



Cite this article: Loisel D, Ramchandra R. 2015 A counterview of 'An investigation of the false discovery rate and the misinterpretation of p -values' by Colquhoun (2014). *R. Soc. open sci.* **2**: 150217.
<http://dx.doi.org/10.1098/rsos.150217>

Received: 19 May 2015

Accepted: 30 June 2015

Subject Category:

Mathematics

Author for correspondence:

D. Loisel

e-mail: ds.loiselle@auckland.ac.nz

The accompanying reply can be viewed at
<http://dx.doi.org/10.1098/rsos.150319>.

Electronic supplementary material is available
at <http://dx.doi.org/10.1098/rsos.150217> or via
<http://rsos.royalsocietypublishing.org>.

A counterview of 'An investigation of the false discovery rate and the misinterpretation of p -values' by Colquhoun (2014)

D. Loisel^{1,2} and R. Ramchandra¹

¹Department of Physiology and ²Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand

In commenting on the instructive, comprehensive and entertainingly written article by Prof. Colquhoun (hereinafter referred to as 'the author'), we state unequivocally that we have no truck with its motivation. Indeed, we too often find the Fisherian approach troubling. Nor do we wish to become involved in the relative merits of 'Bayesian' versus 'Fisherian' methods. Rather we wish to focus on the author's underlying model, reminding the reader that the output of any mathematical model is only as good as its input parameters. In this regard, we find the numeric value of 0.1 for the parameter describing 'the probability that the putative effect is real' to be wholly unrealistic for divining the appropriate p -value to be used as the basis for deciding whether the outcome of an experiment provides evidence 'for' or 'against' rejection of the null hypothesis.

We readily admit that we have no more idea than does the author regarding the true value of the 'prevalence' parameter for experimental science, so we have adopted three distinct approaches to estimate it. (i) First, and with reference to the author's charmingly apposite introductory quote from George Elliot's *Middlemarch*, we state our 'gut instinct' estimate to be 'greater than 50%'. (ii) Second, and widening the scope, we have canvassed the senior investigators in our Department of Physiology for their personal estimates of the fraction of times that their explicit, experimentally testable hypotheses have proven to be supported by experimental results. We are aware of the somewhat circular logic of this undertaking because, in each case, 'classical' hypothesis testing underlies the 'guesstimates'. Nevertheless, we consider that well-informed scientists can do better than a flip-of-the-coin (and certainly better than the roll of a decahedral die) in guessing the pathway along which truth lies. (iii) Finally, we have examined the

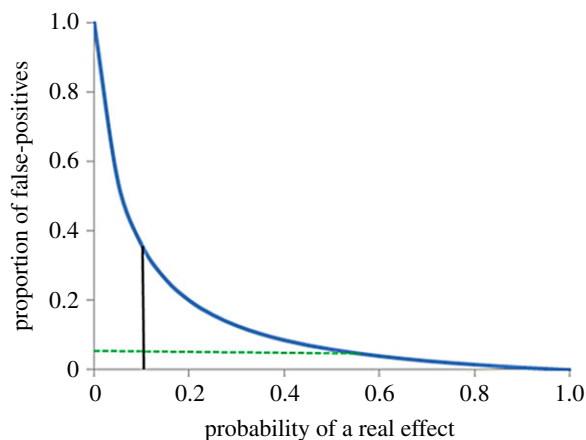


Figure 1. Proportion of ‘false-positive’ rejections of the null hypothesis as a function of the probability that the hypothesized effect exists (i.e. is ‘real’). The curve is drawn for the case when α (the ‘significance level’ or the putative risk of a type I error) is 0.05 and the power of the test (i.e. the probability of correctly rejecting the null hypothesis when it is false) has the value 0.8 (mimicking the value adopted by Colquhoun [32]).

statistical analyses of a selection of published papers ($N = 25$), predominantly in the field of ‘cardiovascular biology’ (our personal areas of interest). Our criteria for selection of papers were as follows: (i) those listed in PubMed and published or pre-published during the months of March or April 2015, and available in full, sans cost; and (ii) those in which an experimentally testable hypothesis was either explicitly stated or strongly implied (‘we hypothesize’, ‘we infer’, ‘we propose’, ‘we aim to test’, etc.). We rejected reviews, meta-analyses, case studies, investigations of genetic associations and those articles of a purely descriptive nature. That is, we focused exclusively on studies based on experimental interventions. In all 25 cases, the Fisherian approach had been adopted by the authors, with the value of α either stated explicitly or implied in *Results* to be 0.05. Analyses of variance and *t*-tests prevailed. In 14 cases, the authors reported *p*-values of <0.05 : <0.02 , <0.01 , <0.001 or <0.0001 (in one novel case, as an undefined sequence of asterisks embedded in graphs presented in *Results*). References [1] to [25] were the articles we surveyed.

The outcomes of both surveys were surprising but comparable. The ‘guesstimates’ of predictive success by our senior investigator co-workers ($N = 11$) ranged from 50% to 90% with the mean \pm standard deviation of $69.6\% \pm 13.1\%$. With respect to the ‘literature survey’, in only three cases were the authors obliged to state that their results did not support their explicitly stated hypothesis—i.e. the null hypothesis could not be rejected or, in plain English, the authors’ scientific hypothesis was declared to be wrong. The complement (22 manuscripts in each of which the null hypothesis was rejected) represents a ‘prevalence’ of 0.88 (a value that exceeds even *our* (probably inflated) ‘gut feelings’).

How are these apparently convincing results to be explained *vis-à-vis* Prof. Colquhoun’s counter-conclusion? Do they represent yet another example of publication bias [26,27] (across some 15 different Journals and journal Editors)? Or have all the investigators succumbed to one or more of: HARKing [28], file drawer-ing [29,30], under-powering [31,32], data-stretching [33], bias [30], over-interpretation, *p*-stretching, Bayes-watching [34] or any of the other sins of which hypothesis-testing is accused? It seems unlikely that ‘circular reasoning’ (reflecting the unavoidable fact that, in every case, classical hypothesis testing provided the decision-basis) could have played a large role, especially given that 14 of the results would have satisfied Berger’s maximum-likelihood criterion (see appendix A5 of Colquhoun [32]). Perhaps we are all unwitting players in a great academic hoax. In this regard, we find it noteworthy that granting agencies commonly favour the presentation of results from ‘pilot studies’. These require the submitter to walk a narrow path between necessarily few observations while avoiding any hint that the study has already been performed. Do such ‘pre-nuptials’ simultaneously dupe both the benefactor and the academic mendicant?

Instead of such speculation, we find it instructive to present an analysis (appendix 1) and graph (figure 1), based on Prof. Colquhoun’s ‘tree diagrams’ (figures 2 and 3 in the original).

In figure 1, the vertical line at 0.1 intersects the curve at a value of 0.36, thereby duplicating the data shown in figure 2 of Colquhoun [32]. Its location is predicated on the author’s implied assumption that biomedical scientists make correct predictions only some 10% of the time. The dashed horizontal line

intersects the curve at a value of 0.55. That is, if a scientist makes hypotheses that are correct at least 55% of the time, then he or she is, in fact, already working at the commonly assumed ‘significance’ level of 0.05 (given by the intercept on the ordinate), so that there would be little justification for its 50-fold reduction, as advocated by Prof. Colquhoun. This is perhaps not unexpected, given that nearly 50% of the 25 papers that we surveyed report p -values very much smaller than their pre-assigned values of α . Furthermore, it accords with the other two of our admittedly ‘free-form’ estimates. Finally, we note that Prof. Colquhoun has examined a specific case of a p -value close to 0.05. Our investigation of this issue (performing simulations using Prof. Colquhoun’s R-based software program) leads us to conclude that the resulting false discovery rate is likewise dependent on input parameters (especially the critical effect size). Because we wish to maintain focus strictly on the input parameter: ‘prevalence’, we present the results of that investigation in appendix 2 in the electronic supplementary material.

In conclusion, we find it difficult to imagine how science could have achieved its manifold successes if scientists have been wrong 90% of the time. Hence, we suspect that a number of behaviours facilitate a high probability of a real effect, thereby rendering scientific hypotheses robust against extreme probabilities of failure. We count among these behaviours the following common practices: achievement of familiarity with the literature and relentless self-criticism, together with willingness to test ideas in the crucible of public debate, to seek direction from the outcome of under-powered pilot studies, to exploit the power of even simple mathematical models and, on occasion, to disregard much of the preceding and, instead, ‘to go with one’s gut-feeling’.

Authors’ contributions. The authors contributed equally to the conception, design, drafting and revision of the manuscript.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

Appendix 1. Derivation of the false discovery rate

Let

x = the unknown (and unknowable) probability of a real effect (P_{real}): $0 \leq x \leq 1$,

$y(x)$ = the proportion of false-positive decisions (F^+),

F^- = the proportion of false-negative decisions = $\beta \times P_{\text{real}}$,

T^+ = the proportion of true-positive decisions = $(1 - \beta) \times P_{\text{real}}$

T^- = the proportion of true-negative decisions = $(1 - \alpha) \times P_{\text{real}}$; and

F^+ = the proportion of false-positive decisions = $\alpha \times (1 - P_{\text{real}})$,

where α = the probability of a type I error (the probability of falsely declaring a true null hypothesis false) and β = the probability of a type II error (the probability of failure to reject a false null hypothesis).

In strict accord with the procedure outlined in figure 2 of Colquhoun [32], where it is labelled ‘the false discovery rate’, the proportion of false-positive decisions is given by

$$y(x) = \frac{F^+}{(F^+ + T^+)}$$

References

1. Abdelhamid DS, Zhang Y, Lewis DR, Moghe PV, Welsh WJ, Uhrich KE. 2015 Tartaric acid-based amphiphilic macromolecules with ether linkages exhibit enhanced repression of oxidized low density lipoprotein uptake. *Biomaterials* **53**, 32–39. (doi:10.1016/j.biomaterials.2015.02.038)
2. Berry E, Hernandez-Anzaldo S, Ghomashchi F, Lehner R, Murakami M, Gelb MH, Kassiri Z, Wang X, Fernandez-Patron C. 2015 Matrix metalloproteinase-2 negatively regulates cardiac secreted phospholipase A2 to modulate inflammation and fever. *J. Am. Heart Assoc.* **4**. (doi:10.1161/jaha.115.001868)
3. Bilet L, Brouwers B, van Ewijk PA, Hesselink MKC, Kooi ME, Schrauwen P, Schrauwen-Hinderling VB. 2015 Acute exercise does not decrease liver fat in men with overweight or NAFLD. *Sci. Rep.* **5**, 9709. (doi:10.1038/srep09709)
4. Cheng M, Huang K, Zhou J, Yan D, Tang Y-L, Zhao TC, Miller RJ, Kishore R, Losordo DW, Qin G. 2015 A critical role of Src family kinase in SDF-1/CXCR4-mediated bone-marrow progenitor cell recruitment to the ischemic heart. *J. Mol. Cell. Cardiol.* **81**, 49–53. (doi:10.1016/j.jmcc.2015.01.024)
5. Dupuis LE, Berger MG, Feldman S, Doucette L, Fowlkes V, Chakravarti S, Thibaudeau S, Alcalá NE, Bradshaw AD, Kern CB. 2015 Lumican deficiency results in cardiomyocyte hypertrophy with altered collagen assembly. *J. Mol. Cell. Cardiol.* **84**, 70–80. (doi:10.1016/j.jmcc.2015.04.007)
6. García-Bermúdez M, López-Mejías R, Genre F, Castañeda S, Corrales A, Llorca J, González-Juanateja C, Ubilla B, Miranda-Fillooy JA, Pina T, Gómez-Vaquero C, Rodríguez-Rodríguez L, Fernández-Gutiérrez B, Balsa A, Pascual-Salcedo D, López-Longo FJ, Carreira P, Blanco R, Martín J, González-Gay MA. 2015 Lack of association between JAK3 gene polymorphisms and cardiovascular disease in Spanish patients with rheumatoid arthritis. *BioMed Research International* **2015**, 318364. (doi:10.1155/2015/318364)
7. Hammer KP, Ljubojevic S, Ripplinger CM, Pieske BM, Bers DM. 2015 Cardiac myocyte alternans in intact heart: Influence of cell–cell coupling and β -adrenergic stimulation. *J. Mol.*

- Cell. Cardiol.* **84**, 1–9. (doi:10.1016/j.yjmcc.2015.03.012)
8. Huang Y-H, Ching-Chang I, Kuo C-H, Hsu Y-Y, Lee F-T, Shi G-Y, Tseng S-H, Wu H-L. 2015 Thrombomodulin Promotes Corneal Epithelial Wound Healing. *PLoS ONE* **10**, e0122491.
 9. Iordache F, Constantinescu A, Andrei E, Maniu H. 2015 Histone acetylation regulates the expression of HoxD9 transcription factor in endothelial progenitor cells. *Romanian J. Morphol. Embryol.* **56**, 107–113.
 10. Kim do Y, Abdelwahab MG, Lee SH, O'Neill D, Thompson RJ, Duff HJ, Sullivan PG, Rho JM. 2015 Ketones prevent oxidative impairment of hippocampal synaptic integrity through KATP channels. *PLoS ONE* **10**, e0119316.
 11. Krishnaswamy PS, Egom EE, Moghtadaei M, Jansen HJ, Azer J, Bogachev O, Mackasey M, Robbins C, Rose RA. 2015 Altered parasympathetic nervous system regulation of the sinoatrial node in Akita diabetic mice. *J. Mol. Cell. Cardiol.* **82**, 125–135.
 12. Kumar SA, Magnusson M, Ward LC, Paul NA, Brown L. 2015 A green algae mixture of *Scenedesmus* and *Schroederiella* attenuates obesity-Linked metabolic syndrome in rats. *Nutrients* **7**, 2771–2787.
 13. Li RWS, Yang C, Chan SW, Hoi MPM, Lee SMY, Kwan YW, Leung GPH. 2015 Relaxation effect of abacavir on rat basilar arteries. *PLoS ONE* **10**, e0123043.
 14. Liu M, Pan Q, Chen Y, Yang X, Zhao B, Jia L, Zhu Y, Han J, Li X, Duan Y. 2015 NaoXinTong inhibits the development of diabetic retinopathy in db/db mice. *Evidence-based Complementary and Alternative Medicine* **2015**, 242517.
 15. Magliano DC, Penna-de-Carvalho A, Vazquez-Carrera M, Mandarim-de-Lacerda CA, Aguila MB. 2015 Short-term administration of GW501516 improves inflammatory state in white adipose tissue and liver damage in high-fructose-fed mice through modulation of the renin-angiotensin system. *Endocrine*.
 16. Muthuramu I, Singh N, Amin R, Nefyodova E, Debasse M, Van Horenbeeck I, Jacobs F, De Geest B. 2015 Selective homocysteine-lowering gene transfer attenuates pressure overload-induced cardiomyopathy via reduced oxidative stress. *J. Mol. Med.* **93**, 609–618.
 17. Previs MJ, Prosser BL, Mun JY, Brevis SB, Gulick J, Lee K, Robbins J, Craig R, Lederer WJ, Warshaw DM. 2015 Myosin-binding protein C corrects an intrinsic inhomogeneity in cardiac excitation-contraction coupling. *Science Advances* **1**, e1400215.
 18. Quttainah M, Al-Hejalain R, Saleh S, Parhar R, Conca W, Bulwer B, Moorjani Z, Catarino P, Elsayed R, Shoukri M, AlJufan M, AlShahid M, Ouban A, Al-Halees Z, Westaby S, Collison K, Al-Mohanna F. 2015 Progression of matrixin and cardiokine expression patterns in an ovine model of heart failure and recovery. *Int. J. Cardiol.* **186**, 77–89. (doi:10.1016/j.ijcard.2015.03.156)
 19. Schoors S, Bruning U, Missaen R, Queiroz KCS, Borgers G, Elia I, Zecchin A, Cantelmo AR, Christen S, Goveia J, Heggermont W, Godde L, Vinckier S, Van Veldhoven PP, Eelen G, Schoonjans L, Gerhardt H, Dewerchin M, Baes M, De Bock K, Ghesquiere B, Lunt SY, Fendt S-M, Carmeliet P. 2015 Fatty acid carbon is essential for dNTP synthesis in endothelial cells. *Nature* **520**, 192–197. (doi:10.1038/nature14362)
 20. Schulkey CE, Regmi SD, Magnan RA, Danzo MT, Luther H, Hutchinson AK, Panzer AA, Grady MM, Wilson DB, Jay PY. 2015 The maternal-age-associated risk of congenital heart disease is modifiable. *Nature* **520**, 230–233. (doi:10.1038/nature14361)
 21. Sun J, Nguyen T, Aponte AM, Menazza S, Kohr MJ, Roth DM, Patel HH, Murphy E, Steenbergen C. 2015 Ischaemic preconditioning preferentially increases protein S-nitrosylation in subsarcolemmal mitochondria. *Cardiovasc. Res.* **106**, 227–236. (doi:10.1093/cvr/cvv044)
 22. Wei M-Y, Xue L, Tan L, Sai W-B, Liu X-C, Jiang Q-J, Shen J, Peng Y-B, Zhao P, Yu M-F, Chen W, Ma L-Q, Zhai K, Zou C, Guo D, Qin G, Zheng Y-M, Wang Y-X, Guanqu J, Liu Q-H. 2015 Involvement of large-conductance Ca²⁺-activated K⁺ channels in chloroquine-induced force alterations in pre-contracted airways smooth muscle. *PLoS ONE* **10**, e0121566. (doi:10.1371/journal.pone.0121566)
 23. Wilson RM, Marshall NE, Jeske DR, Purnell JQ, Thornburg K, Messaoudi I. 2015 Maternal obesity alters immune cell frequencies and responses in umbilical cord blood samples. *Pediatr.*
 24. Yildirim C, Vogel DYS, Hollander MR, Baggen JM, Fontijn RD, Nieuwenhuis S, Haverkamp A, de Vries MR, Quax PHA, Garcia-Vallejo JJ, van der Laan AM, Dijkstra CD, van der Pouw Kraan TCTM, van Royen N, Horrevoets AJG. 2015 Galectin-2 induces a proinflammatory, anti-arteriogenic phenotype in monocytes and macrophages. *PLoS ONE* **10**, e0124347. (doi:10.1371/journal.pone.0124347)
 25. Zhu D, Xie H, Wang X, Liang Y, Yu H, Gao W. 2015 Correlation of plasma catestatin level and the prognosis of patients with acute myocardial infarction. *PLoS ONE* **10**, e0122993. (doi:10.1371/journal.pone.0122993)
 26. Fanelli D. 2010 'Positive' results increase down the hierarchy of the sciences. *PLoS ONE* **5**, e10068. (doi:10.1371/journal.pone.0010068)
 27. Simonsohn U, Nelson LD, Simmons JP. 2014 P-Curve and effect size: correcting for publication bias using only significant results. *Perspect. Psychol. Sci.* **9**, 666–681. (doi:10.1177/174569161453988)
 28. Kerr NL. 1998 HARKing: hypothesizing after the results are known. *Person. Soc. Psychol. Rev.* **2**, 196–217. (doi:10.1207/s15327957pspr0203_4)
 29. Rosenthal R. 1979 The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 638–641. (doi:10.1037/0033-2909.86.3.638)
 30. Ioannidis JPA. 2005 Why most published research findings are false. *CHANCE* **18**, 40–47. (doi:10.1080/09332480.2005.10722754)
 31. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. 2013 Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376. (doi:10.1038/nrn3475)
 32. Colquhoun D. 2014 An investigation of the false discovery rate and the misinterpretation of *p*-values. *R. Soc. open sci.* **1**, 140216. (doi:10.1098/rso.140216)
 33. Greenberg WA, Pollard WA, Alpert WT. 1989 Statistical properties of data stretching. *J. Appl. Econ.* **4**, 383–391. (doi:10.1002/jae.3950040406)
 34. Siegfried T. 2010 Odds are, it's wrong. *Sci. News* **177**, 26–29. (doi:10.1002/scin.5591770721)