

Bibliography

- [1] K. A. Baggerly, K. R. Coombes. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 3:1309–1334, 2009.
- [2] M. Bakker, J. M. Wicherts. The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43:666–678, 2011.
- [3] D. Bassler, M. Briel, V. M. Montori, M. Lane, P. Glasziou, Q. Zhou, D. Heels-Ansdell, S. D. Walter, G. H. Guyatt. Stopping Randomized Trials Early for Benefit and Estimation of Treatment Effects: Systematic Review and Meta-regression Analysis. *JAMA*, 303:1180–1187, 2010.
- [4] P. L. Bedard, M. K. Krzyzanowska, M. Pintilie, I. F. Tannock. Statistical Power of Negative Randomized Controlled Trials Presented at American Society for Clinical Oncology Annual Meetings. *Journal of Clinical Oncology*, 25:3482–3487, 2007.
- [5] C. G. Begley, L. M. Ellis. Drug development: Raise standards for preclinical cancer research. *Nature*, 483:531–533, 2012.
- [6] S. Belia, F. Fidler, J. Williams, G. Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods*, 10:389–396, 2005.
- [7] Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 289–300, 1995.
- [8] C. Bennett, A. Baird, M. Miller, G. Wolford. Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction. *Journal of Serendipitous and Unexpected Results*, 1:1–5, 2010.
- [9] A. F. Bogaert. Biological versus nonbiological older brothers and men’s sexual orientation. *PNAS*, 103:10771–10774, 2006.
- [10] R. Bramwell, H. West. Health professionals’ and service users’ interpretation of screening test results: experimental study. *BMJ*, 2006.
- [11] C. G. Brown, G. D. Kelen, J. J. Ashton, H. A. Werman. The beta error and sample size determination in clinical trials in emergency medicine. *Annals of Emergency Medicine*, 16:183–187, 1987.
- [12] K. S. Button, J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, M. R. Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 2013.
- [13] J. Carp. The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, 63:289–300, 2012.
- [14] A.-W. Chan, A. Hróbjartsson, M. T. Haahr, P. C. Gøtzsche, D. G. Altman. Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles. *JAMA*, 291:2457–2465, 2004.
- [15] A.-W. Chan, A. Hróbjartsson, K. J. Jørgensen, P. C. Gøtzsche, D. G. Altman. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ*, 337:a2299, 2008.
- [16] K. C. Chung, L. K. Kalliainen, R. A. Hayward. Type II (beta) errors in the hand literature: the importance of power. *The Journal of Hand Surgery*, 23:20–25, 1998.
- [17] M. D. D. CS, W. GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA*, 272:122–124, 1994.
- [18] D. Eyding, M. Lelgemann, U. Grouven, M. Härter, M. Kromp, T. Kaiser, M. F. Kerekes, M. Gerken, B. Wieseler. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ*, 341:2010.
- [19] K. R. Gabriel. A simple method of multiple comparisons of means. *Journal of the American Statistical Association*, 73:724–729, 1978.
- [20] J. Galak, R. A. LeBoeuf, L. D. Nelson, J. P. Simmons. Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103:933–948, 2012.
- [21] A. Gelman, P. N. Price. All maps of parameter estimates are misleading. *Statistics in Medicine*, 18:3221–3234, 1999.

- [22] A. Gelman, H. Stern. The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, 60:328–331, 2006.
- [23] F. Gonon, J.-P. Kohnsman, D. Cohen, T. Boraud. Why Most Biomedical Findings Echoed by Newspapers Turn Out to be False: The Case of Attention Deficit Hyperactivity Disorder. *PLoS ONE*, 7:e44275, 2012.
- [24] S. N. Goodman. Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, 130:995–1004, 1999.
- [25] P. C. Gøtzsche. Believability of relative risks and odds ratios in abstracts: cross sectional study. *BMJ*, 333:231–234, 2006.
- [26] P. C. Gøtzsche. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clinical Trials*, 10:31–56, 1989.
- [27] E. Hauer. The harm done by tests of significance. *Accident Analysis & Prevention*, 36:495–500, 2004.
- [28] D. Hemenway. Survey Research and Self-Defense Gun Use: An Explanation of Extreme Overestimates. *The Journal of Criminal Law and Criminology*, 87:1430–1445, 1997.
- [29] K. Huwiler-Müntener, P. Jüni, C. Junker, M. Egger. Quality of Reporting of Randomized Trials as a Measure of Methodologic Quality. *JAMA*, 287:2801–2804, 2002.
- [30] J. P. A. Ioannidis. Why Most Discovered True Associations Are Inflated. *Epidemiology*, 19:640–648, 2008.
- [31] J. P. A. Ioannidis. Why Most Published Research Findings Are False. *PLoS Medicine*, 2:e124, 2005.
- [32] J. P. A. Ioannidis. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*, 294:218–228, 2005.
- [33] J. P. A. Ioannidis, T. A. Trikalinos. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology*, 58:543–549, 2005.
- [34] J. J. Kirkham, K. M. Dwan, D. G. Altman, C. Gamble, S. Dodd, R. Smyth, P. R. Williamson. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*, 340:c365–c365, 2010.
- [35] W. Krämer, G. Gigerenzer. How to Confuse with Statistics or: The Use and Misuse of Conditional Probabilities. *Statistical Science*, 20:223–230, 2005.
- [36] P. A. Kyzas, K. T. Loizou, J. P. A. Ioannidis. Selective Reporting Biases in Cancer Prognostic Factor Studies. *Journal of the National Cancer Institute*, 97:1043–1055, 2005.
- [37] J. R. Lanzante. A cautionary note on the use of error bars. *Journal of climate*, 18:3699–3703, 2005.
- [38] S. E. Lazic. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis?. *BMC Neuroscience*, 11:5, 2010.
- [39] J. LeLorier, G. Gregoire, A. Benhaddad. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine*, 1997.
- [40] M. Marshall, A. Lockwood, C. Bradley, C. Adams, C. Joy, M. Fenton. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *The British Journal of Psychiatry*, 176:249–252, 2000.
- [41] A. M. Metz. Teaching Statistics in Biology: Using Inquiry-based Learning to Strengthen Understanding of Statistical Analysis in Biology Laboratory Courses. *CBE Life Sciences Education*, 7:317–326, 2008.
- [42] E. Mills, P. Wu, J. Gagnier, D. Heels-Ansdell, V. M. Montori. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *Journal of Clinical Epidemiology*, 58:662–667, 2005.
- [43] V. M. Montori, P. J. Devereaux, N. Adhikari. Randomized trials stopped early for benefit: a systematic review. *JAMA*, 294:2203–2209, 2005.
- [44] S. Nieuwenhuis, B. U. Forstmann, E.-J. Wagenmakers. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, 14:1105–1109, 2011.
- [45] T. V. Pereira, J. P. A. Ioannidis. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *Journal of Clinical Epidemiology*, 64:1060–1069, 2011.

- [46] A. C. Plint, D. Moher, A. Morrison, K. Schulz, e. al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical journal of Australia*, 185:263–267, 2006.
- [47] A. P. Prayle, M. N. Hurley, A. R. Smyth. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ*, 344:d7373, 2011.
- [48] D. F. Preusser, W. A. Leaf, K. B. DeBartolo, R. D. Blomberg, M. M. Levy. The effect of right-turn-on-red on pedestrian and bicyclist accidents. *Journal of Safety Research*, 13:45–55, 1982.
- [49] F. Prinz, T. Schlange, K. Asadullah. Believe it or not: how much can we rely on published data on potential drug targets?. *Nature Reviews Drug Discovery*, 10:328–329, 2011.
- [50] N. Schenker, J. F. Gentleman. On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55:182–186, 2001.
- [51] J. D. Schoenfeld, J. P. A. Ioannidis. Is everything we eat associated with cancer? A systematic cookbook review. *American Journal of Clinical Nutrition*, 97:127–134, 2013.
- [52] S. Schroter, N. Black, S. Evans, F. Godlee, L. Osorio, R. Smith. What errors do peer reviewers detect, and does training improve their ability to detect them?. *JRSM*, 101:507–514, 2008.
- [53] J. P. Simmons, L. D. Nelson, U. Simonsohn. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22:1359–1366, 2011.
- [54] D. G. Smith, J. Clemens, W. Crede, M. Harvey, E. J. Gracely. Impact of multiple comparisons in randomized clinical trials. *The American Journal of Medicine*, 83:545–550, 1987.
- [55] A. Tatsioni, N. G. Bonitsis, J. P. A. Ioannidis. Persistence of Contradicted Claims in the Literature. *JAMA*, 298:2517–2526, 2007.
- [56] S. Todd, A. Whitehead, N. Stallard, J. Whitehead. Interim analyses and sequential designs in phase III studies. *British Journal of Clinical Pharmacology*, 51:394–399, 2001.
- [57] R. Tsang, L. Colley, L. D. Lynd. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. *Journal of Clinical Epidemiology*, 62:609–616, 2009.
- [58] E. Wagenmakers, R. Wetzels. Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, 2011.
- [59] H. Wainer. The Most Dangerous Equation. *American Scientist*, 95:249–256, 2007.
- [60] J. M. Wicherts, M. Bakker, D. Molenaar. Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PLoS ONE*, 6:e26828, 2011.
- [61] J. M. Wicherts, D. Borsboom, J. Kats, D. Molenaar. The poor availability of psychological research data for reanalysis. *American Psychologist*, 61:726–728, 2006.