

Sluggish data sharing hampers reproducibility effort

Initiative trying to validate 50 cancer papers finds difficulty in accessing original study data.

Richard Van Noorden

03 June 2015

RIO DE JANEIRO, BRAZIL

An initiative that aims to validate the findings of key cancer papers is being slowed by an unexpected hurdle — problems accessing data from the original studies.

The [Reproducibility Initiative: Cancer Biology](#) consortium aims to repeat experiments from 50 highly-cited studies published in 2010–12 in journals such as *Nature*, *Cell* and *Science*, to see how easy it is to reproduce their findings. Although these journals require authors to share their data on request, it has taken two months on average to get the data for each paper, said William Gunn, a co-leader of the project, at the [4th World Conference on Research Integrity](#) in Rio de Janeiro, Brazil, on 3 June.

For one paper, securing the necessary data took a year. And the authors of four other papers have stopped communicating with the project altogether. In those instances, the journals that published the studies are stepping in to remind researchers of their responsibilities.

Most authors were happy to collaborate with those seeking to validate their findings, but it has taken longer than expected to locate the relevant data. “There’s no interesting reason why it takes so long: academics move around, they don’t keep records of where their data is, or they have to go through old lab journals, or re-analyse files which are in an old format,” said Gunn. “It’s a powerful argument that scientists should deposit their data at the time they submit a manuscript.”

It has also been difficult to identify the laboratory resources that researchers used in their papers, Gunn said. From references given in papers alone, for instance, the replication effort has found less than half the antibodies used can be uniquely identified. He suggests that scientists make clear which resource they used, using standard formats such as the [Research Resource Identifier](#).

Problem hotspots

So far, the Reproducibility Initiative has obtained data for 31 of the 50 papers. Some validation experiments have already been conducted, each at a cost of around US\$25,000 in materials and time. Costs are closer to \$35,000 for mouse studies, Gunn said. The project has a [\\$1.3-million grant](#) from the Laura and John Arnold Foundation of Houston, Texas, to do its work.

Rather than declare a success or failure when it tries to validate papers, the initiative will instead report the statistical significance of the result when it combines its data with that of the original paper — much as a meta-analysis combines the results of different data sets. The project, says Gunn, aims to find ‘hotspots’ for problems in reproducing existing findings — such as whether researchers fail to accurately describe their methods in their original papers, or whether particular types of experiments are prone to difficulties in replication. Its first results are expected towards the end of 2015.

That is a more nuanced approach than is taken by an investigation¹ that is often cited as evidence that biomedical science has a reproducibility problem, in which researchers at the biotech company Amgen, of Thousand Oaks, California, said that they could not confirm the findings of 47 out of 53 ‘landmark’ papers. Because no one knows which papers were investigated, why their findings could not be replicated, or how stringent the replication standard was, it is hard to tell the true extent of the problem, Gunn said.

Gunn says that journals are starting to engage with the problem of reproducibility much more seriously. “From when we started to now, there has really been a dramatic shift,” he said.

Nature | doi:10.1038/nature.2015.17694

References

1. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).