A second look at the fail-safe N method for assessing the potential impact of publication bias on a meta-analysis

Forschungskolloquium Psychologie
Universität Münster

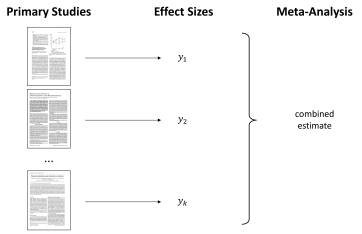
Wolfgang Viechtbauer
Maastricht University
2024-05-28

Overview

- describe **meta-analysis** and the problem of **publication bias**
- provide an overview of various **'fail-safe N'** calculation methods
- describe an **extension** to the random-effects model
- illustrate their **implementation** in the `metafor` package for R
- describe various **other methods** to address publication bias

Meta-Analysis

- a set of statistical methods for combining the results of individual studies examining a common phenomenon
- **general idea:** quantify the phenomenon of interest within each study in terms of an 'effect size' and combine these estimates using some statistical model



Example: Effectiveness of Magnesium Treatment

- used as a treatment for various medical conditions
- also considered as a potential treatment in people with acute myocardial infarction for reducing arrhythmias and mortality

Example: Effectiveness of Magnesium Treatment

- the results from the study of Rasmussen et al. (1986) [1]

	Heart Attack Fatal?		
	Yes	No	
Magnesium	9	126	135
Control	23	112	135
	32	238	270

$$p_T = 9/135 = .0667$$
$$p_C = 23/135 = .1703$$
$$RR = \frac{9/135}{23/135} = .39$$

$$y = \ln\left[\frac{9/135}{23/135}\right] = -.94$$
$$v = \frac{1}{9} - \frac{1}{135} + \frac{1}{23} - \frac{1}{135} = .140$$

Example: Effectiveness of Magnesium Treatment

```r
# load the metafor package
library(metafor)

# copy the magnesium dataset to 'dat'
dat <- dat.egger2001

# compute log risk ratios and corresponding sampling variances
dat <- escalc(measure="RR", ai=ai, n1i=n1i,
                            ci=ci, n2i=n2i,
              data=dat, add=1/2, to="all", subset=-c(8,16))

# inspect the dataset
dat
```

- notes:
  - leaving out a very small study and a study to be discussed
  - using $+\frac{1}{2}$ bias correction for all studies when computing the log risk ratios and sampling variances

Example: Effectiveness of Magnesium Treatment

id	study	year	ai	n1i	ci	n2i	yi	vi
1	Morton	1984	1	40	2	36	-0.61	1.02
2	Rasmussen	1986	9	135	23	135	-0.91	0.13
3	Smith	1986	2	200	7	200	-1.10	0.52
4	Abraham	1987	1	48	1	46	-0.04	1.29
5	Feldstedt	1988	10	150	8	148	0.20	0.20
6	Shechter	1989	1	59	9	56	-1.90	0.74
7	Ceremuzynski	1989	1	25	3	23	-0.93	0.87
9	Singh	1990	6	76	11	75	-0.58	0.21
10	Pereira	1990	1	27	7	27	-1.61	0.73
11	Shechter	1991	2	89	12	80	-1.71	0.46
12	Golf	1991	5	23	13	33	-0.55	0.18
13	Thogersen	1991	4	130	8	122	-0.70	0.32
14	LIMIT-2	1992	90	1159	118	1157	-0.27	0.02
15	Shechter	1995	4	107	17	108	-1.35	0.26

Random-Effects Model

- let y_i denote the estimate, θ_i the corresponding true effect, and v_i the sampling variance in the ith study (and hence $\text{SE}_i = \sqrt{v_i}$ denotes the corresponding standard error)
- the **random-effects model**:

$$y_i = \mu + u_i + \varepsilon_i$$

where $u_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, v_i)$

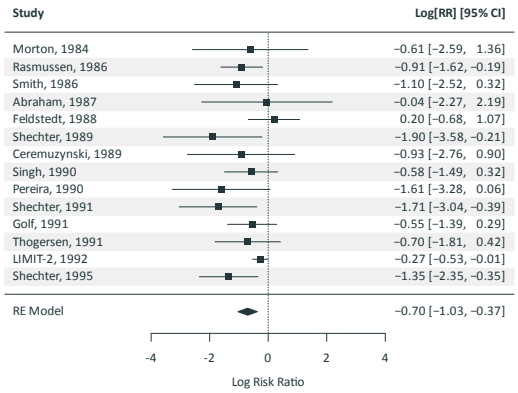
- can be easily fit with ML/REML estimation (and various other approaches for estimating τ^2 followed by WLS for estimating μ)
- the standard 'workhorse' in many MAs for pooling the estimates
- the **equal-effects model** is a special case: $\tau^2 = 0$

Example: Effectiveness of Magnesium Treatment

```
# fit a random-effects model (log risk ratios and variances as input)
res <- rma(yi, vi, data=dat)
res
```

```
## Random-Effects Model (k = 14; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.1087 (SE = 0.1356)
## tau (square root of estimated tau^2 value):      0.3296
## I^2 (total heterogeneity / total variability):   33.61%
## H^2 (total variability / sampling variability):  1.51
##
## Test for Heterogeneity:
## Q(df = 13) = 18.1711, p-val = 0.1511
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb    ci.ub
##  -0.7011  0.1686  -4.1572  <.0001  -1.0316  -0.3706
```


Example: Effectiveness of Magnesium Treatment



Example: Effectiveness of Magnesium Treatment

```
# compute the pooled risk ratio (with corresponding 95% CI/PI)
predict(res, transf=exp, digits=2)
```

```
##  pred ci.lb ci.ub pi.lb pi.ub
##  0.50  0.36  0.69  0.24  1.02
```

- Yusuf et al. (1993): "intravenous magnesium is a safe, effective, widely practicable, and inexpensive intervention" [2]

Example: Effectiveness of Magnesium Treatment

```
# compute the pooled risk ratio (with corresponding 95% CI/PI)
predict(res, transf=exp, digits=2)
```

```
##  pred ci.lb ci.ub pi.lb pi.ub
##  0.50  0.36  0.69  0.24  1.02
```

- Yusuf et al. (1993): "intravenous magnesium is a safe, effective, widely practicable, and inexpensive intervention" [2]
- but then the ISIS-4 trial [3] was published ...
- with 58,000+ participants, it yielded a risk ratio ≈ 1

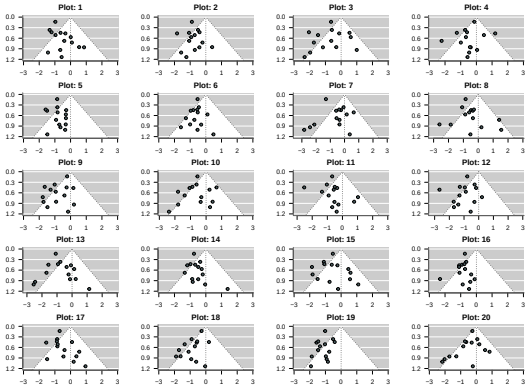
Publication Bias

- when the decision to disseminate the findings of a study depends on the results obtained within the study
- a major concern when studies with large, confirmatory, and/or statistically significant effects are more likely to be published
- published studies are easier to find and hence are more likely to be included in a meta-analysis
- the problem and its biasing impact on meta-analyses were recognized early on [4,5]

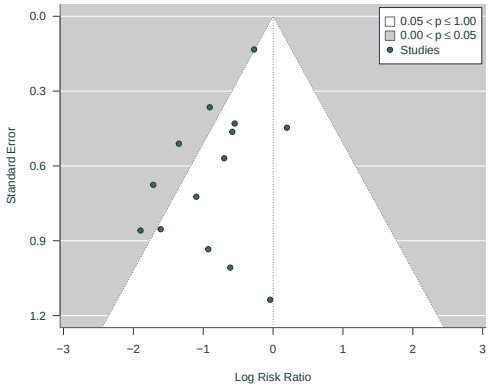
Funnel Plots

- a common graphical device to detect potential publication bias
- plot the estimates (on the x-axis) against their standard errors (on the y-axis) and superimpose a funnel (± SE)
- should look roughly symmetric without publication bias
- gaps or estimates 'hugging' the boundary of statistical significance may suggest the presence of publication bias
- interpretation is often like reading tea leaves ...

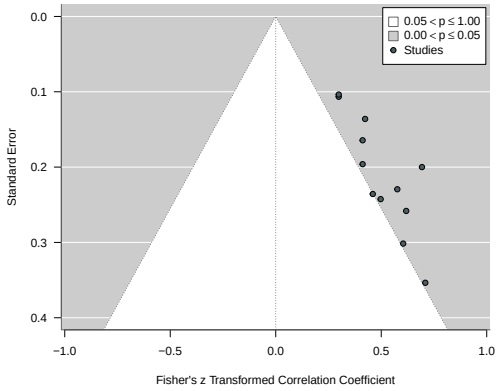
One Real, 19 Simulated Funnel Plots



Example: Effectiveness of Magnesium Treatment



Sometimes it is a bit clearer ...



Regression Test for Funnel Plot Asymmetry

- use the SEs (or some other measure of (im)precision) as a predictor of the estimates in a meta-regression model [6]
- if the plot is symmetric, the slope should be 0

```
regtest(res)
```

```
## Regression Test for Funnel Plot Asymmetry
##
## Model:     mixed-effects meta-regression model
## Predictor: standard error
##
## Test for Funnel Plot Asymmetry: z = -2.7184, p = 0.0066
## Limit Estimate (as sei -> 0):   b = -0.1300 (CI: -0.4557, 0.1957)
```

- the rank correlation test [7] is based on a similar principle

File-Drawer Problem / Analysis

- **general idea:** calculate the minimum number of studies averaging null results that would have to be added to a given set of studies to change the conclussion of a meta-analysis
- if this number (the 'fail-safe N') is unreasonably large, then this should bestow more confidence on the findings obtained
- in the original formulation by Rosenthal (1979) [8], the calculation was based on Stouffer's method for combining p-values [9]

Stouffer Method

- test of H_0: $\theta_i = 0$ in the ith study

$$z_i = \frac{y_i}{\sqrt{v_i}}$$

- test of H_0: $\theta_i = 0$ for all k studies

$$z = \frac{\sum z_i}{\sqrt{k}}$$

```
# calculate combined p-value based on Stouffer's method
dat$zi <- with(dat, yi / sqrt(vi))
z <- with(dat, sum(zi) / sqrt(nrow(dat)))
fmtp(pnorm(z), digits=4)
```

```
## [1] <.0001
```


File-Drawer Problem / Analysis

- suppose there are N additional studies with $\bar{z}_N = 0$, then:

$$z_{(k+N)} = \frac{\sum z_i + 0}{\sqrt{k+N}}$$

- hence

$$N = k \left(\frac{z}{1.645} \right)^2 - k$$

would be the N that turns a significant Stouffer test into a non-significant one (i.e., where $z_{(k+N)} = 1.645$)

Example: Effectiveness of Magnesium Treatment

```
fsn(yi, vi, data=dat)
```

```
## Fail-safe N Calculation Using the Rosenthal Approach
##
## Observed Significance Level: <.0001
## Target Significance Level:   0.05
##
## Fail-safe N: 138
```


Critiques of the Method

- often based on a misunderstanding of its purpose
- **not** a method to detect publication bias
- **not** an estimate of the number of missing studies
- **not** a way to obtain an 'adjusted' estimate
- some valid criticisms:
  - the method does not incorporate sample size information
  - the method is focused purely on statistical significance
  - we do not typically use Stouffer's method for conducting MAs
- these criticisms can and have been addressed

Further Developments of the Idea

- **Orwin's method:** calculate the minimum number of studies averaging null results that would have to be added to a given set of studies to reduce the (unweighted or weighted) average effect size to a target value [10]
- **Rosenberg's method:** calculate the minimum number of studies averaging null results that would have to be added to a given set of studies to reduce the significance level of the average effect size under an equal-effects model [11]

Orwin's Method

- need to specify a target effect size (for a trivial effect)
- for example, consider a 5% reduction in mortality risk

```
fsn(yi, vi, data=dat, type="Orwin", target=log(0.95))
```

```
## Fail-safe N Calculation Using the Orwin Approach
##
## Average Effect Size: -0.4861
## Target Effect Size:  -0.0513
##
## Fail-safe N: 119
```

- sidenote: in the original formulation by Orwin (1983), the calculations are based on an unweighted average

Orwin's Method

- if we add N studies with null results with v_i's equal to the harmonic mean of the v_i's of the original k studies, then we get the target pooled effect size
- for the example (where $N = 119$):

```
N <- 119
yi.all <- c(dat$yi, scale(rnorm(N)))
vi.all <- c(dat$vi, rep(1/mean(1/dat$vi), N))
res.all <- rma(yi.all, vi.all, method="EE")
predict(res.all, transf=exp)
```

```
##    pred  ci.lb  ci.ub
##  0.9501 0.8903 1.0140
```


Rosenberg's Method

- refocuses the calculation on a target significance level
- calculations are done under an equal-effects model

```
fsn(yi, vi, data=dat, type="Rosenberg")
```

```
## Fail-safe N Calculation Using the Rosenberg Approach
##
## Average Effect Size:          -0.4861
## Observed Significance Level:  <.0001
## Target Significance Level:    0.05
##
## Fail-safe N: 69
```


Rosenberg's Method

- if we add N studies with null results with v_i's equal to the harmonic mean of the v_i's of the original k studies, then we get a pooled effect size whose p-value is equal to 0.05
- for the example (where $N = 69$)

```
N <- 69
yi.all <- c(dat$yi, scale(rnorm(N)))
vi.all <- c(dat$vi, rep(1/mean(1/dat$vi), N))
rma(yi.all, vi.all, method="EE")
```

```
## Equal-Effects Model (k = 83)
##
## [...]
##
## estimate      se     zval    pval    ci.lb   ci.ub
##  -0.0820  0.0420  -1.9512  0.0510  -0.1643  0.0004
```


General Method

- the above shows:
  - that there is an inherent flaw in the existing methods: when combining the k original studies and the N studies averaging null results, we should get heterogeneity
  - the generalization to a random-effects model is possible
- assume that the amount of heterogeneity in the N studies is like the amount of heterogeneity in the k original studies
- let $\tilde{y}_1, \ldots, \tilde{y}_N$ denote N random values with mean 0 and variance $\bar{v} + \hat{\tau}^2$, where $\bar{v}$ is the harmonic mean of the v_i's and $\hat{\tau}^2$ is the estimate of τ^2 from a random-effects model
- fit a random-effects model to the $(y_1, \ldots, y_k, \tilde{y}_1, \ldots, \tilde{y}_N)$ with sampling variances $(v_1, \ldots, v_k, \bar{v}, \ldots, \bar{v})$
- find N for some target significance level or pooled effect size

General Method

- find N to reach some target significance level

```
fsn(yi, vi, data=dat, type="General", exact=TRUE)
```

```
## Fail-safe N Calculation Using the General Approach
##
## Average Effect Size:          -0.7011 (with file drawer: -0.1732)
## Amount of Heterogeneity:       0.1087 (with file drawer:  0.1548)
## Observed Significance Level:  <.0001 (with file drawer:  0.0515)
## Target Significance Level:    0.05
##
## Fail-safe N: 29
```

General Method

- find N to reach some target pooled effect size

```
fsn(yi, vi, data=dat, type="General", target=log(0.95), exact=TRUE)
```

```
## Fail-safe N Calculation Using the General Approach
##
## Average Effect Size:         -0.7011 (with file drawer: -0.0507)
## Amount of Heterogeneity:      0.1087 (with file drawer:  0.1233)
## Observed Significance Level:  <.0001 (with file drawer:  0.2733)
## Target Effect Size:          -0.0513
##
## Fail-safe N: 117
```


General Method

- repeatedly fitting the RE model is computationally expensive
- can speed up the calculations via some analytic derivations
- the difference is typically negligible

```
fsn(yi, vi, data=dat, type="General", exact=TRUE)$fsnum
```

```
## [1] 29
```

```
fsn(yi, vi, data=dat, type="General")$fsnum
```

```
## [1] 26
```

```
fsn(yi, vi, data=dat, type="General", target=log(0.95), exact=TRUE)$fsnum
```

```
## [1] 117
```

```
fsn(yi, vi, data=dat, type="General", target=log(0.95))$fsnum
```

```
## [1] 120
```


General Method

- what if the mean effect size of the file drawer studies is not 0?
- if it is opposite in sign to the observed pooled effect size, then this will reduce N, sometimes considerably

```
fsn(yi, vi, data=dat, type="General")$fsnum
```

```
## [1] 26
```

```
fsn(yi, vi, data=dat, type="General", mumiss=log(1.05))$fsnum
```

```
## [1] 18
```

```
fsn(yi, vi, data=dat, type="General", target=log(0.95))$fsnum
```

```
## [1] 120
```

```
fsn(yi, vi, data=dat, type="General", target=log(0.95), mumiss=log(1.05))$fsnum
```

```
## [1] 65
```


Fail-Safe N Results

- the results of the various methods:

method	N
Rosenthal (alpha=.05)	138
Orwin (target=log(0.95))	119
Rosenberg (alpha=.05)	69
General (alpha=.05)	29
General (target=log(0.95))	117
General (alpha=.05, mumiss=log(1.05))	18
General (target=log(0.95), mumiss=log(1.05))	65

- no criterion to judge if N is 'large'
- Rosenthal [8] suggested $5k + 10$ (for the example: 80), but this is of course totally arbitrary

Other Publication Bias Methods

- methods to detect evidence of publication bias
  - funnel plots
  - regression test / rank correlation test
  - test of excess significance
- methods to adjust for publication bias
  - PET/PEESE methods
  - trim and fill method
  - selection models
  - p-uniform method / p-curve method

Test of Excess Significance

- method proposed by Ioannidis and Trikalinos (2007) [12]
- recall: can test H_0: $\theta_i = 0$ with $z_i = y_i / \sqrt{v_i}$
- O: the observed number of significant tests
- compute the power for each test, $1 - \beta_i$, given some (estimated) value of θ or μ
- $E = \sum(1 - \beta_i)$: the expected number of significant tests
- test if O is significantly larger than E (exact test, Pearson test, or binomial test)

Test of Excess Significance

```
tes(yi, vi, data=dat)
```

```
## Test of Excess Significance
##
## Observed Number of Significant Findings: 5 (out of 14)
## Expected Number of Significant Findings: 2.7140
## Observed Number / Expected Number:       1.8423
##
## Estimated Power of Tests (based on theta = -0.4861)
##
##    min      q1  median      q3     max
## 0.0712  0.0875  0.1239  0.1902  0.9542
##
## Test of Excess Significance: p = 0.0814 (exact test)
## Limit Estimate (theta_lim):  -0.5111 (where p = 0.1)
```


PET/PEESE Methods

- proposed by Stanley and Doucouliagos (2014) [13]
- uses the intercept of the model for the regression test (either using the standard errors or sampling variances as predictor) as a test of a 'genuine effect'
- can be thought of as a projection to a study with infinite sample size (i.e., where $\text{SE}_i \to 0$ or $v_i \to 0$)

PET/PEESE Methods

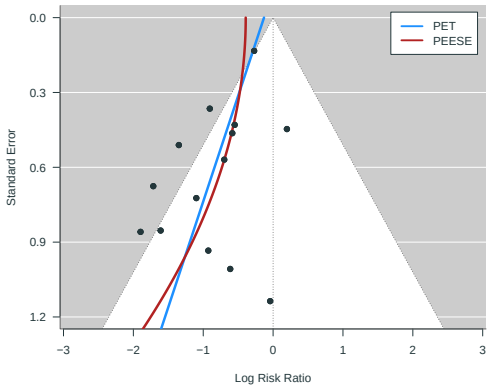
```
regtest(res)
```

```
## Regression Test for Funnel Plot Asymmetry
##
## Model:     mixed-effects meta-regression model
## Predictor: standard error
##
## Test for Funnel Plot Asymmetry: z = -2.7184, p = 0.0066
## Limit Estimate (as sei -> 0):   b = -0.1300 (CI: -0.4557, 0.1957)
```

```
regtest(res, predictor="vi")
```

```
## Regression Test for Funnel Plot Asymmetry
##
## Model:     mixed-effects meta-regression model
## Predictor: sampling variance
##
## Test for Funnel Plot Asymmetry: z = -1.8766, p = 0.0606
## Limit Estimate (as vi -> 0):    b = -0.3918 (CI: -0.7259, -0.0578)
```


PET/PEESE Methods



Trim and Fill Method

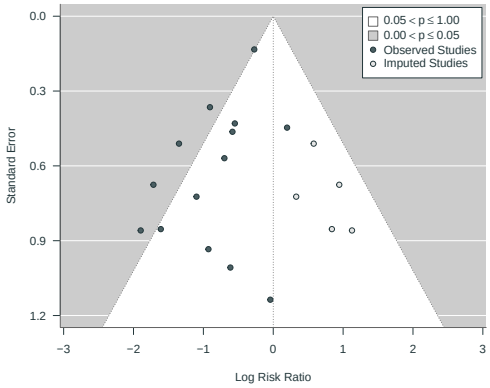
- proposed by Duval and Tweedie (2000a, 2000b) [14,15]
- assumes that the most extreme effects on one side of the funnel plot have been suppressed
- uses information from the other side of the plot to impute the missing effects

Trim and Fill Method

```
trimfill(res)
```

```
## Estimated number of missing studies on the right side: 5 (SE = 2.3774)
##
## Random-Effects Model (k = 19; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.2434 (SE = 0.1889)
## tau (square root of estimated tau^2 value):      0.4933
## I^2 (total heterogeneity / total variability):   50.18%
## H^2 (total variability / sampling variability):  2.01
##
## Test for Heterogeneity:
## Q(df = 18) = 32.6731, p-val = 0.0183
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb    ci.ub
##  -0.4318  0.1799  -2.4002  0.0164  -0.7844  -0.0792
```

Trim and Fill Method



Selection Models

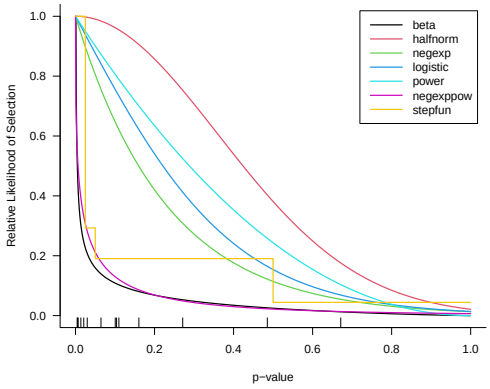
- first proposed in the context of MAs by Hedges (1984) [16] and Iyengar and Greenhouse (1988) [17]
- assume an inverse relationship between the p-value of the test H_0: $\theta_i = 0$ and the probability that a study is included in a MA
- with enough studies, can estimate this relationship and the size of the effect in the entire population of studies
- difficult in practice (k must be quite large, especially when using RE models)

Selection Models

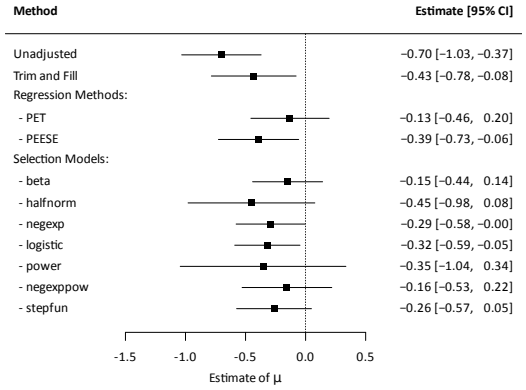
```
selmodel(res, type="logistic", alternative="less")
```

```
## Random-Effects Model (k = 14; tau^2 estimator: ML)
##
## tau^2 (estimated amount of total heterogeneity): 0.0000
## tau (square root of estimated tau^2 value):      0.0069
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb    ci.ub
##  -0.3185  0.1386  -2.2976  0.0216  -0.5902  -0.0468
##
## Test for Selection Model Parameters:
## LRT(df = 1) = 3.6143, p-val = 0.0573
##
## Selection Model Results:
##
## estimate      se     zval    pval   ci.lb   ci.ub
##   4.9695  2.1257  2.3379  0.0194  0.8033  9.1357
```


Selection Models



Selection Models



Comparison of Findings from Different Sources

- an important (but often neglected) idea (e.g., published versus unpublished, journal articles versus theses/reports)
- Smith (1980): "in every one of the ten instances in which the comparison can be made, the average effect from studies published in journals is larger than the corresponding effect estimated from theses and dissertations" [5] (by about 33%)
- cannot make such a comparison for the magnesium example

Publication Bias

- affects all review methods (not a problem exclusive to MAs!)
- due to meta-analysis:
  - increased awareness of the problem
  - development of methods to address the problem
  - increased emphasis on study registration
- gold standard: meta-analysis of registered reports
- same as a 'prospective meta-analysis' [18,19]
- if not possible, acknowledge/examine the multitude of possible results (multiverse analysis [20])

References [1]

1. Rasmussen, H. S., Norregard, P., Lindeneg, O., Mcnair, P., Backer, V., & Balslev, S. (1986). Intravenous magnesium in acute myocardial infarction. *The Lancet*, *327*(8475), 234–236. https://doi.org/10.1016/S0140-6736(86)90773-7
2. Yusuf, S., Teo, K., & Woods, K. (1993). Intravenous magnesium in acute myocardial infarction: An effective, safe, simple, and inexpensive treatment. *Circulation*, *87*(6), 2043–2046. https://doi.org/10.1161/01.cir.87.6.2043
3. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. (1995). ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet*, *345*(8951), 669–685. https://doi.org/10.1016/S0140-6736(95)90865-X
4. 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. https://doi.org/10.2307/2282137
5. Smith, M. L. (1980). Publication bias and meta-analysis. *Evaluation in Education*, *4*, 22–24. https://doi.org/10.1016/0191-765X(80)90004-X

References [2]

6. Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629
7. Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446
8. Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638
9. Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams Jr., R. M. (1949). *The american soldier: Adjustment during army life (vol. 1)*. Princeton University Press.
10. Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. https://doi.org/10.3102/10769986008002157
11. Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468. https://doi.org/10.1111/j.0014-3820.2005.tb01004.x
12. Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, *4*(3), 245–253. https://doi.org/10.1177/1740774507079441

References [3]

13. Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095
14. Duval, S. J., & 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. https://doi.org/10.1080/01621459.2000.10473905
15. Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x
16. 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. https://doi.org/10.3102/10769986009001061
17. Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem (with discussion). *Statistical Science*, *3*(1), 109–135. https://doi.org/10.1214/ss/1177013012
18. Simes, R. J. (1995). Prospective meta-analysis of cholesterol-lowering studies: The prospective pravastatin pooling (PPP) project and the cholesterol treatment trialists (CTT) collaboration. *American Journal of Cardiology*, *76*(9), 122C–126C. https://doi.org/10.1016/s0002-9149(99)80482-2

References [4]

19. Berlin, J. A., & Colditz, G. A. (1999). The role of meta-analysis in the regulatory process for foods, drugs, and devices. *Journal of the American Medical Association*, *281*(9), 830–834. https://doi.org/10.1001/jama.281.9.830
20. Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? *Zeitschrift Für Psychologie / Journal of Psychology*, *227*(1), 64–82. https://doi.org/10.1027/2151-2604/a000357

Thank You for Your Attention!

Questions, Comments, Suggestions?

wolfgang.viechtbauer@maastrichtuniversity.nl

https://www.wvbauer.com

https://www.metafor-project.org

@wviechtb