Universiteit
Leiden
The Netherlands

ALL-IN meta-analysis

Schure, J.A. ter

Citation
Schure, J. A. ter. (2022, April 7). *ALL-IN meta-analysis*. Retrieved from https://hdl.handle.net/1887/3281933

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3281933

Note: To cite this publication please use the final published version (if applicable).

Bibliography

Akl, E. A., Meerpohl, J. J., Elliott, J., Kahale, L. A., Schünemann, H. J., Agoritsas, T., Hilton, J., Perron, C., Akl, E., Hodder, R., et al. (2017). Living systematic reviews: 4. living guideline recommendations. *Journal of clinical epidemiology*, 91:47–53.

Altman, D. G. (1994). The scandal of poor medical research. *BMJ*, 308(6924):283–284. Publisher: British Medical Journal Publishing Group Section: Editorial.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer-Verlag, New York.

Armitage, P. (1984). Controversies and achievements in clinical trials. *Contemporary Clinical Trials*, 5(1):67–72.

Baum, M. L., Anish, D. S., Chalmers, T. C., Sacks, H. S., Smith Jr, H., and Fagerstrom, R. M. (1981). A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *New England Journal of Medicine*, 305(14):795–799.

Berger, J. O. and Berry, D. A. (1988). The relevance of stopping rules in statistical inference. *Statistical decision theory and related topics IV*, 1:29–47.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. DOI: 10.1002/9780470743386.refs.

Borm, G. F. and Donders, A. R. T. (2009). Updating meta-analyses leads to larger type I errors than publication bias. *Journal of clinical epidemiology*, 62(8):825–830.

Branswell, H. (2021). 12 lessons Covid-19 taught us about developing vaccines during a pandemic. https://www.statnews.com/2021/06/30/12-lessons-covid-19-developing-vaccines/. Accessed: 12 July 2021.

Breiman, L. (1961). Optimal gambling systems for favorable games. *Fourth Berkeley Symposium*.

Brok, J., Thorlund, K., Wetterslev, J., and Gluud, C. (2008). Apparently conclusive meta-analyses may be inconclusive—trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. *International journal of epidemiology*, 38(1):287–298.

Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A. M., Howells, D. W., Ioannidis, J. P., and Oliver, S. (2014). How to increase value and reduce waste when research priorities are set. *The Lancet*, 383(9912):156–165.

Chalmers, I. and Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, 114(6):1341–1345.

Chalmers, I. and Glasziou, P. (2016). Systematic reviews and research waste. *The Lancet*, 387(10014):122–123.

Chalmers, T. C. and Lau, J. (1993). Meta-analytic stimulus for changes in clinical trials. *Statistical Methods in Medical Research*, 2(2):161–172.

Claxton, K. P., Sculpher, M., and Drummond, M. (2002). A rational framework for decision making by the national institute for clinical excellence (NICE). *The Lancet*, 360(9334):711–715.

Claxton, K. P. and Sculpher, M. J. (2006). Using value of information analysis to prioritise health research. *Pharmacoeconomics*, 24(11):1055–1068.

Cochrane Scientific Committee, Schmid, C., Senn, S., Sterne, J., Kulinskaya, E., Posch, M., Roes, K., and McKenzie, J. (2018). Should Cochrane apply error-adjustment methods when conducting repeated meta-analyses?

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, 34(2):187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Crane, H. and Shafer, G. (2020). Risk is random: The magic of the d'alembert. Technical report, Working Paper 57, Available from: https://www. proba bilit yandf inance. com.

CureVac AG (2020). Clinical trial protocol a phase 2b/3, randomized, observer-blinded, placebo-controlled, multicenter clinical study evaluating the efficacy and safety of investigational sars-cov-2 mrna vaccine cvncov in adults 18 years of age and older. https://www.curevac.com/wp-content/uploads/2021/06/HERALD_CV-NCOV-004-Protocol.pdf. Accessed: 16 July 2021.

CureVac AG (2021). Curevac final data from phase 2b/3 trial of first-generation Covid-19 vaccine candidate, cvncov, demonstrates protection in age group of 18 to 60. https://www.curevac.com/en/2021/06/30/curevac-final-data-from\-phase-2b-3-trial-of-first-generation-covid-19-vaccine\-candidate-cvncov-demonstrates-protection-in-age-group-of-18-to-60/. Accessed: 16 July 2021.

Darling, D. and Robbins, H. (1967). Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America*, 58(1):66.

Darling, D. and Robbins, H. (1968). Some nonparametric sequential tests with power

one. *Proceedings of the National Academy of Sciences of the United States of America*, 61(3):804.

Davey, J., Turner, R. M., Clarke, M. J., and Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the Cochrane database of systematic reviews: a cross-sectional, descriptive analysis. *BMC medical research methodology*, 11(1):160.

Dawid, A. P. (1984). Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A (General)*, 147(2):278–290.

Dawid, A. P. (1994). Selection paradoxes of Bayesian inference. *Lecture Notes-Monograph Series*, pages 211–220.

van Dongen, N. and van Grootel, L. (2021). Overview on the null hypothesis significance test. https://psyarxiv.com/hwk4n/download?format=pdf.

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B. A., and Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, 112(50):15343–15347.

Edwards, A. F. (1974). *Likelihood: An account of the statistical concept of likelihood and its application to scientific inference*. Cambridge University Press, New York.

Efron, B. (1974). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565. with discussion.

Egger, M. and Smith, G. D. (1998). Bias in location and selection of studies. *BMJ: British Medical Journal*, 316(7124):61.

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., et al. (2017). Living systematic review: 1. introduction—the why, what, when, and how. *Journal of clinical epidemiology*, 91:23–30.

Ellis, S. P. and Stewart, J. W. (2009). Temporal dependence and bias in meta-analysis. *Communications in Statistics—Theory and Methods*, 38(15):2453–2462.

Evans, I., Thornton, H., Chalmers, I., and Glasziou, P. (2011). *Testing treatments: better research for better healthcare*. Pinter & Martin Publishers.

Evbres (2021). Hans lund - EBR - placing research in the context of existing knowledge. https://www.youtube.com/watch?v=dko-35vvJFk&list=PLkIyhRK9IXIOzuqBwNMFl1gyyE4vvdi9u&index=3.

FDA (2020). Development and Licensure of Vaccines to Prevent Covid-19. https://www.fda.gov/media/139638/download. Accessed: 12 July 2021.

FDA (2021). Emergency Use Authorization for Vaccines to Prevent Covid-19. https://www.fda.gov/media/142749/download. Accessed: 12 July 2021.

Fergusson, D., Glass, K. C., Hutton, B., and Shapiro, S. (2005). Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clinical Trials*, 2(3):218–232.

Fisher, R. A. (1938). Presidential address. *Sankhyā: The Indian Journal of Statistics*, pages 14–17.

Gehr, B. T., Weiss, C., and Porzsolt, F. (2006). The fading of reported effectiveness. a meta-analysis of randomised controlled trials. *BMC medical research methodology*, 6(1):25.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2):198–218.

Glasziou, P. and Chalmers, I. (2018). Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. *BMJ*, 363. Publisher: British Medical Journal Publishing Group Section: Feature.

Glasziou, P., Sanders, S., and Hoffmann, T. (2020). Waste in Covid-19 research. *BMJ*, 369. Publisher: British Medical Journal Publishing Group Section: Editorial.

Gøtzsche, P. C. (1987). Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)*, 295(6599):654–656.

Goudie, A. C., Sutton, A. J., Jones, D. R., and Donald, A. (2010). Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. *Journal of Clinical Epidemiology*, 63:983–991.

Grünwald, P. (2021). Peter D. Grünwald's contribution to the Discussion of 'Testing by betting: A strategy for statistical and scientific communication' by Glenn Shafer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(2):440–441.

Grünwald, P., de Heide, R., and Koolen, W. (2019). Safe testing. *arXiv preprint arXiv:1906.07801*.

Grünwald, P. and Mehta, N. (2019). A tight excess risk bound via a unified PAC-Bayesian-Rademacher-Shtarkov-MDL complexity. In *Proceedings of the Thirtieth Conference on Algorithmic Learning Theory (ALT) 2019*.

Grünwald, P. and Roos, T. (2020). Minimum Description Length revisited. *International Journal of Mathematics for Industry*, 11(1).

de Heide, R. and Grünwald, P. D. (2021). Why optional stopping can be a problem for Bayesians. *Psychonomic Bulletin & Review*, 28(3):795–812.

Hendriksen, A., de Heide, R., and Grünwald, P. (2020). Optional stopping with Bayes factors: a categorization and extension of folklore results, with an application to invariant situations. *Bayesian Analysis*.

Henzi, A. and Ziegel, J. F. (2021). Valid sequential inference on probability forecast performance. *arXiv preprint arXiv:2103.08402*.

Higgins, J. P., Whitehead, A., and Simmonds, M. (2011). Sequential methods for random-effects meta-analysis. *Statistics in medicine*, 30(9):903–921.

Howard, S. R. and Ramdas, A. (2019). Sequential estimation of quantiles with applications to A/B-testing and best-arm identification. *arXiv preprint arXiv:1906.09712*.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2018). Exponential line-crossing inequalities. *arXiv:1808.03204 [math]*. arXiv: 1808.03204.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080.

Imberger, G., Thorlund, K., Gluud, C., and Wetterslev, J. (2016). False-positive findings in cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ open*, 6(8):e011890.

Ioannidis, J. P. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *Jama*, 294(2):218–228.

Ioannidis, J. P. (2005b). Why most published research findings are false. *PLoS medicine*, 2(8):e124.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, pages 640–648.

Ioannidis, J. P. (2010). Meta-research: The art of getting it wrong. *Research Synthesis Methods*, 1(3-4):169–184.

Ioannidis, J. P. and Trikalinos, T. A. (2005). Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *Journal of clinical epidemiology*, 58(6):543–549.

Jackson, D. and Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research synthesis methods*, 8(3):290–302.

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. (2021). Always valid inference: Continuous monitoring of A/B tests. *Operations Research*.

Kelly, J. (1956). A new interpretation of information rate. *Bell System Technical Journal*, pages 917–926.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media. Google-Books-ID: aO7xBwAAQBAJ.

Konnikova, M. (2020). *The Biggest Bluff: How I Learned to Pay Attention, Master Myself, and Win*. Penguin.

Krum, H. and Tonkin, A. (2003). Why do phase III trials of promising heart failure drugs often fail? the contribution of "regression to the truth". *Journal of cardiac failure*, 9(5):364–367.

Kulinskaya, E., Huggins, R., and Dogo, S. H. (2016). Sequential biases in accumulating evidence. *Research synthesis methods*, 7(3):294–305.

Kulinskaya, E. and Wood, J. (2014). Trial sequential methods for meta-analysis. *Research synthesis methods*, 5(3):212–220.

Lai, T. L. (1976). On confidence sequences. *The Annals of Statistics*, 4(2):265–280.

Lakens, D. (2021). Sample size justification. https://psyarxiv.com/9d3yf/download?format=pdf.

Lan, K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663.

Lau, J., Antman, E. M., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., and Chalmers, T. C. (1992). Cumulative meta-analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine*, 327(4):248–254.

Lau, J., Schmid, C. H., and Chalmers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of clinical epidemiology*, 48(1):45–57.

Lawrence, J. M., Meyerowitz-Katz, G., Heathers, J. A., Brown, N. J., and Sheldrick, K. A. (2021). The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable. *Nature Medicine*, pages 1–2.

Lehmann, E. (1986). *Testing statistical hypotheses*. Wiley.

Li, J. and Barron, A. (2000). Mixture density estimation. In Solla, S., Leen, T., and Müller, K.-R., editors, *Advances in Neural Information Processing Systems*, volume 12, pages 279–285, Cambridge, MA. MIT Press.

Li, Q. J. (1999). *Estimation of Mixture Models*. PhD Thesis, Yale University, New Haven, CT, USA.

Lund, H., Brunnhuber, K., Juhl, C., Robinson, K., Leenaars, M., Dorch, B. F., Jamtvedt, G., Nortvedt, M. W., Christensen, R., and Chalmers, I. (2016). Towards evidence based research. *Bmj*, 355:i5440.

Mallett, S. and Clarke, M. (2003). How many Cochrane reviews are needed to cover existing evidence on the effects of health care interventions? *ACP journal club*, 139(1):A11–A11.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.*, 50:163–170.

McDonald, A. M., Knight, R. C., Campbell, M. K., Entwistle, V. A., Grant, A. M., Cook, J. A., Elbourne, D. R., Francis, D., Garcia, J., Roberts, I., et al. (2006). What influences recruitment to randomised controlled trials? a review of trials funded by two UK funding agencies. *Trials*, 7(1):1–8.

McShane, B. B. and Gal, D. (2017). Statistical significance and the dichotomization of evidence. *Journal of the American Statistical Association*, 112(519):885–895.

Mehrotra, D. V. and Roth, A. J. (2001). Relative risk estimation and inference using a generalized logrank statistic. *Statistics in Medicine*, 20(14):2099–2113. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.854.

Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., and Altman, D. G. (2007a). Epidemiology and reporting characteristics of systematic reviews. *PLoS medicine*, 4(3):e78.

Moher, D. and Tsertsvadze, A. (2006). Systematic reviews: when is an update an update? *The Lancet*, 367(9514):881–883.

Moher, D., Tsertsvadze, A., Tricco, A., Eccles, M., Grimshaw, J., Sampson, M., and Barrowman, N. (2008). When and how to update systematic reviews. *Cochrane database of systematic reviews*, (1).

Moher, D., Tsertsvadze, A., Tricco, A. C., Eccles, M., Grimshaw, J., Sampson, M., and Barrowman, N. (2007b). A systematic review identified few methods and strategies describing when and how to update systematic reviews. *Journal of clinical epidemiology*, 60(11):1095–e1.

Netea, M. G., Giamarellos-Bourboulis, E. J., Domínguez-Andrés, J., Curtis, N., van Crevel, R., van de Veerdonk, F. L., and Bonten, M. (2020). Trained immunity: a tool for reducing susceptibility to and the severity of sars-cov-2 infection. *Cell*, 181(5):969–977.

Newman, M. (2019). Has cochrane lost its way? *Bmj*, 364.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

Pace, L. and Salvan, A. (2019). Likelihood, replicability and Robbins' confidence sequences. *International Statistical Review*.

Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., Catalá-López, F., Li, L., Reid, E. K., Sarkis-Onofre, R., et al. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS medicine*, 13(5):e1002028.

Pereira, T. V. and Ioannidis, J. P. (2011). Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *Journal of clinical epidemiology*, 64(10):1060–1069.

Peto, R. (1972). Discussion on the paper 'Regression models and Life Tables by Sir David R. Cox. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2):205–208.

Peto, R. (1987). Why do we need systematic overviews of randomized trials?(transcript of an oral presentation, modified by the editors). *Statistics in medicine*, 6(3):233–240.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, 135(2):185–198.

Pfeiffer, T., Bertram, L., and Ioannidis, J. P. (2011). Quantifying selective reporting and the Proteus phenomenon for multiple datasets with similar bias. *PLoS One*, 6(3):e18362.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

Pocock, S. J. (2006). Current controversies in data monitoring for clinical trials. *Clinical trials*, 3(6):513–521.

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Marc, G. P., Moreira, E. D., Zerbini, C., et al. (2020). Safety and efficacy of the NT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*.

Polanin, J. R. and Williams, R. T. (2016). Overcoming obstacles in obtaining individual participant data for meta-analysis. *Research synthesis methods*, 7(3):333–341.

Potthoff, R. F. (2007). Prediction markets, Bayesian priors, and clinical trials. *Journal of statistical planning and inference*, 137(11):3706–3721.

Proschan, M. A., Lan, K. G., and Wittes, J. T. (2006). *Statistical monitoring of clinical trials: a unified approach*. Springer Science & Business Media.

Ramdas, A., Ruf, J., Larsson, M., and Koolen, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*.

Riley, R. D., Higgins, J. P., and Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *Bmj*, 342.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Annals of Mathematical Statistics*, 41:1397–1409.

Roberts, I. and Ker, K. (2015). How systematic reviews cause research waste. *The Lancet*, 386(10003):1536.

Robinson, K. A. and Goodman, S. N. (2011). A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Annals of internal medicine*, 154(1):50–55.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, 86(3):638.

Rothman, K. J., Gallacher, J. E., and Hatch, E. E. (2013). Why representativeness should be avoided. *International journal of epidemiology*, 42(4):1012–1014.

Royall, R. (1997). *Statistical evidence: a likelihood paradigm*, volume 71. Chapman & Hall/CRC.

Royall, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association*, 95(451):760–768.

Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68(1):316–319.

ter Schure, J. (2019). Code for paper Accumulation bias in meta-analysis: The need to consider time in error control. https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:127617.

ter Schure, J. (2021a). Code for blogposts Accumulation Bias: How to handle it. https://osf.io/p2rtw/.

ter Schure, J. (2021b). Code for paper ALL-IN meta-analysis: breathing life into living systematic reviews. https://osf.io/d9jny/.

ter Schure, J. (2021c). Judith ter Schure's contribution to the Discussion of 'Testing by betting: A strategy for statistical and scientific communication' by Glenn Shafer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(2):440–441.

ter Schure, J. and Grünwald, P. (2019). Accumulation Bias in meta-analysis: the need to consider time in error control [version 1; peer review: 2 approved]. *F1000Research*, 8:962.

ter Schure, J., Ly, A., Pérez-Ortiz, M. F., and Grünwald, P. (2020a). Safestats and ALL-IN meta-analysis project page. https://projects.cwi.nl/safestats/.

ter Schure, J. and Pérez-Ortiz, M. F. (2022). Code for paper Safe logrank test. https://osf.io/3n8g2/.

ter Schure, J., Pérez-Ortiz, M. F., Ly, A., and Grünwald, P. (2020b). The safe logrank test: Error control under continuous monitoring with unlimited horizon. *arXiv preprint arXiv:2011.06931*.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika*, 70(2):315–326.

Senn, S. (2000). The many modes of meta. *Drug Information Journal*, 34(2):535–549.

Senn, S. (2008). A note concerning a selection "paradox" of Dawid's. *The American Statistician*, 62(3):206–210.

Senn, S. (2014). A note regarding meta-analysis of sequential trials with stopping for efficacy. *Pharmaceutical Statistics*, 13(6):371–375.

Shafer, G. (2019). The language of betting as a strategy for statistical and scientific communication. http://probabilityandfinance.com/articles/54.pdf. Accessed 16 May 2019.

Shafer, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(2):407–431.

Shafer, G., Shen, A., Vereshchagin, N., and Vovk, V. (2011). Test martingales, Bayes factors and p-values. *Statistical Science*, 26(1):84–101.

Shafer, G. and Vovk, V. (2019). *Game-Theoretic Foundations for Probability and Finance*, volume 455. John Wiley & Sons.

Shamy, M., Dewar, B., and Fedyk, M. (2020). Different meanings of equipoise and the four quadrants of uncertainty. *Journal of Clinical Epidemiology*, 127:248–249.

Simmonds, M., Salanti, G., McKenzie, J., and Elliott, J. (2017). Living systematic reviews: 3. statistical methods for updating meta-analyses. *Journal of clinical epidemiology*, 91:38–46.

Simmonds, M. C., Tierney, J., Bowden, J., and Higgins, J. P. (2011). Meta-analysis of time-to-event data: a comparison of two-stage methods. *Research synthesis methods*, 2(3):139–149.

SIPTA, L. (2021). Game-theoretic foundations for statistical testing: Glenn shafer. https://www.youtube.com/watch?v=pOMTb_x2mxw.

Slud, E. V. (1984). Sequential linear rank tests for two-sample censored survival data. *The Annals of Statistics*, pages 551–571.

Slud, E. V. (1992). Partial likelihood for continuous-time stochastic processes. *Scandinavian journal of statistics*, pages 97–109.

Sutton, A. J., Cooper, N. J., Jones, D. R., Lambert, P. C., Thompson, J. R., and Abrams, K. R. (2007). Evidence-based sample size calculations based upon updated meta-analysis. *Statistics in Medicine*, 26(12):2479–2500. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.2704.

Tierney, J. F., Fisher, D. J., Vale, C. L., Burdett, S., Rydzewska, L. H., Rogozińska, E., Godolphin, P. J., White, I. R., and Parmar, M. K. (2021). A framework for prospective, adaptive meta-analysis (FAME) of aggregate data from randomised trials. *PLoS medicine*, 18(5):e1003629.

Turner, R., Ly, A., and Grünwald, P. (2021). Safe tests and always-valid confidence intervals for contingency tables and beyond. *arXiv preprint arXiv:2106.02693*.

Turner, R., Ly, A., Pérez-Ortiz, M. F., ter Schure, J., and Grünwald, P. (2022). *R-package* `safestats`. R package version 0.8.6, Maintainer: Alexander Ly <a.ly@jaspstats.org>, https://cran.r-project.org/package=safestats.

Turner, R. M., Bird, S. M., and Higgins, J. P. (2013). The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PloS one*, 8(3):e59202.

Ville, J. (1939). *Etude critique de la notion de collectif*. Gauthier-Villars, Paris.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination, and applications. *Annals of Statistics*.

Wald, A. (1947). Sequential analysis. 1947. *Zbl0029*, 15805.

Walters, S. J., dos Anjos Henriques-Cadby, I. B., Bortolami, O., Flight, L., Hind, D., Jacques, R. M., Knox, C., Nadin, B., Rothwell, J., Surtees, M., et al. (2017). Recruitment and retention of participants in randomised controlled trials: a review of trials funded and

published by the united kingdom health technology assessment programme. *BMJ open*, 7(3):e015276.

Wang, M. Q., Yan, A. F., and Katz, R. V. (2018). Researcher requests for inappropriate analysis and reporting: A US survey of consulting biostatisticians. *Annals of internal medicine*, 169(8):554–558.

van Werkhoven, C. H., ter Schure, J., Bonten, M., Netea, M., Grünwald, P., and Ly, A. (2021). Anytime Live and Leading Interim meta-analysis of the impact of Bacillus Calmette-Guérin vaccination in health care workers and elderly during the sars-cov-2 pandemic (ALL-IN-META-BCG-CORONA). https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=213069&VersionID=1473878.

Wetterslev, J., Jakobsen, J. C., and Gluud, C. (2017). Trial sequential analysis in systematic reviews with meta-analysis. *BMC medical research methodology*, 17(1):39.

Wetterslev, J., Thorlund, K., Brok, J., and Gluud, C. (2008). Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *Journal of clinical epidemiology*, 61(1):64–75.

Whitehead, A. (1997). A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Statistics in medicine*, 16(24):2901–2913.

Whitehead, A. (2002). *Meta-analysis of controlled clinical trials*, volume 7. John Wiley & Sons.

van der Wier, M. (2020). Proef: Helpt tbc-vaccin in de strijd tegen het coronavirus? https://www.trouw.nl/nieuws/proef-helpt-tbc-vaccin-in-de-strijd\-tegen-het-coronavirus~b2f670df/.

Wu, J. and Xiong, X. (2017). Group sequential survival trial design and monitoring using the log-rank test. *Statistics in biopharmaceutical research*, 9(1):35–43.

Young, C. and Horton, R. (2005). Putting clinical trials into context. *The Lancet*, 366(9480):107–108. Publisher: Elsevier.

Yu, L.-M., Bafadhel, M., Dorward, J., Hayward, G., Saville, B. R., Gbinigie, O., Van Hecke, O., Ogburn, E., Evans, P. H., Thomas, N. P., et al. (2021). Inhaled budesonide for Covid-19 in people at high risk of complications in the community in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial. *The Lancet*, 398(10303):843–855.

Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases*, 27(5):335–371.