Open Science Monitoring

Impact Case Study – Reproducibility Project
Author information and acknowledgements

This study was delivered by RAND Europe, with the support of Deloitte and Digital Science & Research Solutions, including Altmetric.com and figshare. The project was commissioned by the European Commission Directorate-General for Research and Innovation and received funding under specific contract 2015/RTD/A6/SC/PP-02741-2015.¹

The team at RAND Europe included: Elta Smith (project leader), Salil Gunashekar (project manager), Sarah Parks, Catherine Lichten, Louise Lepeitit, Molly Morgan Jones, Caterina Manville and Calum MacLure. The team at Deloitte included: Valentina Cilli, Maruca De Ramón, Vilma Zotou, Benoit Vandresse and Sebastiaan van der Peijl.

Data sources

We would like to thank the following organisations for generously contributing their data for use on the monitor: 101innovations, Altmetric, arXiv, bioRxiv, Clarivate Analytics, F1000Research, figshare, GenBank, Helmholtz-Centre for Environmental Research – UFZ and German Centre for Integrative Biodiversity Research (iDiv), Nature Publishing Group and Palgrave Macmillan, OpenAIRE, PeerJ preprints, Publons, re3data.org, RePEc, SHERPA/RoMEO, SHERPA/JULIET, Taylor & Francis, and Wiley.

Expert panel

We are grateful to our expert panel members who provided helpful advice and guidance: Aletta Bonn (German Centre for Integrative Biodiversity Research (iDiv) and Helmholtz Center for Environmental Research (UFZ), Friedrich-Schiller-Universität Jena, Germany); Thomas Crouzier (School of Biotechnology, Royal Institute of Technology (KTH), Sweden); Ivo Grigorov (National Institute of Aquatic Resources, Technical University of Denmark); Peter Kraker (Postdoctoral Fellow, Know-Center Social Computing Group, Graz University of Technology, Austria); Eleni Malliou (National Documentation Center, Greece); and Paul Wouters (CWTS-Centre for Science and Technology Studies, Leiden University, The Netherlands).

Acknowledgements:

We would like to thank the following individuals for their contributions to the monitor: Susan Guthrie (RAND Europe), Gavin Cochrane (RAND Europe), Alexandra Pollitt (King’s College London), and Marion Bywater who acted as our quality assurance reviewers and provided thoughtful and timely comments on our analyses; Jessica Plumridge (RAND Europe) who designed the graphics presented in the online monitor; Monica Hertzman (RAND Corporation) who provided technical assistance in relation to the design of the website; Steven Wooding (RAND Europe) who provided advice and support throughout the study; Vincent Larivière (Observatoire des Sciences et des Technologies) for his insights and advice; Joanna Chataway (University of Sussex) for her advice and support; Rebecca Ioppolo (RAND Europe) who provided research assistant support; Jack Melling (RAND Europe) who helped in the dissemination activities related to the launch of the monitor; and the many individuals and organisations who participated in an online consultation we conducted as part of the development of the monitor.

For more information about this document or the Open Science Monitor, please contact:

Elta Smith
RAND Europe | Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
Telephone: +44 (1223) 353 329 | E-mail: eltas@rand.org

Reproducibility Project

Summary

The Reproducibility Project aimed to systematically test the extent to which psychology studies are reproducible, and to do so in an open, transparent way. It consisted of a collaborative effort to repeat 100 different psychology studies through the efforts of 270 collaborators affiliated to 125 institutions and departments around the world. Brian Nosek, a professor of psychology at the University of Virginia who now leads the Virginia-based Center for Open Science, coordinated the project. It was funded by the USA-based Laura and John Arnold Foundation.

Motivating the study was the issue that, although reproducibility is considered a fundamental characteristic of science, evidence suggested that a large proportion of findings published in the scientific literature were not reproducible. However, there was little data available on which replications were attempted, how they were done and what the results were. The Reproducibility Project assessed the situation in the field of psychology – an area of particular concern regarding replicability – by systematically attempting to replicate published studies and exploring factors that affect replicability.

The project aimed to be a model for a fully open methodology, while also assessing reproducibility. This meant that it involved open sharing of research designs and protocols, interactions between the original researchers and those who were replicating their studies, and reuse of original materials, where possible. Raw data and reports on the replications were also made publicly available.

The study findings were striking – only about 40 per cent of 100 original findings could be replicated – and the work prompted discussion and debate about the presence of systemic problems in psychology and other fields, and about how to address them. At the same time, researchers who were involved in the project are helping to drive change in policy and practice. Funders have begun discussing and introducing new initiatives to support reproducibility research, researchers have begun new reproducibility projects in other areas, and journals have introduced policies to support reproducibility and research transparency. While the Reproducibility Project has not been the only factor driving these changes, frequent references to the project and the involvement of project researchers in new initiatives indicate that it is playing an important role.

Background

The Reproducibility Project came about amid a flurry of doubts about the reliability of published scientific findings, particularly in psychology. In an often-cited paper published in 2005, statistician and biomedical researcher John Ioannidis described how it is likely that ‘most published research findings are false’ due to poor study design and biased analysis and reporting (Ioannidis 2005). This finding is consistent with other studies: an attempt to replicate studies on potential drug targets found just one quarter of 67 attempts were successful (Prinz et al. 2011); only 11 per cent of landmark cancer treatment studies (6 out of 53) were reportedly replicable (Begley & Ellis 2012); and a survey of researchers in psychology indicated that only about half of the replications the respondents had attempted were successful (Hartshorne & Schachner 2012).

2 While there was a particular focus on psychology, concerns were not limited to this field. Cases of fraud have arisen in other fields, including stem cell research and cell biology (Jha 2012), as have concerns about the reliability of findings (Ioannidis 2005) and publication bias (Fanelli 2012, 2010).
Other events happening around this time increased these concerns about science. One was the high-profile misconduct cases of Dutch psychologist Diederik Stapel, who was dismissed in 2011 from Tilburg University after being found to have fabricated or manipulated data in at least 55 publications (Levelt Committee et al. 2012). A failure to replicate influential psychology studies on priming – an effect where exposure to one stimulus can unconsciously affect one's response to another, which is an area related to Stapel's work – drew further concerns about the reliability of psychology findings.

With these events came debate about whether there is a ‘crisis’ in psychology (Pashler & Harris 2012). The events prompted prominent psychologist Daniel Kahneman to call for replication studies to address the concerns about credibility (Yong 2012). And while these types of events were not new to science, they incited a strong response that at least one scholar (Spellman 2015) described as a ‘revolution’. One development she identified as part of it is the Reproducibility Project.

The main rationale for the Reproducibility Project was that replication, despite its importance for science, is not a prominent part of scientific practice. A paper describing the project defined reproducibility as the ‘extent to which consistent results are observed when scientific studies are repeated’ (Open Science Collaboration 2012, 657) and cited evidence that the level of reproducibility of published findings may be surprisingly low. Amid growing concerns about the reliability of science and the limited amount of evidence available about which studies cannot be reproduced and why, the Reproducibility Project was set up to systematically test the reproducibility of findings from psychology (ibid.).

Nosek and colleagues, who conducted the Reproducibility Project, attributed the limited focus on replication to the presence of strong incentives for scientists to produce new ideas and relatively weak incentives to carry out replication (ibid.). They have described a ‘disconnect’ between the professional rewards that shape the behaviour of individual scientists and what is good for science as a whole, arguing that self-correction does not occur consistently in science and that invalid findings may go unchallenged for a long time (Nosek et al. 2012). To address this issue, they identified a need for systemic and cultural changes and called for increased openness in scientific practice.

The Reproducibility Project was funded by the Laura and John Arnold Foundation, which became interested in the quality and reproducibility of research through news about the failed attempt to replicate the influential priming studies (Buck, pers.comm. 2016). The organisation aims to base its decisions on data and evidence. It therefore believes that the reliability of this evidence is crucial, and it chose to shift part of its focus to improving research (ibid.). The organisation then began to support work by Nosek, including the Reproducibility Project.

---

3 Although concerns related to scientific misconduct helped fuel discussions about systemic problems in psychology, individuals involved in doing and funding work on reproducibility are keen to point out that the aim of such projects is not to identify misconduct or fraud. Rather, the goal is to gather evidence about reproducibility and to look at ways of improving the reliability of scientific findings. Indeed, the 2012 paper introducing the Reproducibility Project identifies several different reasons why a finding may not be reproducible. Aside from the original finding being false, a failure might arise due to a flawed study design, because the actual effect size is smaller than first reported, because the replication methodology differed from the original in an important way (that may have previously been unappreciated), or simply by chance (Open Science Collaboration 2012).

4 Focusing on psychology, Spellman (2015) identifies four factors that have helped bring about this response: changes in technology, changes in demographics in academia, competition among researchers to publish and win grants, and incentives to publish novel findings being much stronger than those to publish accurate, rigorous incremental work.
Methods and the role of open science

A salient feature of the Reproducibility Project was its open methodology. Building on the idea that a lack of transparency and accountability contributes to biases in the literature, the project had a fully open design. In addition to the announcement of the project in the 2012 paper, the project’s design and protocol were made available online (Open Science Collaboration 2016), along with raw data and reports from the replications. The researchers doing the replications also worked with the authors of original studies to obtain materials and get feedback on their protocols (Open Science Collaboration 2015). The Reproducibility Project was intended as a study to explore the reproducibility of findings, but it was also intended to act as a model for conducting a fully transparent research project. It made use of the Open Science Framework (OSF), an open source software platform that is now being further developed as a tool for collaboration and other research activities (Carp 2014). Together, the Reproducibility Project and OSF served as the foundation for the creation of the Center for Open Science, which has continued initiatives to encourage openness and reproducibility in research (Errington, pers.comm. 2016).

A follow-on project, Reproducibility Project: Cancer Biology (RP:CB), which builds on the open approach developed in the Reproducibility Project, is also being coordinated by the Center for Open Science (Errington et al. 2014). RP:CB includes some adjustments to the original methodology. Whereas the Reproducibility Project involved a step where study designs were pre-defined, reviewed by the authors of the original study and then made available online, the RP:CB goes a step further by using Registered Reports (Errington, pers.comm. 2016). In this approach, the protocol is published and peer reviewed prior to the research being conducted. Expected benefits of using Registered Reports are that the findings can be published regardless of the outcome, thereby reducing publication bias, and that researchers receive feedback to improve the quality of their study design (Royal Society Publishing 2017). The RP:CB also works with the Science Exchange, a company that sources labs that can be contracted to do research.

Outputs and findings

Two academic publications are direct outputs of the Reproducibility Project.

The first paper, published in 2012 in the journal Perspectives on Psychological Science (Open Science Collaboration 2012), described the project, the motivations for it and the methodology being used.

The second academic paper, published in 2015 in the journal Science (Open Science Collaboration 2015), reported the results. The project found that, among the 100 replications carried out, just 39 were judged to have replicated the original result. The project assessed replicability in multiple ways, including comparison of effect sizes, calculation of statistical significance and a subjective judgement by the research teams. The conclusion that 39 per cent of the replications were successful is based on the subjective assessment. While 97 of the original studies had statistically significant results (p<0.05), just 36 of the replications did. An important aim of the study was to go beyond an assessment of the level of reproducibility by also exploring which factors contribute to it. The study found that a larger effect size was a better predictor of replicability than were other factors, such as perceived importance of the effect and expertise of the original team or the team doing the replication (ibid.). It also found that surprising effects were less reproducible.

Impacts

The Reproducibility Project has attracted attention, inspired further research, and sparked dialogue and changes in policy and scientific practice. The research, publishing and funding communities have taken a strong interest in efforts to encourage reproducibility. While engagement has focused on the need for reliable research, openness – including the sharing
of protocols, materials and data, and the use of fully transparent methodology – has been central to discussions and changes, and was a crucial element of the approach exhibited in the Reproducibility Project.

Within academia, the two papers published about the project are among the most cited in their field for their age, and papers have been published that reanalyse and discuss the project’s findings (Anderson et al. 2016; Etz & Vandekerckhove 2016; Gilbert et al. 2016). In addition, at least three other reproducibility projects that have been started or are under development in other fields build on the example and methodology set out by the Reproducibility Project (Buck, pers.comm. 2016; Errington, pers.comm. 2016).5

The 2015 paper is among the most popular scientific articles tracked by Altmetric,6 ranking 45th out of more than 6 million, and ranking 5th among articles published in 2015. The project has also attracted attention beyond academia. The article was cited in 160 news stories in such outlets as Slate, the New York Times, 20 minutes, El pais, The Guardian and Wired, and it was tweeted about by nearly 2,000 users (Altmetric 2015).

In terms of impacts on policy and practice, both publishers and funders have made changes. Many, if not most, of the editorials and commentaries written in the scientific literature about improving scientific reproducibility reference the project (Buck, pers.comm. 2016). Changes have been made to journal policy and practice, particularly in psychology journals,7 to better support reproducibility. These changes include support for practices that were used in the Reproducibility Project, for instance, by requiring or acknowledging the sharing of data materials and requiring study preregistration.

Among funders, the Laura and John Arnold Foundation has continued to support the Center for Open Science and is also funding the RP:CB and a reproducibility project in the area of transcranial direct stimulation (ibid.). The funders indicate that they may continue supporting work on reproducibility, but they have said this type of work should become part of standard scientific practice (ibid.). They have therefore called on other funders – both major public funders and private philanthropy – to support replication studies (Buck 2015).

Public funders have taken an interest in this area. In 2016, the Netherlands Organisation for Scientific Research (known by acronym NWO) announced they would make €3 m available for a pilot programme of funding for researchers to replicate studies in social sciences or health research (NWO 2016). While NWO may be the first public funder to run such a call, NWO staff said that many funders are thinking about what they can do in this area, but are grappling with such questions as what specifically to fund, because some studies are more important to replicate than others (Hillebrink and Macor, pers.comm. 2016). NWO has focused on what it calls ‘cornerstone research’, which has particular impacts on further scientific work, policy and public debate. It invites researchers to identify the findings they would replicate and to present an argument about why replication should be done (ibid.). The first of the three planned annual rounds of pilot funding opened in September 2016. It offers two streams of funding, for either full replications or reanalysis of data from original studies (NWO 2016).

In the USA, the National Institutes of Health (NIH) included reproducibility as a focus area in their strategic plan for the period 2016-2020, saying that they, together with a range of

---

5 In cancer biology, transcranial direct stimulation and tropical ecology.

6 Altmetric.com tracks article mentions on such sources as social media, news sources and blogs.

stakeholders, have ‘launched a series of initiatives to enhance the rigor and reproducibility of the conduct of science and reporting of scientific results. Each of these initiatives will be evaluated over the next 5 years for beneficial effects, as well as for any unanticipated, negative consequences’ (NIH 2015, 44). The National Science Foundation has also taken an interest in this area, publishing a report titled Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science (Cacioppo et al. 2015) and opening a call for proposals related to the reproducibility, robustness and reliability of research in 2016 (Cook 2016).

Among other funders, the Prostate Cancer and Movember foundations have also set up a reproducibility initiative in collaboration with Science Exchange (Prostate Cancer Foundation 2014).

There are varying degrees of evidence linking the funders’ activities described above to the project. For NWO, there were multiple factors that led it to start its pilot, and the Reproducibility Project was among them (Hillebrink and Macor, pers.comm. 2016). A researcher who had contributed to the Reproducibility Project is a member of the programme committee for the pilot, and she provided insights about how to structure the programme, the types of replication studies to fund, and how to phrase the call (ibid.). The NIH consulted individuals, including Nosek, about reproducibility during the development of its overarching strategy (Anderson 2014), and Nosek was also involved in an NSF workshop on robust research (Cacioppo et al. 2015).

Sources

Interviews

We gratefully acknowledge the following interviewees, who provided information for this case study:

- Timothy Errington, the Center for Open Science
- Stuart Buck, the Laura and John Arnold Foundation
- Carlien Hillebrink and Guillaume Macor, the Netherlands Organisation for Scientific Research

Bibliography


---

The Initiative for Impact Evaluation (3ie) also runs an internal replication programme, where the validity and robustness of original estimates and recommendations are checked, and it has announced plans to begin using the OSF as a platform to publish its ‘push-button replication’ reports (Brown & Wood 2016).


Levelt Committee, Noort Committee & Drenth Committee. 2012. 'Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel'. tilburguniversity.edu,


Yong, E. 2012. ‘Nobel Laureate Challenges Psychologists to Clean up Their Act’. nature.com, News and Comment, 3 October. doi:10.1038/nature.2012.11535