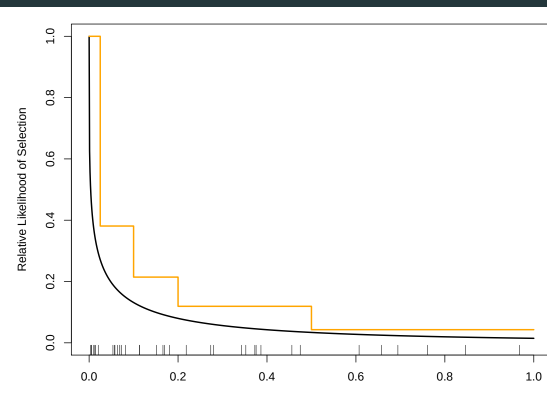


## Example: Hackshaw (1998)

```
## Fixed-Effects Model (k = 37)
##
## Model Results:
##
## estimate      se    zval    pval    ci.lb   ci.ub
##   0.0001  0.0647  0.0012  0.9990  -0.1268  0.1269
##
## Test for Selection Model Parameters:
## LRT(df = 4) = 10.6790, p-val = 0.0304
##
## Selection Model Results:
##
##                      k  estimate      se      zval    pval   ci.lb   ci.ub
## 0     < p <= 0.025   7    1.0000     ---       ---     ---     ---     ---
## 0.025 < p <= 0.1     8    0.3810  0.2108   -2.9356  0.0033  0.0000  0.7943
## 0.1   < p <= 0.2     6    0.2144  0.1382   -5.6845  <.0001  0.0000  0.4852
## 0.2   < p <= 0.5    10    0.1191  0.0827  -10.6523  <.0001  0.0000  0.2812
## 0.5   < p <= 1       6    0.0429  0.0414  -23.0958  <.0001  0.0000  0.1241
```



## Example: Hackshaw (1998)



## References [1]


Assen, M. A. van, Aert, R. C. van, & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101.

Citkowicz, M., & Vevea, J. L. (2017). A parsimonious weight function for modeling publication bias. *Psychological Methods*, *22*(1), 28–41.

Copas, J. (1999). What works? Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A*, *162*(1), 95–109.

Copas, J., & Shi, J. Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research*, *10*(4), 251–265.

Dear, K. B. G., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science*, *7*(2), 237–245.

Hackshaw, A. K. (1998). Lung cancer and passive smoking. *Statistical Methods in Medical Research*, *7*(2), 119–136.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, *9*(1), 61–85.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, *7*(2), 246–255.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem (with discussion). *Statistical Science*, *3*(1), 109–135.




## References [2]


Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 107–112.

Preston, C., Ashby, D., & Smyth, R. (2004). Adjusting for publication bias: Modelling the selection process. *Journal of Evaluation in Clinical Practice*, *10*(2), 313–322.

Rufibach, K. (2011). Selection models with monotone weight functions in meta analysis. *Biometrical Journal*, *53*(4), 689–704.

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.

Smith, M. L. (1980). Publication bias and meta-analysis. *Evaluation in Education*, 4, 22–24.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428–443.




# Thank You!

Questions, Comments, Suggestions?


wolfgang.viechtbauer@maastrichtuniversity.nl

http://www.wvbauer.com/

@wviechtb