The international journal of science / 16 December 2021

nature

Replicating scientific results is tough — but crucial

A high-profile replication study in cancer biology has had disappointing results. Scientists must redouble their efforts to find out why.

Replicabillity — the ability to obtain the same result when an experiment is repeated — is foundational to science. But in many research fields it has proved difficult to achieve. An important and much-anticipated brace of research papers now show just how complicated, time-consuming and difficult it can be to conduct and interpret replication studies in cancer biology[1,2] (see page 368).

Nearly a decade ago, research teams organized by the non-profit Center for Open Science in Charlottesville, Virginia, and ScienceExchange, a research-services company based in Palo Alto, California, set out to systematically test whether selected experiments in highly cited papers published in prestigious scientific journals could be replicated. The effort was part of the high-profile Reproducibility Project: Cancer Biology (RPCB) initiative. The researchers assessed experimental outcomes or 'effects' by seven metrics, five of which could apply to numerical results. Overall, 46% of these replications were successful by three or more of these metrics, such as whether results fell within the confidence interval predicted by the experiment or retained statistical significance.

The project was launched in the wake of reports from drug companies that they could not replicate findings in many cancer-biology papers. But those reports did not identify the papers, nor the criteria for replication. The RPCB was conceived to bring research rigour to such retrospective replication studies.

Initial findings

One of the clearest findings was that the effects of an experimental treatment — such as killing cancer cells or shrinking tumours — were drastically smaller in replications, overall 85% smaller, than what had been reported originally. It's hard to know why. There could have been statistical fluke, for example; bias in the original study or in the replication; or lack of know-how by the replicators that caused the repeated study to miss some essential quality of the original.

The project also took more than five years longer than expected, and, despite taking the extra time, the teams were able to assess experiments in only one-quarter of the experiments they had originally planned to cover. This underscores the fact that such assessments take much more time and effort than expected.

> **"Researchers, research funders and publishers must take replication studies much more seriously."**

The RPCB studies were budgeted to cost US$1.3 million over three years. That was increased to $1.5 million, not including the costs of personnel or project administration.

None of the 53 papers selected contained enough detail for the researchers to repeat the experiments. So the replicators had to contact authors for information, such as how many cells were injected, by what route, or the exact reagent used. Often, these were details that even the authors could not provide because the information had not been recorded or laboratory members had moved on. And one-third of authors either refused requests for more information or did not respond. For 136 of the 193 experimental effects assessed, replicators also had to request a key reagent from the original authors (such as a cell line, plasmid or model organism) because they could not buy it or get it from a repository. Some 69% of the authors were willing to share their reagents.

Openness and precision

Since the reproducibility project began, several efforts have encouraged authors to share more-precise methodological details of their studies. *Nature*, along with other journals, introduced a reproducibility checklist in 2013. It requires that authors report key experimental data, such as the strain, age and sex of animals used. Authors are also encouraged to deposit their experimental protocols in repositories, so that other researchers can access them.

Furthermore, the 'Landis 4' criteria were published in 2012 to promote rigorous animal research. They include the requirement for blinding, randomization and statistically assessed sample sizes. Registered Reports, an article format in which researchers publish the design of their studies before doing their experiments, is another key development. It means that 'null effects' are more likely to be published than buried in a file drawer. The project team found that null effects were more likely to be replicated; 80% of such studies passed by three metrics, compared with only 40% of 'positive effects'.

Harder to resolve is the fact that what works in one lab might not work in another, possibly because of inherent variation or unrecognized methodological differences. Take the following example: one study tracked whether a certain type of cell contributes to blood supply in tumours[3]. Tracking these cells required that they express a 'reporter' molecule (in this case, green fluorescent protein). But, despite many attempts and tweaks, the replicating team couldn't make the reporter sufficiently active in the cells to be tracked[4], so the replication attempt was stopped.

The RPCB teams vetted replication protocols with the original authors, and also had them peer reviewed. But detailed advance agreement on experimental designs will not necessarily, on its own, account for setbacks encountered when studies are repeated — in some cases, many years after the originals. That is why another approach to replication is used by the US Defense Advanced Research Projects Agency (DARPA). In one DARPA programme, research teams are assigned independent verification teams. The research teams must help to troubleshoot and provide support for the verification teams so that key

nature

results can be obtained in another lab even before work is published. This approach is built into programme requirements: 3–8% of funds allocated for research programmes go towards such verification efforts[5].

Such studies also show that researchers, research funders and publishers must take replication studies much more seriously. Researchers need to engage in such actions, funders must ramp up investments in these studies, and publishers, too, must play their part so that researchers can be confident that this work is important. It is laudable that the press conference announcing the project's results included remarks and praise by the leaders of the US National Academies of Sciences, Engineering, and Medicine and the National Institutes of Health. But the project was funded by a philanthropic investment fund, Arnold Ventures in Houston, Texas.

The entire scientific community must recognize that replication is not for replication's sake, but to gain an assurance central to the progress of science: that an observation or result is sturdy enough to spur future work. The next wave of replication efforts should be aimed at making this everyday essential easier to achieve.

1. Errington, T. M., Denis, A., Perfito, N., Iorns, E. & Nosek, B. A. *eLife* **10**, e67995 (2021).
2. Errington, T. M. *et al. eLife* **10**, e71601 (2021).
3. Ricci-Vitiani, L. *et al. Nature* **468**, 824–828 (2010).
4. Errington, T. M. *et al. eLife* **10**, e73430 (2021).
5. Raphael, M. P., Sheehan, P. E. & Vora, G. J. *Nature* **579**, 190–192 (2020).

Gender balance at Nature Conferences: an update

Nature has made progress in improving representation and participation of women at scientific conferences – but there is much more to do.

At the end of 2019, *Nature* pledged to work harder to help to address the entrenched gender inequity at scientific conferences (*Nature* **576**, 182; 2019). We looked closely at gender diversity at Nature Conferences (conferences curated by editors of the Nature Portfolio journals), and what we saw was simply not good enough. We introduced a code of conduct, including pledges to have no all-male panels and to invite an equal percentage of women (including all those who identify as women) and men as speakers at all our conferences.

Two years on and these decisions have yielded results. Women comprised 29% of keynote speakers at Nature Conferences between 2016 and 2018 (15 out of 51 speakers across 27 events). That number increased to 48% (14 out of 29) in the 20 meetings that took place in 2020 and 2021.

> **"*Nature* is committed to achieving inclusivity for the good of science and society."**

During the same period, women represented 43% of conference organizing committees (59 out of 130) and 51% of speakers for short talks (33 of 65). It's an encouraging start, but the trends cannot stop there. In 2021, all Nature Conferences were virtual, like almost all other scientific events. It's early days, but initial reports suggest that online formats can be more inclusive than in-person events (S. Sarabipour *eLife* **9**, e62668; 2020). It is absolutely essential that these modest gains are not reversed once in-person events return.

Best-practice guidance is being refined all the time. The advocacy group 500 Women Scientists is working with several major scientific organizations, including the Aspen Global Change Institute, the American Geophysical Union, Colorado State University, the Earth Science Women's Network, Georgia Sea Grant and the team behind the virtual seminar series Pal(a)eoPERCS, to update their inclusive scientific meetings guide (see go.nature.com/3ilz3e5). The guide aims to share good practice, including tools to help ensure that events are more inclusive. Nature Conferences will strive to use these tools, and we hope that other conference organizers will too.

Data from the UK Society for Endocrinology's annual national conference are the latest to show that even when meetings have roughly equal numbers of male and female delegates, women attendees participate less and tend to ask fewer and shorter questions (V. Salem *et al. Lancet Diabetes Endocrinol.* **9**, 556–559; 2021). A simple intervention improved things: when conference organizers were asked to have more female session chairs, and those chairs were asked to invite opening questions from women wherever possible, questions from women rose to 35%, from 24% the previous year.

"If women are not visible at conferences, they cannot act as role models for junior academics, creating a self-perpetuating cycle," the paper's authors point out. Nature's insistence on gender parity, where possible, at its conferences is just one step towards encouraging more women to take part in their communities' discussions.

Codes of conduct for scientific meetings, such as that introduced by Nature Conferences and others before us, are crucial. They are becoming the norm, with many large learned/academic societies making commitments to improve diversity across their activities – and there are encouraging signs that they are starting to pay off.

Codes of conduct need to be enforced, and there need to be checks in place to ensure that they are followed and that they continue to be effective at truly improving diversity – not just in attendance, but in participation, too. We also acknowledge that our efforts need to be broader, and not focused solely on gender.

There's a very long way to go to achieve full equity, inclusion and diversity at scientific conferences. The Nature journals are committed to achieving inclusivity for the good of science and society. We are proud to have made a small change and understand and accept that there is much more that we can and must do.

Correction

The article 'Replicating scientific results is tough — but crucial' originally mischaracterized the RPCB's analysis of replication attempts. Rather than recording seven experimental outcomes, it assessed experimental effects using seven metrics, and it also assessed 193 experimental effects not 193 experiments.