Automated report generation for meta-analyses using the R package metafor

Evidence Synthesis Technology Webinar

September 8, 2020

Wolfgang Viechtbauer

Maastricht University

2020-09-08

The `metafor` Package

- **metafor**: a package for conducting meta-analyses with R
- grew out of code I wrote for my dissertation research (~2000)
- turned into a full R package in 2009
- 25 updated releases since then

- **paper**: Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03
- CRAN: https://cran.r-project.org/package=metafor
- package website: http://www.metafor-project.org/
- documentation: https://wviechtb.github.io/metafor/

Quick Demo: The Usual Workflow

- compute observed outcomes / effect size estimates
- analyze with some meta-analytic model
- further analysis steps
    - model diagonostics
    - forest/funnel plots
    - checks for publication bias
    - ...
- write up methods and results for paper

Example: BCG Vaccine

- BCG: Bacillus Calmette-Guérin
- BCG is a vaccine against tuberculosis
- effectiveness studies: compare proportion of TB positive cases in vaccinated and non-vaccinated group

---

Results from One Trial

	Positive	Negative	Total
Vaccinated	4	119	123
Not Vaccinated	11	128	139

$$p_T = 4/123 = .0325$$

$$p_C = 11/139 = .0791$$

$$RR = \frac{4/123}{11/139} = .41$$

$$y = \ln[RR] = \ln\left[\frac{4/123}{11/139}\right] = -.89$$

$$v = \frac{1}{4} - \frac{1}{123} + \frac{1}{11} - \frac{1}{139} = .326$$

Results from 13 Trials

```
# load metafor package
library(metafor)

# examine BCG dataset
dat.bcg
```

trial	author	year	tpos	tneg	cpos	cneg	ablat	alloc
1	Aronson	1948	4	119	11	128	44	random
2	Ferguson & Simes	1949	6	300	29	274	55	random
3	Rosenthal et al	1960	3	228	11	209	42	random
4	Hart & Sutherland	1977	62	13536	248	12619	52	random
5	Frimodt-Moller et al	1973	33	5036	47	5761	13	alternate
6	Stein & Aronson	1953	180	1361	372	1079	44	alternate
7	Vandiviere et al	1973	8	2537	10	619	19	random
8	TPT Madras	1980	505	87886	499	87892	13	random
9	Coetzee & Berjak	1968	29	7470	45	7232	27	random
10	Rosenthal et al	1961	17	1699	65	1600	42	systematic
11	Comstock et al	1974	186	50448	141	27197	18	systematic
12	Comstock & Webster	1969	5	2493	3	2338	33	systematic
13	Comstock et al	1976	27	16886	29	17825	33	systematic

Quick Demo: The Usual Workflow

```r
# calculate log risk ratios and corresponding sampling variances
dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos, di=cneg,
              slab=paste(author, ", ", year, sep=""), data=dat.bcg)

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

## Random-Effects Model (k = 13; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.3132 (SE = 0.1664)
## tau (square root of estimated tau^2 value):      0.5597
## I^2 (total heterogeneity / total variability):   92.22%
## H^2 (total variability / sampling variability):  12.86
##
## Test for Heterogeneity:
## Q(df = 12) = 152.2330, p-val < .0001
##
## Model Results:
##
## estimate      se     zval    pval    ci.lb    ci.ub
## -0.7145  0.1798  -3.9744  <.0001  -1.0669  -0.3622  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
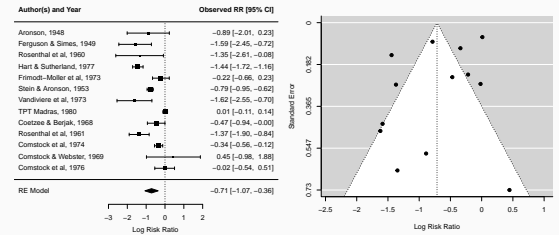

Quick Demo: The Usual Workflow

```r
# then do lots more stuff ...
forest(res)
funnel(res)
influence(res)
ranktest(res)
regtest(res)
```

The reporter() Function

- automatically generates an analysis report
- describes the statistical methods used
- gives a natural language summary of the results
- includes a forest and a funnel plot
- gives references for all methods used
- output can be html, pdf, or docx

Quick Demo: Using reporter()

```r
reporter(res)
```

Directory for generating the report is: /tmp/RtmpcDgVja

Copying references.bib and apa.csl to report directory ...
Saving model object to report_res.rdata ...
Creating report_res.rmd file ...
Rendering report_res.rmd file ...
Generated /tmp/RtmpcDgVja/report_res.html ...
Opening report ...

Quick Demo: Using reporter()

Analysis Report

Generated with the reporter() Function of the metafor Package

07 September, 2020

Methods

The analysis was carried out using the log risk ratio as the outcome measure. A random-effects model was fitted to the data. The amount of heterogeneity (i.e., τ^2), was estimated using the restricted maximum-likelihood estimator (Viechtbauer, 2005). In addition to the estimate of τ^2, the Q-test for heterogeneity (Cochran, 1954) and the I^2 statistic (Higgins & Thompson, 2002) are reported. In case any amount of heterogeneity is detected (i.e., $\hat{\tau}^2 > 0$, regardless of the results of the Q-test), a prediction interval for the true outcomes is also provided (Riley et al., 2011). Studentized residuals and Cook's distances are used to examine whether studies may be outliers and/or influential in the context of the model (Viechtbauer & Cheung, 2010). Studies with a studentized residual larger than the $100 \times (1 - 0.05/(2 \times k))$th percentile of a standard normal distribution are considered potential outliers (i.e., using a Bonferroni correction with two-sided $\alpha = 0.05$ for k studies included in the meta-analysis). Studies with a Cook's distance larger than the median plus six times the interquartile range of the Cook's distances are considered to be influential. The rank correlation test (Begg & Mazumdar, 1994) and the regression test (Sterne & Egger, 2005), using the standard error of the observed outcomes as predictor, are used to check for funnel plot asymmetry. The analysis was carried out using R (version 4.0.2) (R Core Team, 2020) and the **metafor** package (version 2.5.47) (Viechtbauer, 2010).

Results

A total of $k = 13$ studies were included in the analysis. The observed log risk ratios ranged from -1.6209 to 0.4459, with the majority of estimates being negative (85%). The estimated average log risk ratio based on the random-effects model was $\hat{\mu} = -0.7145$ (95% CI: -1.0669 to -0.3622). Therefore, the average outcome differed significantly from zero ($z = -3.9744$, $p < 0.0001$). A forest plot showing the observed outcomes and the estimate based on the random-effects model is shown in Figure 1.
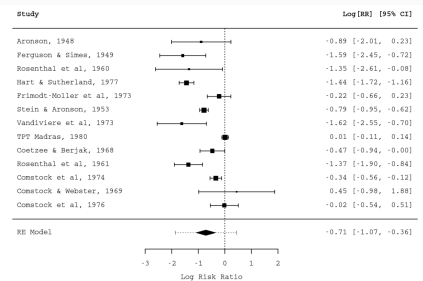

Quick Demo: Using reporter()

Figure 1: Forest plot showing the observed outcomes and the estimate of the random-effects model

According to the Q-test, the true outcomes appear to be heterogeneous ($Q(12) = 152.2330$, $p < 0.0001$, $\hat{\tau}^2 = 0.3132$, $I^2 = 92.2214\%$). A 95% prediction interval for the true outcomes is given by -1.8667 to 0.4376. Hence, although the average outcome is estimated to be negative, in some studies the true outcome may in fact be positive.

Slide 13

An examination of the studentized residuals revealed that none of the studies had a value larger than ± 2.8905 and hence there was no indication of outliers in the context of this model. According to the Cook's distances, none of the studies could be considered to be overly influential.

A funnel plot of the estimates is shown in Figure 2. Neither the rank correlation nor the regression test indicated any funnel plot asymmetry ($p = 0.9524$ and $p = 0.4218$, respectively).
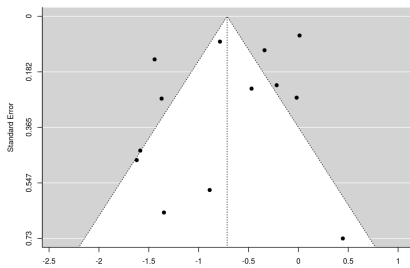

Figure 2: Funnel plot

13

Slide 14

Notes

This analysis report was dynamically generated for model object `res` with the `reporter()` function of the **metafor** package. The model call that was used to fit the model was `rma(yi = yi, vi = vi, data = dat)`. This report provides an illustration of how the results of the model can be reported, but is not a substitute for a careful examination of the results.

References

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

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101–129. https://doi.org/10.2307/3001666

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

R Core Team. (2020). *R: A language and environment for statistical computing* (Issues). R Foundation for Statistical Computing. https://www.R-project.org/

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *British Medical Journal*, *342*, d549. https://doi.org/10.1136/bmj.d549

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustment* (Issues, pp. 99–110). Wiley.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

14

Slide 15

To-Do's

· make `reporter()` work with meta-regression models

· other ideas:

  · option to suppress forest/funnel plots?

  · allow transformation of results?

  · add explanatory footnotes?

  · extend to `rma.mh`, `rma.peto`, and `rma.glmm` objects?

  · more customization? (group names, outcome name, …)

  · …

15

Slide 16

wolfgang.viechtbauer@maastrichtuniversity.nl

wvb@wvbauer.com

http://www.wvbauer.com/

http://www.metafor-project.org/

https://twitter.com/wviechtb

16