A fail-safe N computation based on the random-effects model

Society for Research Synthesis Methodology
2024 Annual Conference
Amsterdam, The Netherlands
June 26-28, 2024

Wolfgang Viechtbauer
Maastricht University

2024-06-28

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
- various meta-analyses (and re-analyses) have been conducted on this topic [1–7] and it has often been used as an example to discuss publication bias

Example: Effectiveness of Magnesium Treatment

```
# 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 the ISIS-4 trial [8]
  - using $+\frac{1}{2}$ bias correction for all studies when computing the log risk ratios and sampling variances

Example: Effectiveness of Magnesium Treatment

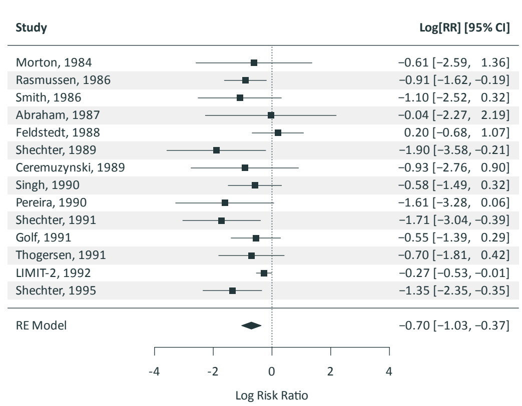
id	study	year	ai	n1i	ci	n2i	yi	vi
1	Morton	1984	1	40	2	36	-0.61	1.02
2	Rasmussen	1986	9	135	23	135	-0.91	0.13
3	Smith	1986	2	200	7	200	-1.10	0.52
4	Abraham	1987	1	48	1	46	-0.04	1.29
5	Feldstedt	1988	10	150	8	148	0.20	0.20
6	Shechter	1989	1	59	9	56	-1.90	0.74
7	Ceremuzynski	1989	1	25	3	23	-0.93	0.87
9	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

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



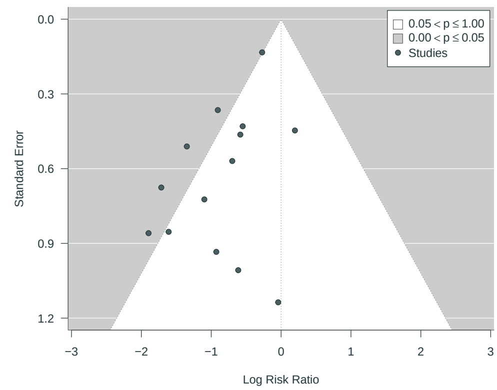
Example: Effectiveness of Magnesium Treatment

```
# 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" [3]
- but then the ISIS-4 trial [8] was published …
- with 58,000+ participants, it yielded a risk ratio ≈ 1
- which raised discussions around potential **publication bias**

Example: Effectiveness of Magnesium Treatment



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) [9], the calculation was based on Stouffer's method for combining p-values [10]

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 [11]
- **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 [12]

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, digits=2)
```

```
## pred ci.lb ci.ub
## 0.95  0.89  1.01
```


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", digits=2)
```

```
## Equal-Effects Model (k = 83)
##
## [...]
##
## estimate    se    zval  pval  ci.lb  ci.ub
##    -0.08  0.04  -1.95  0.05  -0.16   0.00
```


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:         -0.70 (with file drawer: -0.17)
## Amount of Heterogeneity:      0.11 (with file drawer:  0.15)
## Observed Significance Level:  <.01 (with file drawer:  0.05)
## 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:         -0.70 (with file drawer: -0.05)
## Amount of Heterogeneity:      0.11 (with file drawer:  0.12)
## Observed Significance Level:  <.01 (with file drawer:  0.27)
## Target Effect Size:          -0.05
##
## Fail-safe N: 117
```


General Method

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

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

```
## Fail-safe N: 29
```

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

```
## Fail-safe N: 26
```

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

```
## Fail-safe N: 117
```

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

```
## Fail-safe N: 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")
```

```
## Fail-safe N: 26
```

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

```
## Fail-safe N: 18
```

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

```
## Fail-safe N: 120
```

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

```
## Fail-safe N: 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 [9] suggested $5k + 10$ (for the example: 80), but this is of course totally arbitrary

Final Comments

- some have suggested to abandon fail-safe N calculations
- I think if it is used and interpreted correctly, it is one of various useful methods in our 'meta-analytic tool belt'
- the real solution to publication bias is to get rid of it, but ...

References [1]

1. Teo, K. K., Yusuf, S., Collins, R., Held, P. H., & Peto, R. (1991). Effects of intravenous magnesium in suspected acute myocardial infarction: Overview of randomised trials. *British Medical Journal*, *303*(6816), 1499–1503. https://doi.org/10.1136/bmj.303.6816.1499
2. Horner, S. M. (1992). Efficacy of intravenous magnesium in acute myocardial infarction in reducing arrhythmias and mortality: Meta-analysis of magnesium in acute myocardial infarction. *Circulation*, *86*(3), 774–779. https://doi.org/10.1161/01.cir.86.3.774
3. 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
4. Egger, M., & Davey Smith, G. (1995). Misleading meta-analysis: Lessons from "an effective, safe, simple" intervention that wasn't. *British Medical Journal*, *310*(6982), 752–754. https://doi.org/10.1136/bmj.310.6982.752
5. Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, *54*(10), 1046–1055. https://doi.org/10.1016/s0895-4356(01)00377-8

References [2]

6. Higgins, J. P. T., & Spiegelhalter, D. J. (2002). Being sceptical about meta-analyses: A Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology*, *31*(1), 96–104. https://doi.org/10.1093/ije/31.1.96
7. Li, J., Zhang, Q., Zhang, M., & Egger, M. (2007). Intravenous magnesium for acute myocardial infarction. *Cochrane Database of Systematic Reviews*, *2*, CD002755. https://doi.org/10.1002/14651858.CD002755.pub2
8. 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
9. 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
10. 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.
11. 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

References [3]

12. 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

Thank You for Your Attention!

Questions, Comments, Suggestions?

wolfgang.viechtbauer@maastrichtuniversity.nl

https://www.wvbauer.com

https://www.metafor-project.org

@wviechtb