

Fixing flaws in science must be professionalized

Ben Goldacre^{a,*}, Tracey Brown^b

^aDepartment Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT

^bSense About Science, 14a Clerkenwell Green, London EC1R 0DP

Accepted 17 June 2015; Published online 3 July 2015

It is heartening to see calls for greater transparency around data and analytic strategies, including in this issue, from such senior academic figures as Robert West. Science currently faces multiple challenges to its credibility. There is an ongoing lack of public trust in science and medicine, often built on weak conspiracy theories about technology such as vaccines [1]. At the same, however, there is clear evidence that we have failed to competently implement the scientific principles we espouse.

The most extreme illustration of this is our slow progress on addressing publication bias. The best current evidence, from dozens of studies on publication bias, shows that only around half of all completed research goes on to be published [2], and studies with positive results are twice as likely to be published [2,3]. Given the centrality of systematic review and meta-analysis to decision making in medicine, this is an extremely concerning finding. It also exemplifies the perversity of our failure to address structural issues in evidence-based medicine: we spend millions of dollars on trials specifically to exclude bias and confounding; but then, at the crucial moment of evidence synthesis, we let all those biases flood back in, by permitting certain results to be withheld.

Progress toward addressing this issue has been slow, with a series of supposed fixes that have often given little more than false reassurance [4]. Clearest among these is the FDA Amendment Act 2007, which required all trials after 2008 on currently available treatments with at least one site in the United States to post results at clinicaltrials.gov within 12 months of completion. There was widespread celebration that this had fixed the problem of publication bias, but no routine public audit of compliance. When one was finally published, in 2012, it showed that the rate of compliance with this legislation was 22% [5].

Conflict of interest: B.G. and T.B. are cofounders of the AllTrials campaign. B.G. has been supported by grants from the Wellcome Trust, NIHR BRC, and the Arnold Foundation. B.G. receives income from speaking and writing on problems in science. T.B. is employed to campaign on science policy issues by Sense About Science, a UK charity.

* Corresponding author.

E-mail address: ben.goldacre@lshtm.ac.uk (B. Goldacre).

It is also worth noting the growing evidence that—despite peer review and editorial control—academic journals are rather bad places to report the findings of clinical trials. Information on side effects in journal publications is known to be incomplete [6], for example; and primary outcomes of trials are routinely switched out and replaced with other outcomes [7,8], which in turn exaggerates the apparent benefits of the treatment. Initiatives such as CONSORT [9] have attempted to standardize trial reporting with best practice guidelines, but perhaps, most encouraging is the instantiation of such principles in clear proformas for reporting. All trials from any region and era can now post results on clinicaltrials.gov, and completion of simple pre-specified fields is mandated, leaving less opportunity for editorializing and manipulation. A recent cohort study of 600 trials at clinicaltrials.gov found that reporting was significantly more complete on this registry than in the associated academic journal article—where one was even available—for efficacy results, adverse events, serious adverse events, and participant flow [10].

Academic articles appear even less informative when compared with clinical study reports (CSRs). These are lengthy documents, with a stereotyped and prespecified structure, which are generally only prepared for trials sponsored by the pharmaceutical industry. For many years, they were kept from view, but are increasingly being sought for use in systematic reviews, and to double-check analyses. A recent study by the German government's cost-effectiveness agency compared 101 CSRs against the academic articles reporting the same trials and found that CSRs were significantly more informative, with important information on benefits and harms absent from the academic articles that many would regard as the canonical document on a trial [6]. Sadly, progress toward greater transparency on CSRs has been hindered by lobbying and legal action from drug companies [11].

Lastly, there have been encouraging recent calls for greater transparency around individual patient data (IPD) from trials, which offers considerable opportunity for IPD meta-analysis and checking initial analyses; although these calls have also been tempered by overstatement of the

privacy risks and administrative challenges (even though both have long been managed for the sharing of large datasets of patients' electronic health records for observational epidemiology research), and position statements on data sharing that exclude most retrospective data on the treatments that are currently in widespread use [4].

The AllTrials campaign [12] calls for all trials—on all uses, of all currently prescribed interventions—to be registered, with their full methods and results reported, and CSRs shared where they have been produced. This is a simple, clear “ask,” and the campaign has had significant policy impact, with extensive lobbying in the United Kingdom and European Union, and a US launch to follow in 2015. It is sobering to note that the first robust quantitative evidence demonstrating the presence of publication bias was published in 1986 and was accompanied by calls for full trial registration [13], which have still not been answered.

This is especially concerning because transparency around the methods and results of whole trials is one of the simplest outstanding structural flaws we face, and there is clear evidence of more subtle and interesting distortions at work throughout the scientific literature. Masicampo and Lalande [14], for example, examined all the *P*-values in 1 year of publications from three highly regarded psychology journals and found an unusually high prevalence of *P*-values just below 0.05 (a conventional cutoff for regarding a finding as statistically significant). A similar study examined all *P*-values in economics journals and found a “hump” of *P*-values just below the traditional cutoff and a “trough” just above [15]. It is highly unlikely that these patterns emerged by chance. On the contrary, anyone who has analyzed data themselves will be well aware of exactly how a marginal *P*-value could be improved, with the judicious application of small changes to the analytic strategy. “Perhaps it might look a little better if we split age into quintiles, rather than 5 year bands?”; “Maybe we should sense-check the data again for outliers?”; “If we took that co-variate out of the model things might change? We can always justify it in the discussion.” Whether we like it or not, the evidence suggests that sentences like these echo through the corridors of academia.

There is also the related problem of selective outcome reporting within studies, and people finding their hypothesis in their results, a phenomenon which has also been detected through close analysis of large quantities of research. One unusual landmark study inferred power calculations from a large sample of brain imaging studies which were looking for correlations between anatomical and behavioral features: overall, this literature reported almost twice as many positive findings as can plausibly be supported by the number of hypotheses the authors claim to have tested [16]. The most likely explanation for this finding is highly concerning: it appears that large numbers of researchers have effectively gone on fishing expeditions, comparing multiple anatomical features against multiple behavioral features, then selectively reported the positive findings,

and misleadingly presented their work as if they had only set out to examine that one correlation.

For epidemiology, all this raises important questions. It is clear that there are discretionary decisions made by researchers that can affect the outcomes of research, whether observational studies or randomized trials. Entire studies that go unpublished—the crudest form of outcome reporting bias—is in many ways the simplest to address, with universal registration and disclosure of results, accompanied by close monitoring of compliance through universities, ethics committees, sponsors, and journals. Addressing the other distortions may prove more challenging. One option is to demand extensive and laborious prespecification of analytic strategy, with similarly extensive documentation of any deviations. This may help but may also miss much salient information, as there are so many small decisions (or “researcher degrees of freedom” [17]) in an analysis, and some may not even be thought of before the analysis begins. A more extreme option might be to demand a locked and publicly accessible log of every command and program ever run on a researcher's installation of their statistics package, providing a cast iron historical record of any attempt to exaggerate a finding through multiple small modifications to the analysis. The Open Science Framework offers tools to facilitate something approaching this [18]: although even such extreme approaches are still vulnerable to deliberate evasion. Clearly, a trade-off must emerge between what is practical, what is effective, what the culture of science will bear, and perfection: but at the moment, these problems are largely unaddressed and underdiscussed. Meanwhile, the experience of publication bias suggests progress will be slow.

Is there a way forward? We think so. The flaws we see today, in the structures of evidence-based medicine, are a significant public health problem. It is remarkable that we should have identified such widespread problems, with a demonstrable impact on patient care, documented them meticulously, and then left matters to fix themselves. It is as if we had researched the causes of cholera, and then sat proudly on our publications, doing nothing about cleaning the water or saving lives. Yet, all too often efforts to improve scientific integrity, and fix the flaws in our implementation of the principles of evidence-based medicine, are viewed as a hobby, a side project, subordinate to the more important business of publishing academic articles.

We believe that this is the core opportunity. Fixing structural flaws in science are labor intensive. It requires extensive lobbying of policy makers and professional bodies; close analysis of evidence on flaws and opportunities; engaging the public to exert pressure back on professionals; creating digital infrastructure to support transparency; open, public audit of best and worst practice; and more. If we do not regard this as legitimate professional activity—worthy of grants, salaries, and foreground attention from a reasonable number of trained scientists and medics—then it will not happen. The public, and the patients of the future, may not judge our inaction kindly.

References

- [1] Goldacre B. *Bad science*. London, UK: Fourth Estate Ltd; 2008.
- [2] Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS ONE* 2014;9:e114023.
- [3] Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:1–193. iii, ix – xi.
- [4] Goldacre B. Commentary on Berlin et al. “Bumps and bridges on the road to responsible sharing of clinical trial data”. *Clin Trials* 2014; 11:15–8.
- [5] Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 2012;344:d7373.
- [6] Wieseler B, Wolfram N, McGauran N, Kerekes MF, Vervölgyi V, Kohlepp P, et al. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished clinical study reports with publicly available data. *Plos Med* 2013;10:e1001526.
- [7] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977–84.
- [8] Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
- [9] Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Trial* 2010;11:32.
- [10] Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *Plos Med* 2013;10:e1001566.
- [11] Groves T, Godlee F. The European Medicines Agency’s plans for sharing data from clinical trials. *BMJ* 2013;346:f2961.
- [12] AllTrials. The AllTrials Campaign [Internet] 2013: Available at <http://www.alltrials.net/> Accessed January 9, 2014.
- [13] Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529–41.
- [14] Masicampo EJ, Lalande DR. A peculiar prevalence of p values just below .05. *Q J Exp Psychol* 2012;65:2271–9.
- [15] Brodeur A, Lé M, Sangnier M, Zylberberg Y. *Star Wars: The Empirics Strike Back* [Internet]. Rochester, NY: Social Science Research Network; 2012: Report No.: ID 2089580. Available at <http://papers.ssrn.com/abstract=2089580> Accessed June 11, 2015.
- [16] Ioannidis JPA. Excess significance bias in the literature on brain volume abnormalities. *Arch Gen Psychiatry* 2011;68:773–80.
- [17] Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant [Internet]. Rochester, NY: Social Science Research Network; 2011: Report No.: ID 1850704. Available at <http://papers.ssrn.com/abstract=1850704> Accessed February 11, 2015.
- [18] Centre for Open Science. *Open Science Framework* [Internet] 2015: Available at <https://osf.io/> Accessed February 11, 2015.