Statistical methodology for social science meta-analysis: Outstanding needs and directions

Society for Research Synthesis Methodology
2026 Annual Meeting

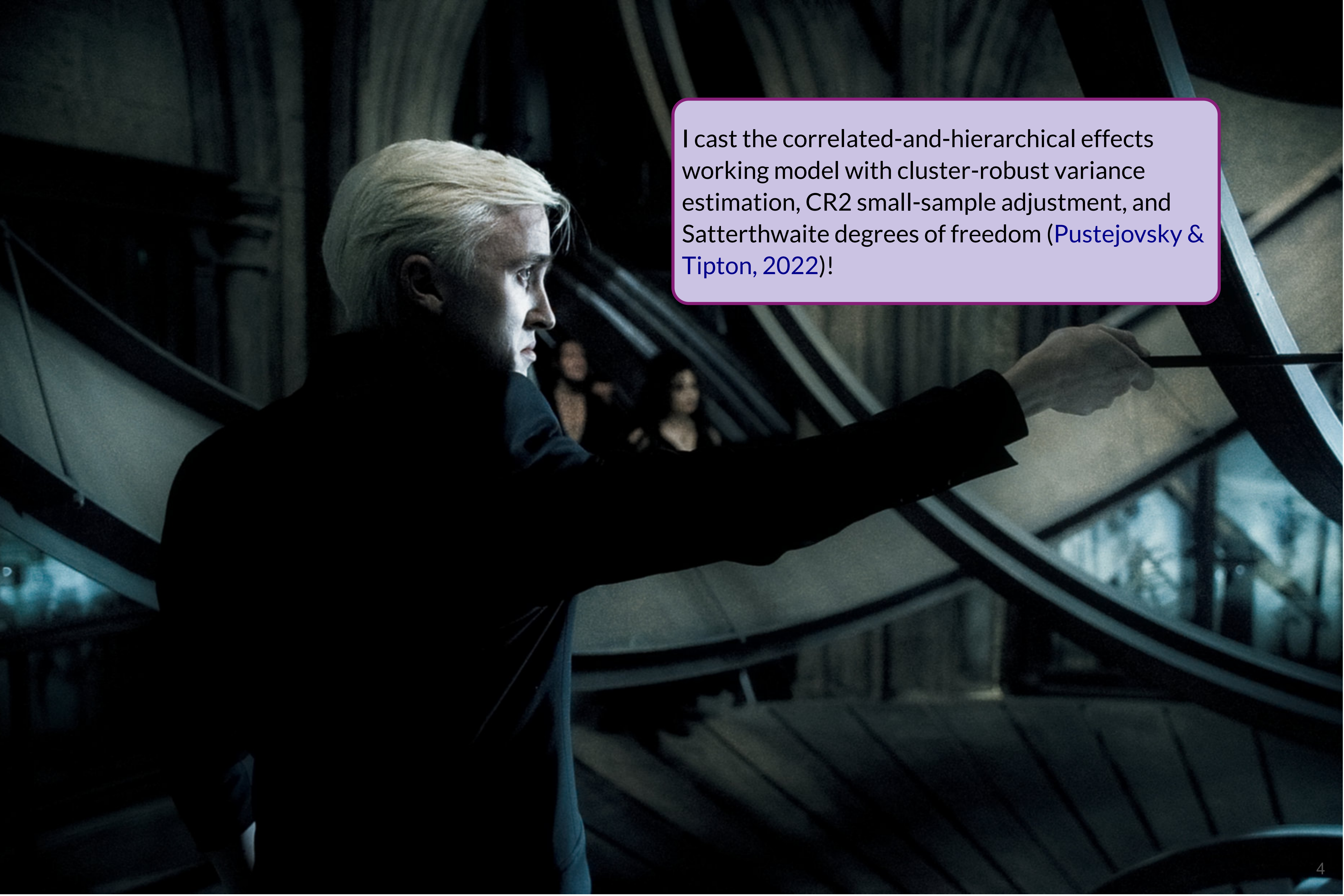
James E. Pustejovsky
University of Wisconsin - Madison

June 10, 2026

For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt.... All in all, I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of statistics which apply to analyzing data (Tukey, 1962).

Trend 1: More complex models for more complex data

- Hierarchical meta-analysis (Konstantopoulos, 2011; Van den Noortgate et al., 2013)
- Correlated-and-hierarchical effects models (Pustejovsky & Tipton, 2022)
- Multivariate, multilevel meta-analysis (McShane & Böckenholt, 2018; Sera et al., 2019)
- Network meta-analysis (Higgins & Whitehead, 1996; Lu & Ades, 2006)
- Phylogenetic meta-analysis (Lajeunesse, 2009)
- Location-scale random effects models (Viechtbauer & López-López, 2022)
- Robust Bayesian model-averaged meta-analysis (Bartoš et al., 2023, 2026; Maier et al., 2023)



I cast the correlated-and-hierarchical effects working model with cluster-robust variance estimation, CR2 small-sample adjustment, and Satterthwaite degrees of freedom (Pustejovsky & Tipton, 2022)!


I cast Robust Bayesian Model-Averaged Meta-Analysis with Selection Models (Bartoš et al., 2023)!

Preliminary data analysis methods

- Initial data analysis as a formal process (Baillie et al., 2022; Heinze et al., 2024; Huebner et al., 2018; Lusa et al., 2024).

PReliminary Investigation of MEta-analytic Databases (PRIMED, Pustejovsky et al., 2026)

Describe the amount of data and its dependence structure

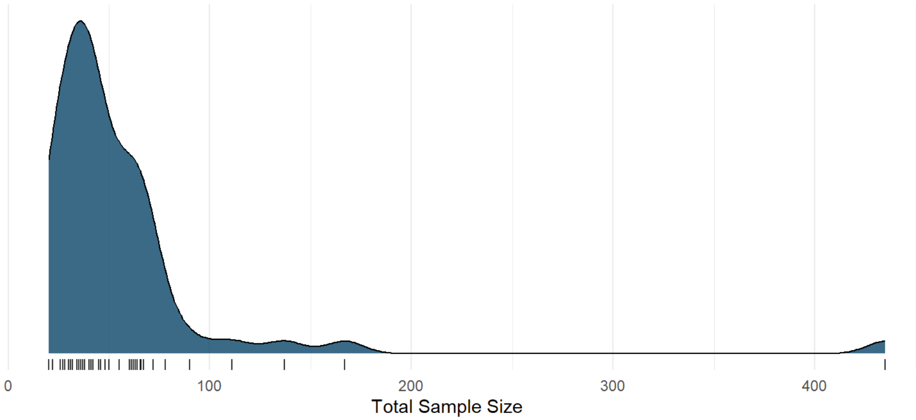
Explore study characteristics and potential moderators

Inspect standard errors and other auxiliary data

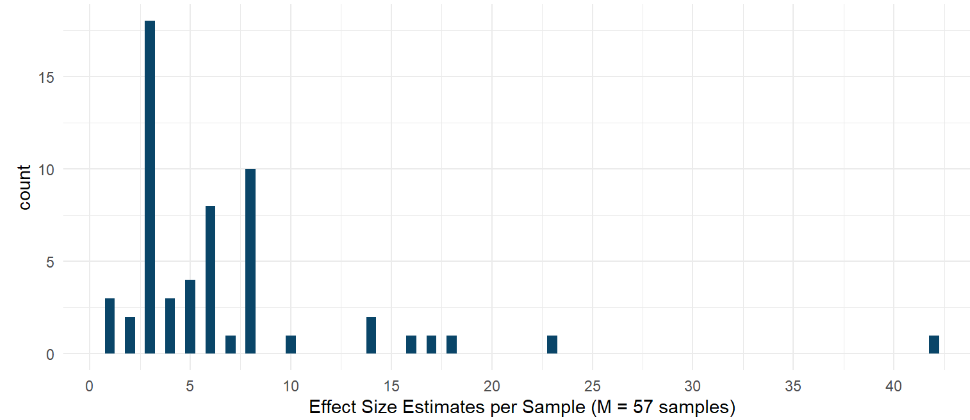
Visualize the distribution of effect size estimates

Describe the amount of data and its dependence structure

Sample size distribution



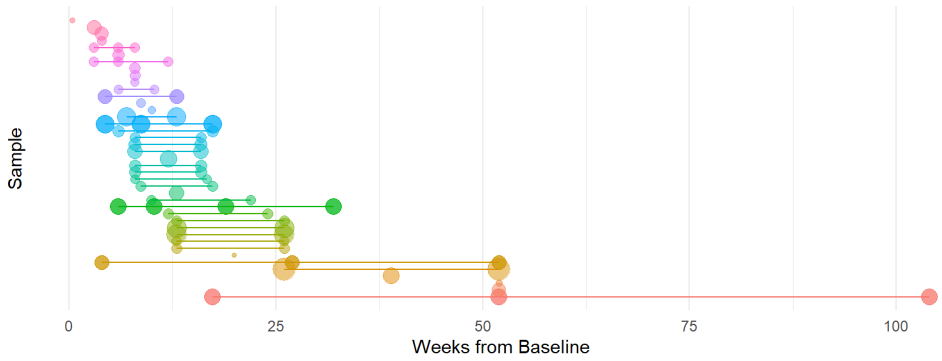
Effect size estimates per sample



Source: Winters et al. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis.

Explore study characteristics and potential moderators

Follow-up times vary within and between studies



Source: McLouth et al. (2021). A systematic review and meta-analysis of effects of psychosocial interventions on spiritual well-being in adults with cancer.

Within-study co-occurrence of assessment types

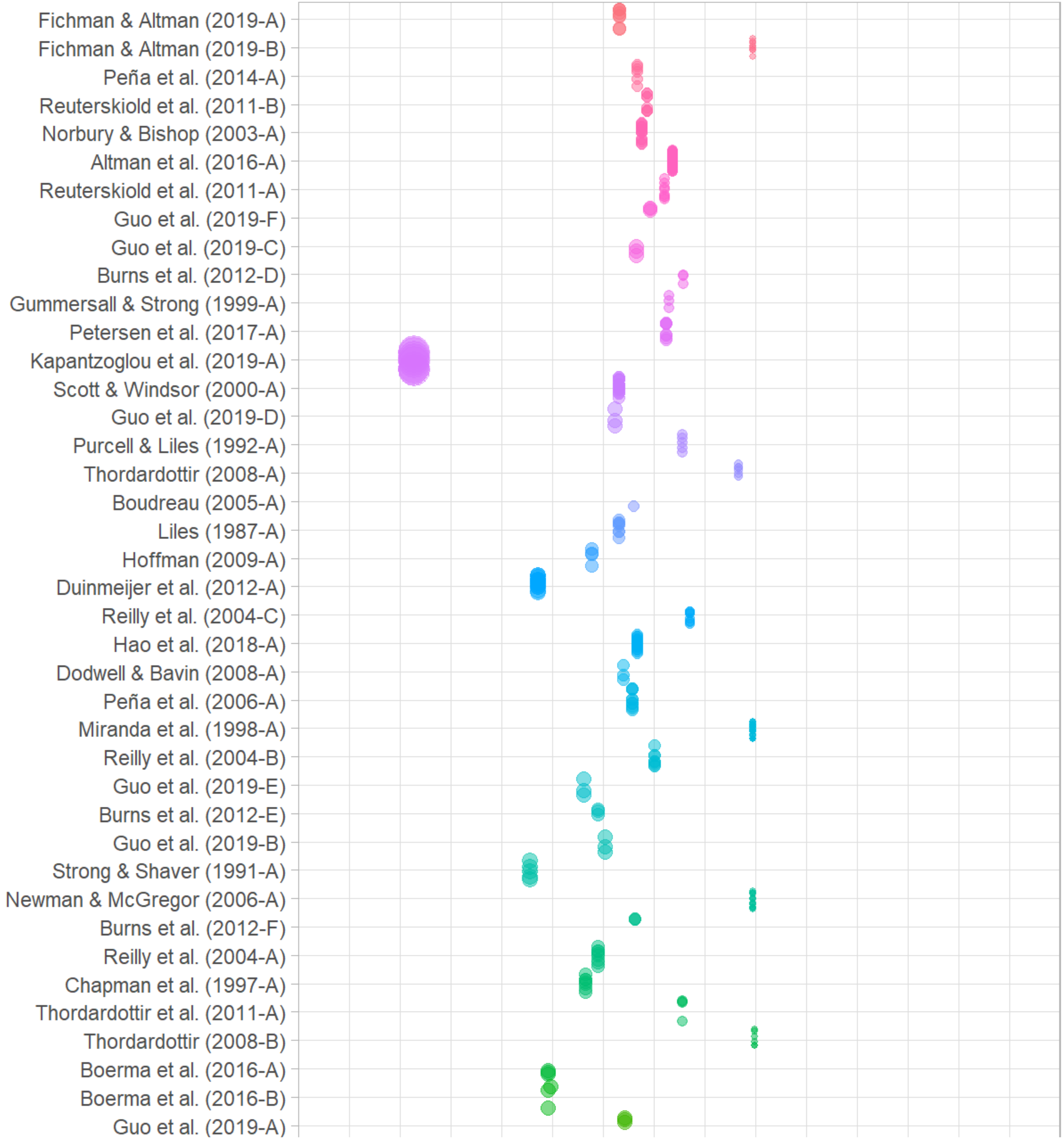
Assessment type	(A)	(B)	(C)	(D)
(A) ISL	3 (4, 20)	3 (4, 20)	2 (2, 18)	
(B) Macro	3 (4, 13)	26 (39, 111)	22 (33, 87)	4 (11, 11)
(C) Micro	2 (2, 31)	22 (33, 153)	31 (49, 235)	4 (11, 15)
(D) Mixed		4 (11, 14)	4 (11, 14)	6 (13, 16)

Values outside parentheses indicate the number of studies;

Values in parentheses indicate the number of samples and number of effect size estimates.

Source: Winters et al. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis.

Inspect standard errors and other auxiliary data



Fichman & Altman (2019-A)
Fichman & Altman (2019-B)
Peña et al. (2014-A)
Reuterskiold et al. (2011-B)
Norbury & Bishop (2003-A)
Altman et al. (2016-A)
Reuterskiold et al. (2011-A)
Guo et al. (2019-F)
Guo et al. (2019-C)
Burns et al. (2012-D)
Gummersall & Strong (1999-A)
Petersen et al. (2017-A)
Kapantzoglou et al. (2019-A)
Scott & Windsor (2000-A)
Guo et al. (2019-D)
Purcell & Liles (1992-A)
Thordardottir (2008-A)
Boudreau (2005-A)
Liles (1987-A)
Hoffman (2009-A)
Duinmeijer et al. (2012-A)
Reilly et al. (2004-C)
Hao et al. (2018-A)
Dodwell & Bavin (2008-A)
Peña et al. (2006-A)
Miranda et al. (1998-A)
Reilly et al. (2004-B)
Guo et al. (2019-E)
Burns et al. (2012-E)
Guo et al. (2019-B)
Strong & Shaver (1991-A)
Newman & McGregor (2006-A)
Burns et al. (2012-F)
Reilly et al. (2004-A)
Chapman et al. (1997-A)
Thordardottir et al. (2011-A)
Thordardottir (2008-B)
Boerma et al. (2016-A)
Boerma et al. (2016-B)
Guo et al. (2019-A)

Source: Winters et al. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis.

Visualize the distribution of effect size estimates

Hierarchical forest plot

Source: Winters et al. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis.

Trend 2: Method Proliferation

Many available tools for investigating selective reporting

- Tests/adjustments for funnel plot asymmetry
  - Trim-and-fill (Duval & Tweedie, 2000)
  - Limit meta-analysis (Rücker et al., 2011)
  - Egger's regression (Egger et al., 1997)
  - PET/PEESE (Stanley & Doucouliagos, 2014)
  - Skewness of standardized deviates (Lin & Chu, 2018)
  - LFK index (Furuya-Kanamori et al., 2018)
  - Kinked meta-regression (Bom & Rachinger, 2019)
  - Weighted average of adequately powered studies (WAAP, Stanley et al., 2017)
  - Weighted-and-iterated least squares (WILS, Stanley & Doucouliagos, 2024)

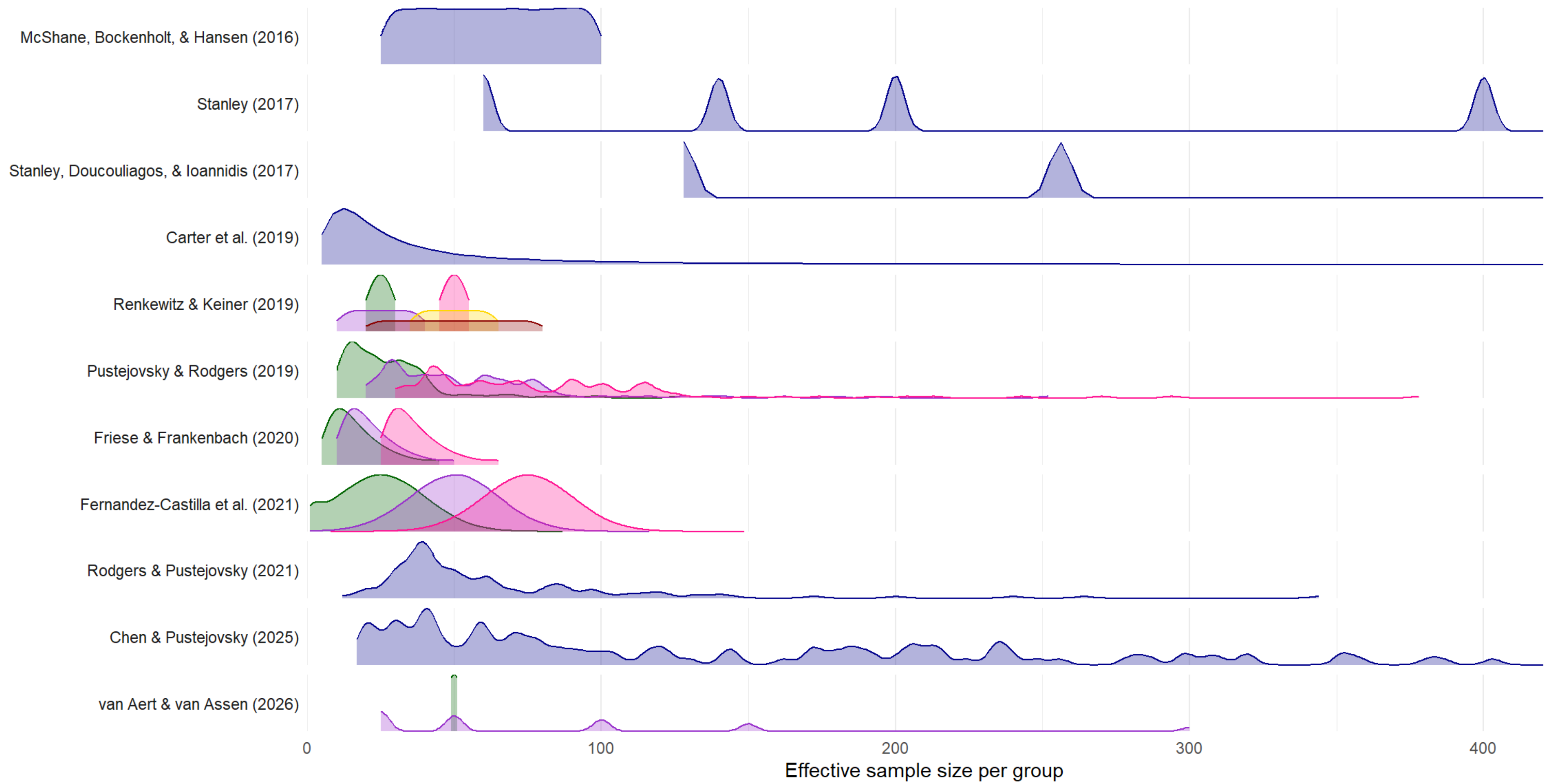
- Selection models
  - Weight-function models (Vevea & Hedges, 1995)
  - Copas models (Copas & Shi, 2000)
  - Worst-case sensitivity analysis (Mathur & VanderWeele, 2020)
- p-value diagnostics
  - Test of Excess Significance (Ioannidis & Trikalinos, 2007)
  - p-curve (Simonsohn et al., 2014)
  - p-uniform (van Assen et al., 2015)
  - $p\text{-uniform}^*$ (van Aert & van Assen, 2026)
  - Z-curve (Bartoš & Schimmack, 2022; Brunner & Schimmack, 2020)
- Model Ensembles
  - Robust Bayesian Meta-Analysis (Bartoš et al., 2023, 2026; Maier et al., 2023)

Can simulation save us?

- Hong & Reed (2021) propose that analysts should
  - Identify conditions within past simulations that most closely resemble their data
  - Only use methods that perform well under those conditions
- This would require:
  - Knowledge about which features of your data are important
  - Past simulations that examined similar conditions and all relevant methods
  - Method performance does not depend strongly on unknown parameters

Sample size distributions

In simulation studies on publication bias in meta-analysis of continuous outcomes



What is needed?

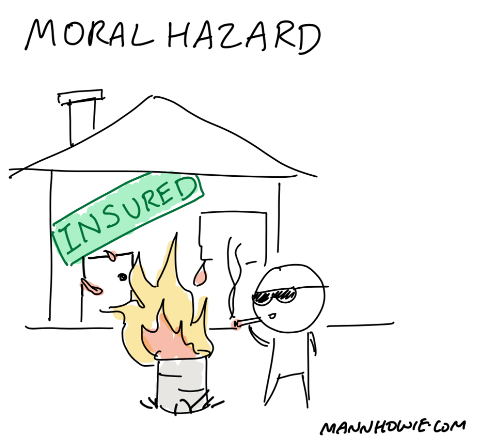
- Methodology that provides contra-indications and diagnostics for discontinuing use of a method.

- Greater emphasis on **model fit** assessment in real data analysis.

- Better simulations
  - Living synthetic benchmarks (Bartoš et al., 2025) that seperate work on specifying benchmarking conditions from work on method/guidance development.
  - Calibrated simulations, specific to one application (a.k.a. plasmode simulations, Franklin et al., 2014; Schreck et al., 2024)

Trend 3: Questing for robustness

- Fixed effects weighting for robustness to publication bias (Henmi & Copas, 2010)
  - Multiplicative heterogeneity models (Mawdsley et al., 2017; Thompson & Sharp, 1999)
  - Unrestricted weighted least squares (Stanley & Doucouliagos, 2015)
- Robust variance estimation
  - for heteroskedasticity (Sidik & Jonkman, 2006)
  - for dependent effects (Hedges et al., 2010)
  - small-sample corrections (Joshi et al., 2022; Tipton, 2015; Tipton & Pustejovsky, 2015)
- Pseudo-likelihood methods for multivariate meta-analysis (Chen et al., 2015)



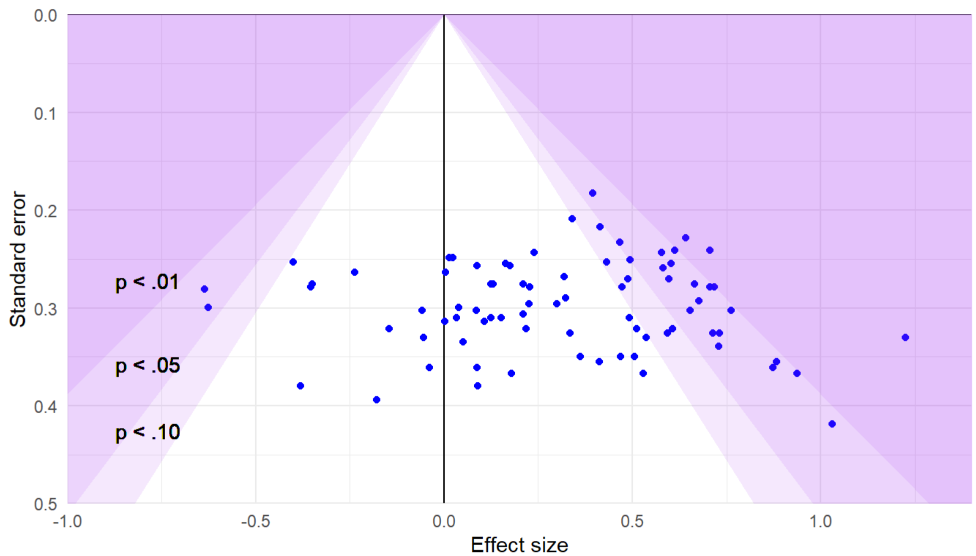
Robust methods shift focus away from developing models that we would actually put stock in.

What is needed?

Generative model

A model that provides a (relatively) more extensive description of a data-generating process, defined by parameters that have a (relatively) stable interpretations across contexts.

Publication bias



Random effects model

Param	Est	95% CI	
μ	0.30	0.23	0.38
τ	0.19	0.08	0.30

Regression adjustment (heuristic)

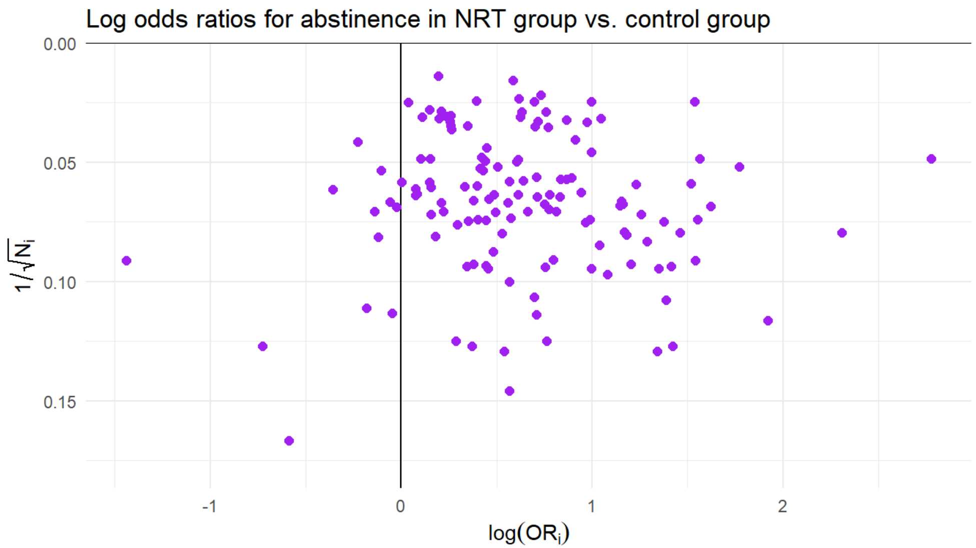
Param	Est	95% CI	
β_0	0.25	-0.25	0.75
β_1	0.19	-1.53	1.90
τ	0.19	0.09	0.31

Step-function selection model (generative)

Param	Est	95% CI	
μ	0.05	-0.13	0.23
τ	0.16	0.00	0.29
λ_1 ($p >$ .025)	0.22	0.02	0.43
λ_2 ($p >$ .500)	0.07	0.00	0.17
$k_{missing}$	322	61	1258

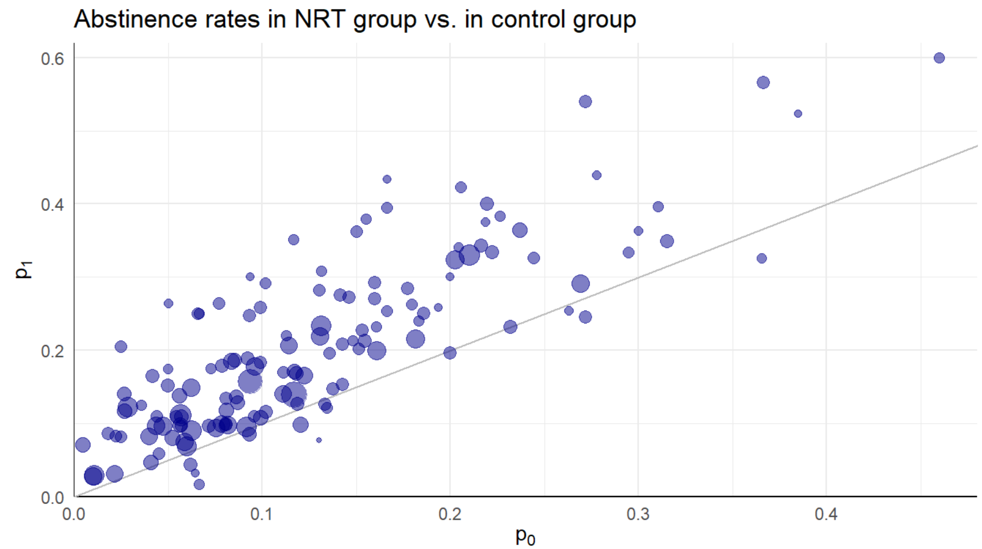
Meta-analysis of binary outcomes

Random effects model of log-odds ratios



Param	Est	95% CI		80% PI	
μ_{LOR}	0.56	0.49	0.63	0.26	0.87
OR	1.75	1.63	1.88	1.29	2.38
τ	0.24	0.15	0.35		

Bivariate model of arm-specific risks



Param	Est	95% CI		80% PI		τ
$\mathrm{logit}(p_0)$	-2.05	-2.18	-1.92	-2.97	-1.12	0.
μ_{LOR}	0.56	0.49	0.63	0.26	0.85	0.
OR	1.75	1.63	1.87	1.30	2.34	

- How well this method works depends on the distribution of π_{0i} (Pateras et al., 2018).

Meta-analysis vs. computational modeling

- Pillny et al. (2026) synthesized research on **effort-based decision-making** (EBDM) tasks to understand differences in motivation across individuals with different psychological disorders.

Meta-analysis

- Meta-analysis of standardized mean differences comparing diagnostic groups versus health controls on EBDM tasks:
  - Dual choice paradigms
  - Effort-discounting paradigms
  - Progressive ratio paradigms

Computational models

- Narrative review of computational models that aim to understand component processes in decision-making:
  - Reward valuation
  - Effort sensitivity
  - Probability weighting

> While computational modeling may provide deeper insights into EBDM

> mechanisms, its parameters cannot be directly aggregated in classical meta-analyses due to variations in model structures, parameter definitions, and task designs. Unlike standardized effect sizes used in traditional meta-analyses, computational parameters are model-dependent and often require study-specific assumptions, making direct comparisons across studies challenging or unfeasible.

Aspire for more generative models

- More generative models…
  - Can be used for simulating data
  - Can be used to make predictions
  - Can be more thoroughly evaluated for fit / misfit
  - Can be more closely tied to scientific phenomena / substantive theory

Trends and Directions

Naive use of complex models

- Preliminary data analysis routines

Method proliferation

- Method contra-indications
- Better / more model fit assessment
- Living synthetic benchmarking studies
- Calibrated simulations

Questing for robustness

- Aspire towards more generative models

References

Baillie, M., Le Cessie, S., Schmidt, C. O., Lusa, L., Huebner, M., & for the Topic Group “Initial Data Analysis” of the STRATOS Initiative. (2022). Ten simple rules for initial data analysis. *PLOS Computational Biology*, *18*(2), e1009819. https://doi.org/10.1371/journal.pcbi.1009819

Bartoš, F., Maier, M., & Wagenmakers, E.-J. (2026). Robust Bayesian multilevel meta-analysis: Adjusting for publication bias in the presence of dependent effect sizes. *Behavior Research Methods*, *58*(6), 165. https://doi.org/10.3758/s13428-026-03023-y

Bartoš, F., Maier, M., Wagenmakers, E.-J., Doucouliagos, H., & Stanley, T. D. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, *14*(1), 99–116. https://doi.org/10.1002/jrsm.1594

Bartoš, F., Pawel, S., & Siepe, B. S. (2025, October 22). *Living Synthetic Benchmarks: A Neutral and Cumulative Framework for Simulation Studies*. arXiv.org. https://arxiv.org/abs/2510.19489v1

Bartoš, F., & Schimmack, U. (2022). Z-curve 2.0: Estimating replication rates and discovery rates. *Meta-Psychology*, *6*. https://doi.org/10.15626/MP.2021.2720

Bom, P. R. D., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research Synthesis Methods*, *10*(4), 497–514. https://doi.org/10.1002/jrsm.1352

Brunner, J., & Schimmack, U. (2020). Estimating Population Mean Power Under Conditions of Heterogeneity and Selection for Significance. *Meta-Psychology*, *4*. https://doi.org/10.15626/MP.2018.874

Chen, Y., Hong, C., & Riley, R. D. (2015). An alternative pseudolikelihood method for multivariate random-effects meta-analysis: An alternative pseudolikelihood method for multivariate random-effects meta-analysis. *Statistics in Medicine*, *34*(3), 361–380. https://doi.org/10.1002/sim.6350

Copas, J. B., & Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, *1*(3), 247–262. https://doi.org/10.1093/biostatistics/1.3.247

Duval, S., & Tweedie, R. (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/10.2307/2669529

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634.

Franklin, J. M., Schneeweiss, S., Polinski, J. M., & Rassen, J. A. (2014). Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. *Computational Statistics & Data Analysis*, *72*, 219–226. https://doi.org/10.1016/j.csda.2013.10.018

Furuya-Kanamori, L., Barendregt, J. J., & Doi, S. A. R. (2018). A new improved graphical and quantitative method for detecting bias in meta-analysis. *JBI Evidence Implementation*, *16*(4), 195. https://doi.org/10.1097/XEB.0000000000000141

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Heinze, G., Baillie, M., Lusa, L., Sauerbrei, W., Schmidt, C. O., Harrell, F. E., & Huebner, M. (2024). Regression without regrets –initial data analysis is a prerequisite for multivariable regression. *BMC Medical Research Methodology*, *24*(1, 1), 1–17. https://doi.org/10.1186/s12874-024-02294-3

Henmi, M., & Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine*, *29*(29), 2969–2983.

Higgins, J. P. T., & Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, *15*(24), 2733–2749. https://doi.org/10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0

Hong, S., & Reed, W. R. (2021). Using Monte Carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, *12*(2), 192–215. https://doi.org/10.1002/jrsm.1467

Huebner, M., Le Cessie, S., Schmidt, C. O., & Vach, W. (2018). A contemporary conceptual framework for initial data analysis. *Observational Studies*, *4*(1), 171–192. https://doi.org/10.1353/obs.2018.0014

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

Joshi, M., Pustejovsky, J. E., & Beretvas, S. N. (2022). Cluster Wild Bootstrapping to Handle Dependent Effect Sizes in Meta-Analysis with a Small Number of Studies. *Research Synthesis Methods*, *13*(4), 457–477. https://doi.org/10.1002/jrsm.1554

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis:

Three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. https://doi.org/10.1002/jrsm.35

Lajeunesse, M. J. (2009). Meta-analysis and the comparative phylogenetic method. *The American Naturalist*, *174*(3), 369–381. https://doi.org/10.1086/603628

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, *74*(3), 785–794. https://doi.org/10.1111/biom.12817

Lu, G., & Ades, A. E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, *101*(474), 447–459. https://doi.org/10.1198/016214505000001302

Lusa, L., Proust-Lima, C., Schmidt, C. O., Lee, K. J., Cessie, S. le, Baillie, M., Lawrence, F., Huebner, M., & Initiative, on behalf of T. of the S. (2024). Initial data analysis for longitudinal studies to build a solid foundation for reproducible analysis. *PLOS ONE*, *19*(5), e0295726. https://doi.org/10.1371/journal.pone.0295726

Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, *28*(1), 107–122. https://doi.org/10.1037/met0000405

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

Mawdsley, D., Higgins, J. P. T., Sutton, A. J., & Abrams, K. R. (2017). Accounting for heterogeneity in meta-analysis using a multiplicative model—an empirical study. *Research Synthesis Methods*, *8*(1), 43–52. https://doi.org/10.1002/jrsm.1216

McLouth, L. E., Ford, C. G., Pustejovsky, J. E., Park, C. L., Sherman, A. C., Trevino, K., & Salsman, J. M. (2021). A systematic review and meta-analysis of effects of psychosocial interventions on spiritual well-being in adults with cancer. *Psycho-Oncology*, *30*(2), 147–158. https://doi.org/10.1002/pon.5562

McShane, B. B., & Böckenholt, U. (2018). Multilevel multivariate meta-analysis with application to choice overload. *Psychometrika*, *83*(1), 255–271. https://doi.org/10.1007/s11336-017-9571-z

Pateras, K., Nikolakopoulos, S., & Roes, K. (2018). Data-generating models of dichotomous outcomes: Heterogeneity in simulation studies for a random-effects meta-analysis. *Statistics in Medicine*, *37*(7), 1115–1124. https://doi.org/10.1002/sim.7569

Pillny, M., Renz, K. E., Hay, A., Fulford, D., Lincoln, T. M., Barch, D. M., Gold, J. M., & Kaiser, S. (2026). Effort-based decision making in psychopathology: A transdiagnostic multilevel meta-analysis and systematic review of behavioral patterns and mechanisms underlying amotivational psychopathology. *Psychological Bulletin*,

152(1), 33–65. https://doi.org/10.1037/bul0000510
Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3
Pustejovsky, J. E., Zhang, J., & Tipton, E. (2026). A preliminary data analysis workflow for meta-analysis of dependent effect sizes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *384*(2321), 20240604. https://doi.org/10.1098/rsta.2024.0604
Rücker, G., Schwarzer, G., Carpenter, J. R., Binder, H., & Schumacher, M. (2011). Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, *12*(1), 122–142. https://doi.org/10.1093/biostatistics/kxq046
Schreck, N., Slynko, A., Saadati, M., & Benner, A. (2024). Statistical plasmode simulations–Potentials, challenges and recommendations. *Statistics in Medicine*, *43*(9), 1804–1825. https://doi.org/10.1002/sim.10012
Sera, F., Armstrong, B., Blangiardo, M., & Gasparrini, A. (2019). An extended mixed-effects framework for meta-analysis. *Statistics in Medicine*, *38*(29), 5429–5444. https://doi.org/10.1002/sim.8362
Sidik, K., & Jonkman, J. N. (2006). Robust variance estimation for random effects meta-analysis. *Computational Statistics & Data Analysis*, *50*(12), 3681–3701. https://doi.org/10.1016/j.csda.2005.07.019
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. https://doi.org/10.1177/1745691614553988
Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/https://doi.org/10.1002/jrsm.1095
Stanley, T. D., & Doucouliagos, H. (2015). Neither fixed nor random: Weighted least squares meta-analysis. *Statistics in Medicine*, *34*(13), 2116–2127. https://doi.org/10.1002/sim.6481
Stanley, T. D., & Doucouliagos, H. (2024). Harnessing the power of excess statistical significance: Weighted and iterative least squares. *Psychological Methods*, *29*(2), 407–420. https://doi.org/10.1037/met0000502
Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. A. (2017). Finding the power to reduce publication bias. *Statistics in Medicine*, *36*(10), 1580–1598. https://doi.org/10.1002/sim.7228
Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, *18*(20), 2693–2708. https://doi.org/10.1002/

(SICI)1097-0258(19991030)18:20<2693::AID-SIM235>3.0.CO;2-V

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634. https://doi.org/10.3102/1076998615606099

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, *33*(1), 1–67. https://doi.org/10.1214/aoms/1177704711

van Aert, R. C. M., & van Assen, M. A. L. M. (2026). Correcting for publication bias in a meta-analysis with the p-uniform* method. *Psychonomic Bulletin & Review*, *33*(3), 102. https://doi.org/10.3758/s13423-025-02812-4

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*(3), 293–309. https://doi.org/http://dx.doi.org/10.1037/met0000025

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

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. https://doi.org/10.1007/BF02294384

Viechtbauer, W., & López-López, J. A. (2022). Location-scale models for meta-analysis. *Research Synthesis Methods*, jrsm.1562. https://doi.org/10.1002/jrsm.1562

Winters, K. L., Jasso, J., Pustejovsky, J. E., & Byrd, C. T. (2022). Investigating narrative performance in children with developmental language disorder: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3908–3929. https://doi.org/10.1044/2022_JSLHR-22-00017