Modeling science trustworthiness under publish or perish pressure - Supplementary Material

David Robert Grimes[1,2], Chris T. Bauch[3], John P.A. Ioannidis[4]

[1] Queen's University Belfast, UK, BT7 1NN, UK
[2] University of Oxford, Old Road Campus, Oxford, OX3 7DQ, UK
[3] University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
[4] Meta-Research Innovation Center at Stanford (METRICS) / Departments of Medicine, Health Research and Policy, Biomedical Data Science, and Statistics, Stanford University, SPRC, MSOB X306, 1265 Welch Rd, Stanford, CA 94305, USA

d.r.grimes@qub.ac.uk

* E-mail: d.r.grimes@qub.ac.uk

Parameter estimation

To simulate the trends that would occur under these assumptions requires that we select appropriate parameters. These are detailed in table 1. which are used in all simulations unless otherwise stated in the text. It can be seen through inspection that discovery rate per unit resource D_R cancel in the analysis for $x(t)$,$y(t)$,$z(t)$ and $T(t)$, and accordingly this can be ascribed any real positive value without skewing analysis. When there is no fraud detection funding penalization ($\eta = 0$), journal carrying capacity J also cancels in the analysis and does not impact results. Initially we assume also that $G = 0$ so that funding levels remain constant. Estimation of fraudulent submission fraction per unit resource δ requires some elaboration, as this is notoriously difficult to ascertain and field-specific. A 1996 analysis by Fuchs and Westervelt [1] extrapolated from known cases to estimate that approximately 0.01% of published papers were fraudulent, though this is considered exceptionally conservative [2]. Empirical estimates of plagiarism vary markedly from 0.02% 25% of all publications [3].The frequency of paper retractions from the PubMed data base for misconduct is about 0.02%, suggesting that fraud might be present in 0.02%-0.2% of papers therein [4]. An investigation in the Journal of Cell Biology found inappropriate image-manipulation occurring in 1% of papers [5]. More alarmingly perhaps, a 1992 data audit by the United States Food and Drug Administration found deficiencies a in 1020% of studies published between 1977 and 1990, with 2% of investigators deemed guilty of severe scientific misconduct [6–8]. For the purposes of this work, we'll assume that FFP violations are present in 1% of the published literature so that $J_F = \frac{J}{100}$. Defining $x_o = x(0)$, $y_o = y(0)$, $z_o = z(0)$ and $\nu_{pO} = \nu_p(0)$, then for any selected values of D_R and J, we can readily define the required rate by

$$\delta = \frac{J_F(x_o S_{D+} + y_o S_{C+} + z_o D_R(p_T + \epsilon p_F))}{z_o D_R(BJ - J_F)} \tag{1}$$

This is dependent on the true / false positive of the field, and we initially take an optimistic assumption that the 1% published fraud occurs in fields with high levels of false positives, and will be less in fields with less ambiguity in results, so that the same value

Table 1: Parameters for initial simulations - Values in this table comprise the default initial assumptions, which are varied to investigate different conditions, as outlined in the respective relevant section.

Parameter	Value
Journal carrying capacity J	120 / cycle
Total discovery rate per unit funding D_R	15
Initial proportion Diligent researchers f_D	0.65 [7]
Initial proportion Careless researchers f_C	0.33 [7]
Initial proportion Unethical researchers f_U	0.02 [7]
Reasonable error rate ϵ	0.05 [9] [a]
Fraudulent submission rate per unit resource δ	0.0574 [b]
Positive publication bias B	0.9
Multiplicative factor for careless cohort - c	2
Null / Negative submission rates - $\beta_D/\beta_C/\beta_U$	0.40
Resource growth rate - G	0
Fraud detection proportion - η	0
Field-specific true positive fraction p_T	0.2
Field-specific false positive fraction p_F	0.2

[a] Strictly speaking, Prof. Colquhoun puts forward an eloquent argument in the cited work that $p < 0.05$ is a frequently abused metric, leading to false positives. For simplicity however, we'll presume that $\epsilon = 0.05$ reflects best reasonable practice in this simulation.

[b] See text for origin of this value and implicit assumptions.

of δ is used for all simulations. This is calculated assuming $p_F = 0.32$ and $p_T = 0.08$ so that $\delta = 0.057$ as per table 1.

Simplifications and other results

The model presented is a much simplified picture of reality, but it allows us to examine how different factors might influence the trustworthiness of published science, and potentially suggest strategies to improve it. As the motivations of and pressures on scientists are incredibly complex, its important to recognize the limitations in the model too. The three cohorts presented here would in reality constitute a spectrum. The sub-divisions in this work are relatively arbitrary and informed by the available data on researcher populations, though it would be easily possible to extend this to consider more subpopulations if desired Scientific conduct is notoriously difficult to quantify, and the assumptions we've used in this work reflect the best estimates to date [7].

We can also envision a situation where authors are awarded solely on the basis of positive findings, so that negative findings have no funding benefit. We can apply the model to these circumstances too, with the realization that under such a scheme, there would be no incentive for authors to submit negative results. In this case, $B = 1$ and all β terms reduce to zero. Essentially then, one gets a similar result to the one shown in paper figure 5(a), with even further reduction in trustworthiness. Finally, measures that can

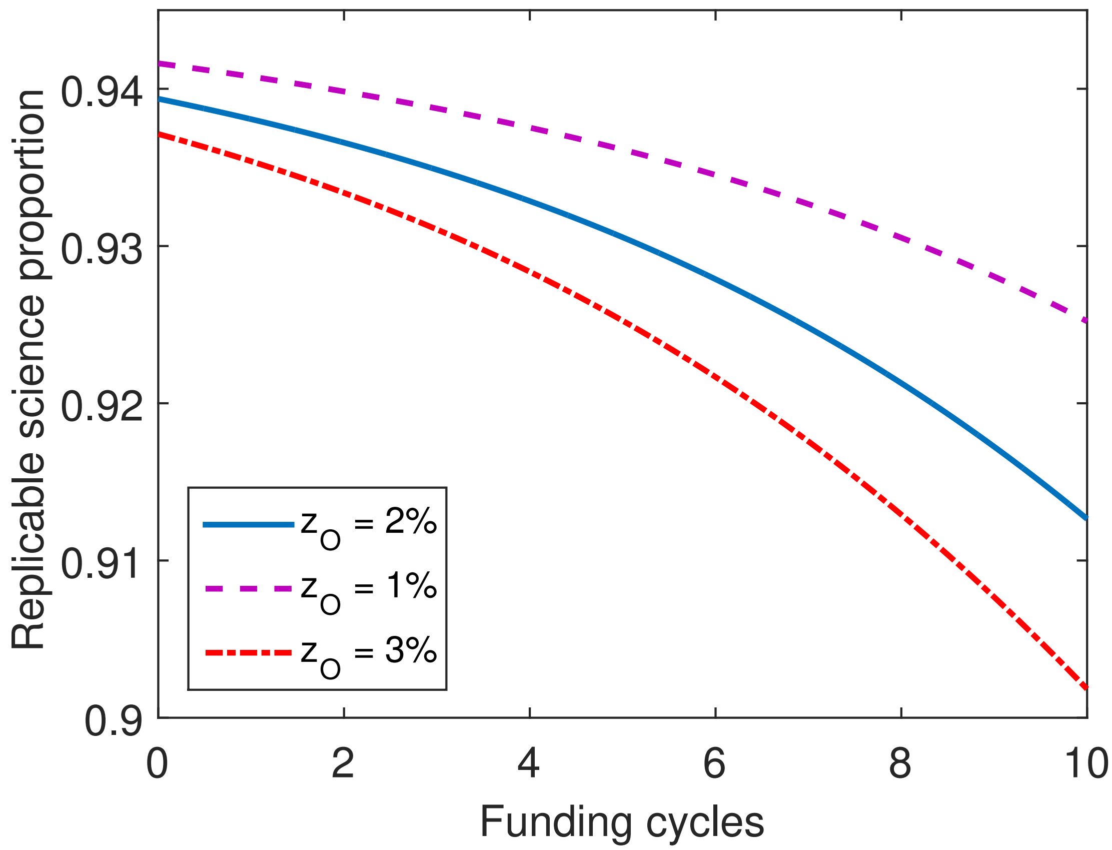

Figure 1: The trustworthiness of science in a field where $p_T = p_F = 0.2$, with varying values of z_O - the less funds are initially allocated to unethical cohorts, the better the resultant science trustworthiness.

be adopted to begin changing the culture of fixation on novel positive results include the establishment of awards by academic societies designed to recognize methodological rigor rather than positive results, as well as the explicit recognition of material published in online repositories as relevant material in university tenure and promotion guidelines.

It's also worth considering how the positive publication weighing might impact on the 'file-drawer' problem [10]. This was the observation first articulated by Rosenthal in 1979 that researchers tended to not invest their energy trying to publish null findings, instead burying them in a file-drawer. The great tragedy of this is that essential null results are often disregarded by the scientists who discover them, meaning others labor down fruitless avenues. In the model, we have implicitly assumed a version of this by assuming researchers submit only a portion of their negative findings (β) for consideration. It would be useful to know precisely how much is never submitted, and to gauge the extent of the file-drawer problem. One approach might be to consider the issue from an energy-expenditure point of view or game-theory approach which could be coupled with the model to estimate how much vital science never reaches the public domain, though this is beyond the scope of this investigation.

Impact of initial unethical funding proportion

Figure 1 depicts the sensitivity of trustworthiness to different assumptions of initial unethical funding proportion z_O. As might be expected, increasing z_O has negative implications for published trustworthiness.

Future extensions

A more sophisticated future analysis might include variables that respond to the available funding. For example, the fraudulent publication rate δ is treated as a constant in this work for the most part, but it is easy to imagine a situation where this increases with shrinking funding, or where the number of investigators willing to engage in such practices is a function of available funding. This is not considered here, but the model presented could be easily adapted to probe this further.

References

[1] S. Fuchs and S. D. Westervelt, "Fraud and trust in science," *Perspectives in biology and medicine*, vol. 39, no. 2, pp. 248–269, 1996.

[2] M. Franzen, S. Rödder, and P. Weingart, "Fraud: causes and culprits as perceived by science and the media," *EMBO reports*, vol. 8, no. 1, pp. 3–7, 2007.

[3] J. Giles, "Special Report: Taking on the cheats," *Nature*, vol. 435, no. 7040, pp. 258–259, 2005.

[4] L. D. Claxton, "Scientific authorship: Part 1. A window into scientific fraud?," *Mutation Research - Reviews in Mutation Research*, vol. 589, no. 1, pp. 17–30, 2005.

[5] N. Steneck, "Fostering integrity in research: Definition, current knowlege, and future directions," *Science and Engineering Ethics*, vol. 12, no. 1, pp. 53–74, 2006.

[6] J. L. Glick, "Scientific data auditA key management tool," *Accountability in Research*, vol. 2, no. 3, pp. 153–168, 1992.

[7] D. Fanelli, "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data," *PLoS ONE*, vol. 4, no. 5, 2009.

[8] D. Fanelli, "How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data," *PloS one*, vol. 4, no. 5, p. e5738, 2009.

[9] D. Colquhoun, "An investigation of the false discovery rate and the misinterpretation of P values," *Royal Society Open Science*, pp. 1–15, 2014.

[10] R. Rosenthal, "The file drawer problem and tolerance for null results.," *Psychological bulletin*, vol. 86, no. 3, p. 638, 1979.