Visualization of heterogeneity in forest plots

European Congress of Methodology
Tenerife, Spain
July 23-25, 2025

Wolfgang Viechtbauer
Maastricht University

2025-07-24

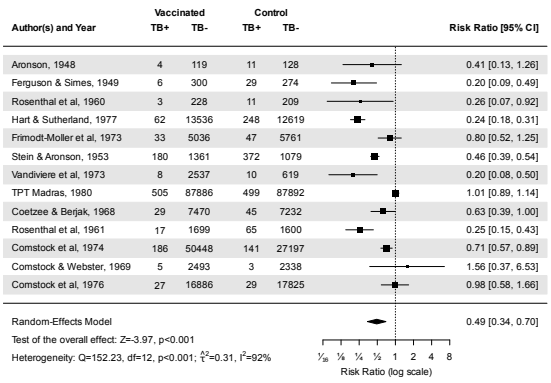
Forest Plots

- started to appear in meta-analyses in the early 1980s [1]
- the **most commonly used** visualization in meta-analyses [2]
- provides information about the findings of the individual studies and the results of the meta-analysis [3]
- **recommended** in various reporting guidelines: PRISMA [4], MOOSE [5], JARS [6], Cochrane [7], Campbell [8], ...

Forest Plots

- combination of **textual info** and various **graphical elements**:
  - study identifier (typically first author + year)
  - summary statistics to compute the effect sizes
    - means, standard deviations, study/group sizes
    - counts (e.g., of the cells in 2x2 tables)
  - effect sizes of the studies with corresponding CIs
    - as points/boxes and horizontal lines
    - often also added textually
  - study weights (in %)
  - a polygon ('diamond') for the pooled estimate ($\hat{\mu}$) and its CI
  - info about the statistical significance of the pooled estimate
  - info about potential heterogeneity (e.g., Q-test, I^2, $\hat{\tau}^2$)
  - risk of bias assessment for the individual studies

Example: BCG Vaccine Meta-Analysis [9]

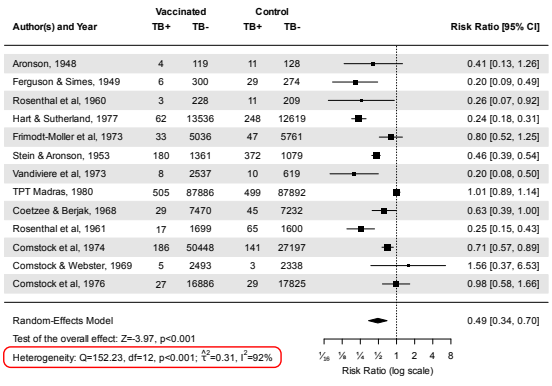


Author(s) and Year	Vaccinated TB+	Vaccinated TB-	Control TB+	Control TB-	Risk Ratio [95% CI]
Aronson, 1948	4	119	11	128	0.41 [0.13, 1.26]
Ferguson & Simes, 1949	6	300	29	274	0.20 [0.09, 0.49]
Rosenthal et al, 1960	3	228	11	209	0.26 [0.07, 0.92]
Hart & Sutherland, 1977	62	13536	248	12619	0.24 [0.18, 0.31]
Frimodt-Moller et al, 1973	33	5036	47	5761	0.80 [0.52, 1.25]
Stein & Aronson, 1953	180	1361	372	1079	0.46 [0.39, 0.54]
Vandiviere et al, 1973	8	2537	10	619	0.20 [0.08, 0.50]
TPT Madras, 1980	505	87886	499	87892	1.01 [0.89, 1.14]
Coetzee & Berjak, 1968	29	7470	45	7232	0.63 [0.39, 1.00]
Rosenthal et al, 1961	17	1699	65	1600	0.25 [0.15, 0.43]
Comstock et al, 1974	186	50448	141	27197	0.71 [0.57, 0.89]
Comstock & Webster, 1969	5	2493	3	2338	1.56 [0.37, 6.53]
Comstock et al, 1976	27	16886	29	17825	0.98 [0.58, 1.66]
Random-Effects Model					0.49 [0.34, 0.70]

Test of the overall effect: Z=-3.97, p<0.001
Heterogeneity: Q=152.23, df=12, p<0.001; $\hat{\tau}^2$=0.31, I^2=92%

1/32 1/16 1/8 1/4 1/2 1 2 4 8
Risk Ratio (log scale)

Example: BCG Vaccine Meta-Analysis [9]



Author(s) and Year	Vaccinated TB+	Vaccinated TB-	Control TB+	Control TB-	Risk Ratio [95% CI]
Aronson, 1948	4	119	11	128	0.41 [0.13, 1.26]
Ferguson & Simes, 1949	6	300	29	274	0.20 [0.09, 0.49]
Rosenthal et al, 1960	3	228	11	209	0.26 [0.07, 0.92]
Hart & Sutherland, 1977	62	13536	248	12619	0.24 [0.18, 0.31]
Frimodt-Moller et al, 1973	33	5036	47	5761	0.80 [0.52, 1.25]
Stein & Aronson, 1953	180	1361	372	1079	0.46 [0.39, 0.54]
Vandiviere et al, 1973	8	2537	10	619	0.20 [0.08, 0.50]
TPT Madras, 1980	505	87886	499	87892	1.01 [0.89, 1.14]
Coetzee & Berjak, 1968	29	7470	45	7232	0.63 [0.39, 1.00]
Rosenthal et al, 1961	17	1699	65	1600	0.25 [0.15, 0.43]
Comstock et al, 1974	186	50448	141	27197	0.71 [0.57, 0.89]
Comstock & Webster, 1969	5	2493	3	2338	1.56 [0.37, 6.53]
Comstock et al, 1976	27	16886	29	17825	0.98 [0.58, 1.66]
Random-Effects Model					0.49 [0.34, 0.70]

Test of the overall effect: Z=-3.97, p<0.001
Heterogeneity: Q=152.23, df=12, p<0.001; $\hat{\tau}^2$=0.31, I^2=92%

1/32 1/16 1/8 1/4 1/2 1 2 4 8
Risk Ratio (log scale)

Reporting/Quantifying Heterogeneity

- **Q-test:** only indicates whether we can reject H_0: $\tau^2 = 0$
- **I^2 statistic:** indicates (in %) how much of the total variability in the observed effects is due to heterogeneity, but it is not an absolute measure of heterogeneity [10]
- **estimate of τ^2:** is an absolute measure but hard to interpret
- **95% prediction interval:** indicates where most of the true effects of similar/new studies are expected to fall [11,12]
  - directly reflects the degree of heterogeneity (in effect size units)
  - basic computation:

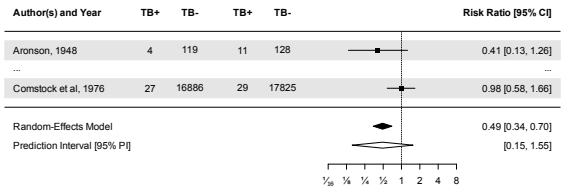
$$\hat{\mu} \pm 1.96\hat{\tau}$$

  - various refinements to account for uncertainty in $\hat{\mu}$ and $\hat{\tau}^2$
  - recommended as the best indicator of heterogeneity [10,13]
  - 95% PI for the BCG vaccine meta-analysis: 0.15 to 1.55

Visualizing Heterogeneity

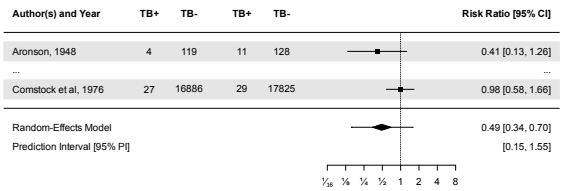
- want to **show the prediction interval** visually in a forest plot
- **various proposals** for this have been made [11,12,14–16]:
  - diamond shape
  - horizontal line through the diamond
  - horizontal bar under the diamond
- two (somewhat) **novel approaches**:
  - horizontal bar with shading
  - plot of the predictive distribution

Diamond Shape for the Prediction Interval



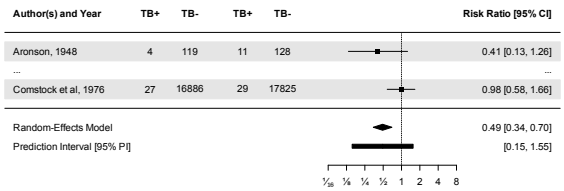
- suggested by Higgins et al. (2009) [11]
- sometimes the PI diamond is distinguished by color (see above)
- **problem:** easy to confuse the PI with the CI for μ

Horizontal Line Through the Diamond



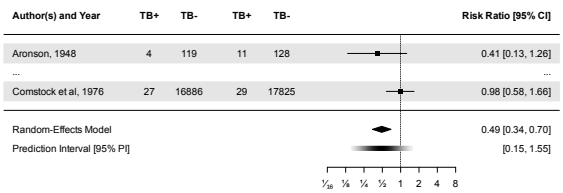
- suggested by Borenstein et al. (2009) [14] (and others [12,15])
- **problems:**
  - easy to confuse the PI with the CIs for the individual effect sizes
  - PI bounds in the annotations are in a different line than the horizontal line for the PI

Horizontal Bar under the Diamond



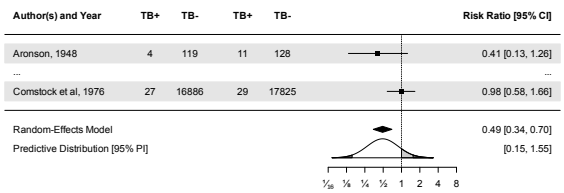
- suggested by Guddat et al. (2012) [16]
- PI uses a clearly distinguishable shape
- **problem:** not every value in the interval is equally likely

Horizontal Bar with Shading Under the Diamond



- not suggested for PIs before, but based on Jackson (2008) [17]
- shading indicates the density of the predictive distribution
- **problems:**
  - might not be clear what the shading indicates
  - need a legend to be able to interpret the shades

Plot of the Predictive Distribution

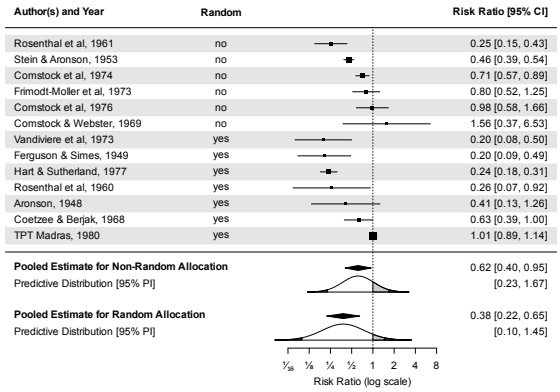


- clarifies that we are estimating an entire distribution of effects
- (optional) shading of regions:
  - **darker gray:** tail regions correspond to the 95% PI
  - **lighter gray:** area on the opposite side of the pooled effect (corresponds to $P(\theta > 1) = 0.11$ in this case)

Other Applications / Models

- illustrate **differences in heterogeneity** across:
  - subgroups (subgrouping / meta-regression models)
  - values of a continuous predictor (location-scale models)
  - different outcomes (in multivariate models)
  - ...

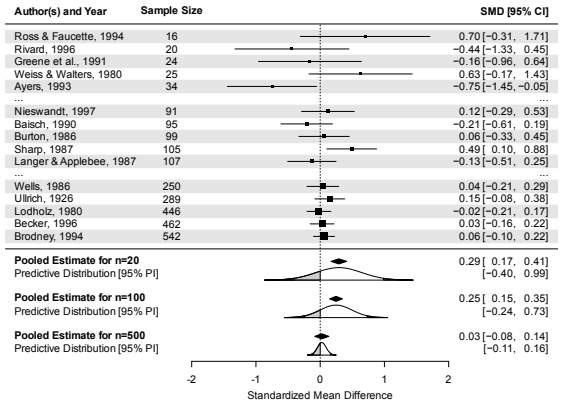
Subgrouping



Location-Scale Models

- Bangert-Drowns et al. (2004) [18] meta-analyzed $k = 48$ studies examining the effectiveness of an intervention for improving educational achievement
- effect sizes are given as standardized mean differences
- studies differed in their size (from $n = 16$ to $n = 542$)
- can use a **location-scale model** [19] to examine if the size of the pooled effect and the amount of heterogeneity differs for smaller versus larger studies

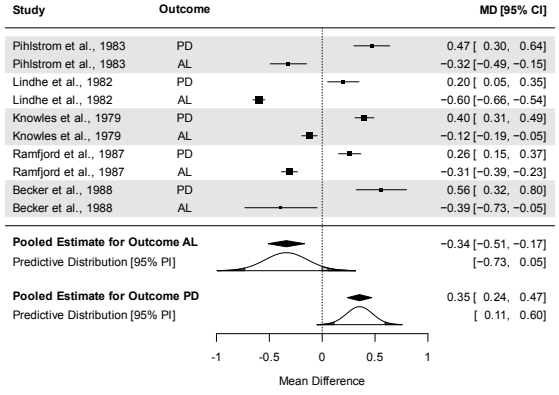
Location-Scale Models



Multivariate Models

- studies often report multiple outcomes
- can use a **multivariate model** to analyze multiple outcomes simultaneously [20–22]
- Berkey et al. (1998) [21] examined the effectiveness of surgical versus non-surgical treatment for periodontal disease for outcomes 'probing depth' (PD) and 'attachment level' (AL)

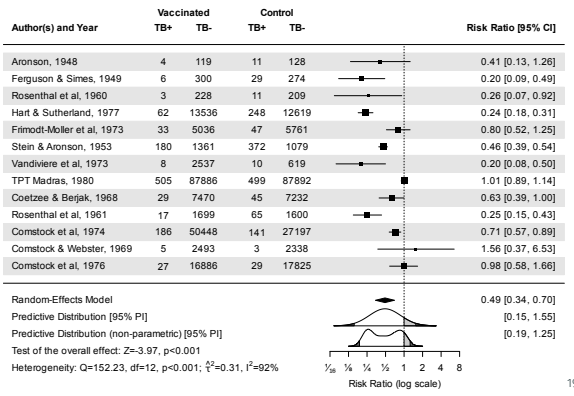
Multivariate Models



Non-Normal True Effects

- distribution of true effects may not be normal
- can either assume different distributions [23–27] or estimate the distribution non-parametrically [28]
- can use the **same visualization** when using these methods
- note: other visualizations hide the shape of the distribution (except when using shading, but this is difficult to process)

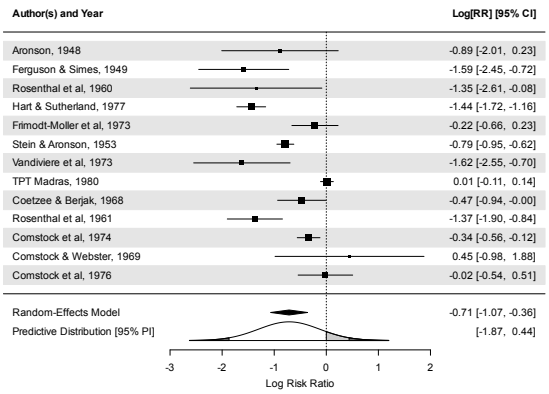
Example: BCG Vaccine Meta-Analysis



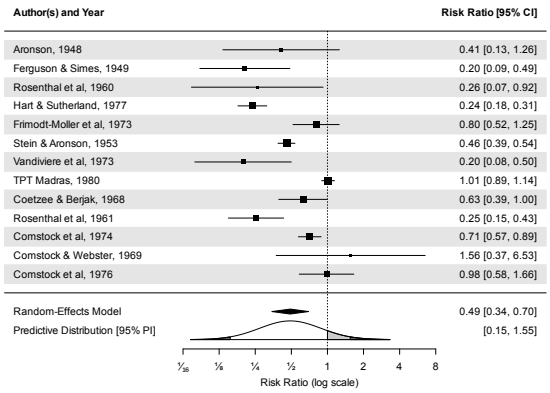
Back-Transformation

- often a meta-analysis is conducted on a **transformed scale** (e.g., log risk/odds ratios, r-to-z transformed correlation coefficients)
- for easier interpretation, the results from the meta-analysis are **back-transformed** (e.g., exponentiation, z-to-r transformation)
- in the forest plot, we can:
  - show the results on the transformed scale
  - use an axis transformation
  - directly back-transform all values
- when directly back-transforming all values, the predictive distribution needs to be transformed accordingly
- this implies a **non-normal distribution** for the true effects on the back-transformed scale (e.g., a log-normal distribution)

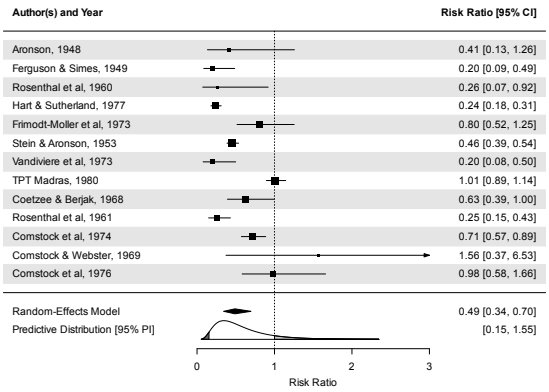
BCG Example: Show Results on the Transformed Scale



BCG Example: Use an Axis Transformation



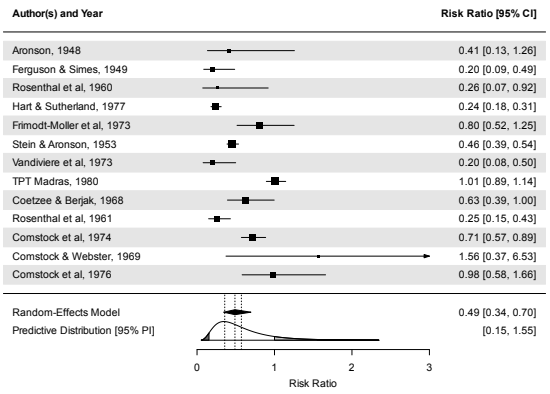
BCG Example: Directly Back-Transform All Values



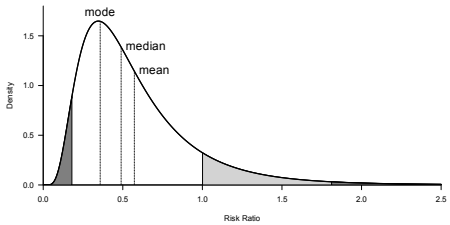
Back-Transformed Estimates

- due to Jensen's inequality, the **back-transformed mean effect** is not an estimate of the mean effect on the back-transformed scale (it is an estimate of the **median effect**)
- can also estimate the **mean** of the back-transformed predictive distribution and its **mode**

BCG Example: Directly Back-Transform All Values



BCG Example: Back-Transformed Predictive Distribution



- mode: 0.36 (95% CI: 0.25 – 0.51)
- median: 0.49 (95% CI: 0.34 – 0.70)
- mean: 0.57 (95% CI: 0.40 – 0.81)
- 95% prediction interval: 0.18 – 1.81
- $P(\theta > 1) = 0.11$

Conclusions

- predictive distribution is **visually distinct** from the polygon
- clarifies that we are estimating an **entire distribution** of effects
- visualizes the **different densities** under the distribution
- can use **shading** to emphasize different regions of interest
- can also be used for **non-normal** predictive distributions
- easily possible with the `metafor` **package** in R
- **code** corresponding to the examples given on my website

References [1]

1. Lewis, S., & Clarke, M. (2001). Forest plots: Trying to see the wood and the trees. *British Medical Journal*, *322*(7300), 1479–1480. https://doi.org/10.1136/bmj.322.7300.1479
2. Schild, A. H., & Voracek, M. (2013). Less is less: A systematic review of graph use in meta-analyses. *Research Synthesis Methods*, *4*(3), 209–219. https://doi.org/10.1002/jrsm.1076
3. Anzures-Cabrera, J., & Higgins, J. P. T. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods*, *1*(1), 66–80. https://doi.org/10.1002/jrsm.6
4. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*, n71. https://doi.org/10.1136/bmj.n71
5. Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., & Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association*, *283*(15), 2008–2012. https://doi.org/10.1001/jama.283.15.2008

References [2]

6. Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191
7. Cumpston, M., Lasserson, T., Flemyng, E., & J., P. M. (2024). Chapter III: Reporting the review [last updated august 2023]. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions (version 6.5)*. https://cochrane.org/handbook
8. Wilson, D. B., Pigott, T., Welch, V., Stewart, G., Hennessy Emily, A., & Dewidar, O. (2023). Methodological expectations of campbell collaboration intervention reviews (MECCIR): 2023 update. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/KCSPX
9. Colditz, G. A., Brewer, T. F., Berkey, C. S., Wilson, M. E., Burdick, E., Fineberg, H. V., & Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: Meta-analysis of the published literature. *Journal of the American Medical Association*, *271*(9), 698–702. https://doi.org/10.1001/jama.1994.03510330076038

References [3]

10. Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230
11. Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society, Series A*, *172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x
12. 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
13. IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, *6*(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247
14. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
15. Viechtbauer, W. (2010). *metafor: A meta-analysis package for R* (Version 1.0.1). https://www.metafor-project.org/

References [4]

16. Guddat, C., Grouven, U., Bender, R., & Skipka, G. (2012). A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Systematic Reviews*, *1*(34). https://doi.org/10.1186/2046-4053-1-34
17. Jackson, C. H. (2008). Displaying uncertainty with shading. *The American Statistician*, *62*(4), 340–347. https://doi.org/10.1198/000313008x370843
18. Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58. https://doi.org/10.3102/00346543074001029
19. Viechtbauer, W., & López-López, J. A. (2022). Location-scale models for meta-analysis. *Research Synthesis Methods*, *13*(6), 697–715. https://doi.org/10.1002/jrsm.1562
20. Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, *1*(3), 227–235. https://doi.org/10.1037/1082-989X.1.3.227
21. Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, *17*(22), 2537–2550. https://doi.org/10.1002/(sici)1097-0258(19981130)17:22%3C2537::aid-sim953%3E3.0.co;2-c
22. Mavridis, D., & Salanti, G. (2013). A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research*, *22*(2), 133–158. https://doi.org/10.1177/0962280211432219

References [5]

23. Lee, K. J., & Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, *27*(3), 418–434.
24. Baker, R., & Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science*, *11*(2), 121–131.
25. Beath, K. J. (2014). A finite mixture method for outlier detection and robustness in meta-analysis. *Research Synthesis Methods*, *5*(4), 285–293. https://doi.org/10.1002/jrsm.1114
26. Baker, R., & Jackson, D. (2016). New models for describing outliers in meta-analysis. *Research Synthesis Methods*, *7*(3), 314–328. https://doi.org/10.1002/jrsm.1191
27. Noma, H., Nagashima, K., Kato, S., Teramukai, S., & Furukawa, T. A. (2022). Meta-analysis using flexible random-effects distribution models. *Journal of Epidemiology*, *32*(10), 441–448. https://doi.org/10.2188/jea.JE20200376
28. Wang, C. C., & Lee, W. C. (2019). A simple method to estimate prediction intervals and predictive distributions: Summarizing meta-analyses beyond means and confidence intervals. *Research Synthesis Methods*, *10*(2), 255–266. https://doi.org/10.1002/jrsm.1345

Thank You for Your Attention!

Questions, Comments, Suggestions?

wolfgang.viechtbauer@maastrichtuniversity.nl

https://www.wvbauer.com

https://www.metafor-project.org

@wviechtb @wviechtb