

COMMENT

SOCIETY Is Steven Pinker too optimistic about our troubled times? **p.420**



BIOLOGY From fly room to fly boom: how labs filled up with *Drosophila* **p.421**

EARTH OBSERVATION History shapes where, when, why and how we collect data **p.423**

METRICS Abolish author position — it is too crude a guide to contribution **p.423**

ZOONAR GMBH/ALAMY



Efforts to replicate research studies are distorted by inherent conflicts between the authors of the original work and those trying to reproduce the results.

How to make replication the norm

The publishing system builds in resistance to replication. **Paul Gertler, Sebastian Galiani and Mauricio Romero** surveyed economics journals to find out how to fix it.

Replication is essential for building confidence in research studies¹, yet it is still the exception rather than the rule^{2,3}. That is not necessarily because funding is unavailable — it is because the current system makes original authors and replicators antagonists. Focusing on the fields of economics, political science, sociology and psychology, in which ready access to raw data and software code are crucial to replication efforts, we survey

deficiencies in the current system.

We propose reforms that can both encourage and reinforce better behaviour — a system in which authors feel that replication of software code is both probable and fair, and in which less time and effort is required for replication.

Current incentives for replication attempts reward those efforts that overturn the original results. In fact, in the 11 top-tier economics journals we surveyed, we could find only

11 replication studies — in this case, defined as reanalyses using the same data sets — published since 2011. All claimed to refute the original results. We also surveyed 88 editors and co-editors from these 11 journals. All editors who replied (35 in total, including at least one from each journal) said they would, in principle, publish a replication study that overturned the results of an original study. Only nine of the respondents said that they would consider publishing a replication ▶

► study that confirmed the original results.

We also personally experienced antagonism between replicators and authors in a programme sponsored by the International Initiative for Impact Evaluation (3ie), a non-governmental organization that actively funds software-code replication. We participated as authors of original studies (P.G. and S.G.) and as the chair of 3ie's board of directors (P.G.).

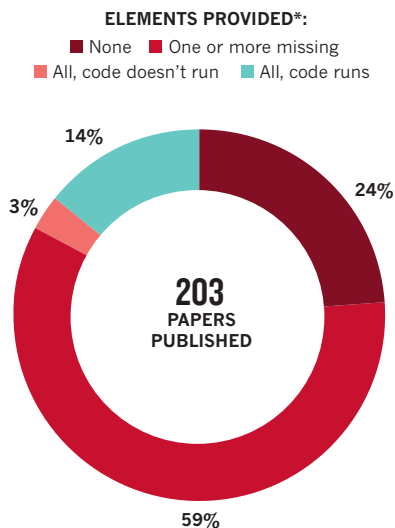
In our experience, the programme worked liked this: 3ie selected influential papers to be replicated and then held an open competition, awarding approximately US\$25,000 for the replication of each study⁴. The organization also offered the original authors the opportunity to review and comment on the replications. Of 27 studies commissioned, 21 were completed, and 7 (33%) reported that they were unable to fully replicate the results in the original article. The only replication published in a peer-reviewed journal⁵ claimed to refute the results of the original paper.

Despite 3ie's best efforts, adversarial relationships developed between original and replication researchers. Original authors of five studies wrote in public comments that the replications actively sought to refute their results and were nitpicking.

One group stated that the incentives of replicators to publish "could lead to overstatement of the magnitude of criticism" (see go.nature.com/2gecz3b). Others made similar points. Although one effort replicated all the results in the original paper, the originating authors wrote, "we disagree with the unnecessarily aggressive tone of some statements in the replication report (particularly in the abstract)" (see go.nature.com/2esdjkr).

REPLICATION RARELY POSSIBLE

An analysis of 203 economics papers found that fewer than one in seven supplied the materials needed for replication.



*The elements assessed were raw data, raw code, estimation data and estimation code.

DATA CHECKED?

In a survey of 67 journals, most of the political-science and top-tier economics titles required authors to submit software code and data to editors before publication. Journals in sociology and psychology rarely did so.



Another group felt that "the statement that our original conclusions were robust was buried" (see go.nature.com/2siufet).

The 3ie-sponsored replication⁶ of one highly cited paper⁶ resulted in an acrimonious debate that became known as the Worm Wars. Several independent scholars speculated that assumptions made by the replicators had more to do with overturning results than with any scientific justification.

REPLICATION COSTS

A first step to getting more replications is making them easier. This could be done by requiring authors to publicly post the data and code used to produce the results in their studies — although this in itself would not redress the incentive to gain a publication by overturning results.

Ready access to data and code would ease replication attempts. In our survey of journal websites, the mid-tier economics journals (as ranked by impact factor) and those in sociology and psychology rarely asked for these resources to be made available. By contrast, almost all of the top-tier journals in economics have policies that require software code and data to be made available to editors before publication.

This is also true of most of the political-science journals we assessed, and all three of the broad science journals in our survey (see Supplementary information). In addition, many of the top-tier economics journals explicitly ask authors to post raw data as well as estimation data — the final data set used to produce the results after data clean-up and manipulation of variables. These are usually placed on a publicly accessible website that is designated or maintained by the journal. There seems to be no such norm in psychology and

sociology journals (see 'Data checked?').

To see how often the posted data and code could readily replicate original results, we attempted to recreate the tables and figures of a number of papers using the code and data provided by authors. Of 415 articles published in 9 leading economics journals in May 2016, 203 were empirical papers that did not contain proprietary or otherwise restricted data. We checked these to see which sorts of files were downloadable and spent up to four hours per paper trying to execute the code to replicate the results (not including code runtime).

We were able to replicate only a small minority of these papers. Overall, of the 203 studies, 76% published at least one of the 4 files required for replication: the raw data used in the study (32%); the final estimation data set produced after data cleaning and variable manipulation (60%); the data-manipulation code used to convert the raw data to the estimation data (42%), but only 16% had both raw data and usable code that ran); and the estimation code used to produce the final tables and figures (72%).

The estimation code was the file most frequently provided. But it ran in only 40% of these cases. We were able to produce final tables and figures from estimation data in only 37% of the studies analysed. And in only 14% of 203 studies could we do the same starting from the raw data (see 'Replication rarely possible'). We do not think that spending significantly more time would have boosted our success — usually the code called for data or variables that had not been supplied.

Tellingly, the tables and figures could almost always be replicated if the code ran without major modifications. This was a significant hurdle: only 54% and 61% of articles

had data-manipulation code or estimation code, respectively, that did not require major modifications.

Our results align with previous findings in the literature². A study of the articles published in the *Journal of Money, Credit and Banking* found that only 35.7% of articles met data-archive requirements, and only 20% of studies could be replicated using the information available in the archive⁷. Another study attempted to replicate 67 papers published in 13 well-regarded general-interest and macroeconomics journals, and could replicate only 29 (ref. 8). This problem goes beyond economics. In 2013, only 18 of 120 political-science journals had replication policies⁹, and a 2016 study found that only 58% of the articles in top political-science journals publish their data and code¹⁰.

However, some progress is being made. Notably, a few political-science journals verify that posted data and code produce the results in a publication. A handful of journals in statistics and information sciences have appointed ‘reproducibility editors’ to ensure that analyses can be replicated. The American Economic Association appointed a data editor last year to oversee reproducibility in its journals.

“Usually the code called for data or variables that had not been supplied.”

CREDIBLE THREAT

We think that the way forward is for more journals to take on this kind of responsibility, using data editors to help implement the following replication policy.

Journals could oversee the replication exercise after conditional acceptance of a manuscript but before publication. Journals would then verify that all raw data used in the paper and code (that is, sample and variable construction, as well as estimation code) are included and executable. They would then commission academic experts, advanced graduate students or their own staff to verify that the code reproduces the tables and figures in the article. If not, editors could ask authors to correct their errors and, if necessary, have papers reviewed again.

In addition, for a random sample of papers, journals should attempt to reconstruct the code from scratch or search the executable code for errors. In this way, all papers would have some positive probability of being fully replicable.

This simple procedure has the winning combination of four desirable characteristics. First, it is unbiased: editors would have no incentive to overturn results. Second, it creates a credible expectation in authors that their work will be replicated, motivating

them to be careful and to put effort into constructing their code and not report false results. Third, data and code would be easily available to outside researchers to explore the robustness of the original results using alternative specifications, measurements and methods of identification and estimation. Fourth, there is little cost associated with getting a research associate to perform replication exercises, especially because authors have strong incentives to cooperate with journals at the pre-publication stage. Although we are writing from our experience as economists, we think that similar practices could be adopted in other disciplines, particularly the social sciences.

Initially, these steps might slow down the time from acceptance to publication for some papers. However, authors will eventually internalize them and submit accurate, error-free materials so that the study replication will be done efficiently. This will help to restore confidence in the credibility of science. ■

Paul Gertler is a professor of economics in the Haas School of Business at the University of California, Berkeley, USA. **Sebastian Galiani** is a professor of economics at the University of Maryland, College Park, USA. **Mauricio Romero** is a graduate student in economics at the University of California, San Diego, USA.
e-mail: gertler@berkeley.edu

1. Nosek, B. A. et al. *Science* **348**, 1422–1425 (2015).
2. Christensen, G. & Miguel, E. Preprint at <http://dx.doi.org/10.17605/OSF.IO/9A3RW> (2017).
3. Berry, J., Coffman, L. C., Hanley, D., Gihleb, R. & Wilson, A. J. *Am. Econ. Rev.* **107**, 27–31 (2017).
4. International Initiative for Impact Evaluation. *3ie Replication Programme* (3ie, 2017); available at <http://go.nature.com/2eoe7o5>
5. Davey, C., Aiken, A. M., Hayes, R. J. & Hargreaves, J. R. *Int. J. Epidemiol.* **44**, 1581–1592 (2015).
6. Miguel, E. & Kremer, M. *Econometrica* **72**, 159–217 (2004).
7. McCullough, B. D., McGeary, K. A. & Harrison, T. D. *J. Money Credit Bank.* **38**, 1093–1107 (2006).
8. Chang, A. C. & Li, P. *Finance Econ. Disc. Ser.* <http://dx.doi.org/10.17016/FEDS.2015.083> (2015).
9. Gherghina, S. & Katsanidou, A. *Eur. Polit. Sci.* **12**, 333–349 (2013).
10. Key, E. M. *Polit. Sci. Politics* **49**, 268–272 (2016).

Supplementary information and funding details accompany this article online; see go.nature.com/2gejxbd.

CORRECTION

The Comment article ‘Join the disruptors of health science’ (*Nature* **551**, 23–26; 2017) should have disclosed that *Nature’s* Editor-in-chief, Philip Campbell, serves on an unpaid basis as a member of the science advisory board for the start-up firm Mindstrong Health.

CORRECTION

In the Comment 'How to make replication the norm' (*Nature* **554**, 417–419; 2018) mistakenly stated that only authors of non-replicated articles from the 3ie project described antagonism. In fact, authors of both replicated and non-replicated studies did. Also, the number of replications was 21, not 20. The data set is now available as supplementary information.