



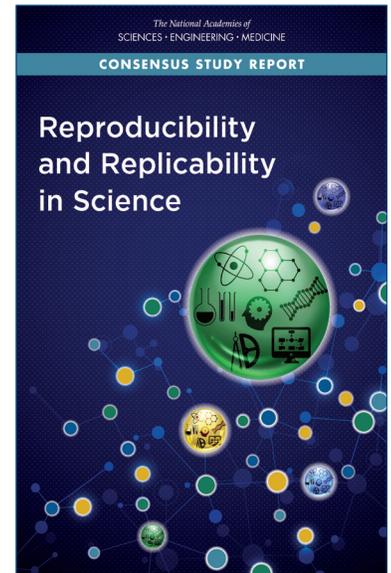
May 2019

Reproducibility and Replicability in Science

One of the pathways by which the scientific community confirms the validity of a new scientific discovery is by repeating the research that produced it. When a scientific effort fails to independently confirm the computations or results of a previous study, some fear that it may be a symptom of a lack of rigor in science, while others argue that such an observed inconsistency can be an important precursor to new discovery.

Concerns about reproducibility and replicability have been expressed in both scientific and popular media. As these concerns came to light, Congress requested that the National Academies of Sciences, Engineering, and Medicine conduct a study to assess the extent of issues related to reproducibility and replicability and to offer recommendations for improving rigor and transparency in scientific research.

The National Academies' study resulted in a report, *Reproducibility and Replicability in Science* (2019), that offers definitions of reproducibility and replicability and examines the factors that may lead to non-reproducibility and non-replicability in research. While reproducibility is straightforward and should generally be expected, the report says, replicability is more nuanced, and in some cases a lack of replicability can aid the process of scientific discovery. The report provides recommendations to researchers, academic institutions, journals, and funders on steps they can take to improve reproducibility and replicability in science.



DEFINING REPRODUCIBILITY AND REPLICABILITY

The terms “reproducibility” and “replicability” are often used interchangeably, but the report uses each term to refer to a separate concept.

Reproducibility means computational reproducibility—obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis.

Replicability means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

In short, reproducing research involves using the *original* data and code, while replicating research involves *new* data collection and similar methods used by previous studies. These two processes also differ in the type of results that should be expected. In general, when a researcher transparently reports a study and makes available the underlying digital artifacts, such as data and code, the results should be computationally reproducible. In contrast, even when a study was rigorously conducted according to best practices, correctly analyzed, and transparently reported, it may fail to be replicated.

REPRODUCIBILITY

The committee's definition of reproducibility focuses on computation because most scientific and engineering research disciplines use computation as a tool, and the abundance of data and widespread use of computation have transformed many disciplines. However, this revolution is not yet uniformly reflected in how scientists use software and how scientific results are published and shared. These shortfalls have implications for reproducibility, because scientists who wish to reproduce research may lack the information or training they need to do so.

When results are produced by complex computational processes using large volumes of data, the traditional methods section of a scientific paper is insufficient to convey the necessary information for others to reproduce the results. Additional information related to data, code, models, and computational analysis is needed.

To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis.

If sufficient data, code, and description of methods are available and a second researcher follows the methods described by the first researcher, one expects in many cases full bitwise reproduction of the original results—obtaining the same exact numeric values. For some research questions, bitwise reproduction may not be attainable and reproducible results could be obtained within an accepted range of variation.

How common is non-reproducibility in research? The evidence base is incomplete, and determining the extent of issues related to computational reproducibility across or within fields of science would be a massive undertaking with a low probability of success. However, a number of systematic efforts to reproduce computational results across a variety of fields have failed in more than half of the attempts made—mainly due to insufficient detail on digital artifacts such as data, code, and computational workflow.

REPLICABILITY

One important way to confirm or build on previous results is to follow the same methods, obtain new data, and see if the results are consistent with the original. A successful replication does not guarantee that the original scientific results of a study were correct, however, nor does a single failed replication conclusively refute the original claims.

Non-replicability can arise from a number of sources. The committee classified sources of non-replicability into those that are potentially helpful to gaining knowledge, and those that are unhelpful.

Potentially helpful sources of non-replicability include inherent but uncharacterized uncertainties in the system being studied. These sources of non-replicability are a normal part of the scientific process, due to the intrinsic variation or complexity in nature, the scope of current scientific knowledge, and the limits of our current technologies. In such cases, a failure to replicate may lead to the discovery of new phenomena or new insights about variability in the system being studied.

In other cases, non-replicability is due to shortcomings in the design, conduct, and communication of a study. Whether arising from lack of knowledge, perverse incentives, sloppiness, or bias, these unhelpful sources of non-replicability reduce the efficiency of scientific progress.

Unhelpful sources of non-replicability can be minimized through initiatives and practices aimed at improving research design and methodology through training and mentoring, repeating experiments before publication,

rigorous peer review, utilizing tools for checking analysis and results, and better transparency in reporting. Efforts to minimize avoidable and unhelpful sources of non-replicability warrant continued attention.

Researchers who knowingly use questionable research practices with the intent to deceive are committing misconduct or fraud. It can be difficult in practice to differentiate between honest mistakes and deliberate misconduct, because the underlying action may be the same while the intent is not. Scientific misconduct in the form of misrepresentation and fraud is a continuing concern for all of science, even though it accounts for a very small percentage of published scientific papers.

IMPROVING REPRODUCIBILITY AND REPLICABILITY IN RESEARCH

The report recommends a range of steps that stakeholders in the research enterprise should take to improve reproducibility and replicability, including:

- All researchers should include a clear, specific, and complete description of how the reported results were reached. Reports should include details appropriate for the type of research, including:
 - ◇ a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study;
 - ◇ a clear description of the analysis of data and decisions for exclusion of some data or inclusion of other;
 - ◇ for results that depend on statistical inference, a description of the analytic decisions and when these decisions were made and whether the study is exploratory or confirmatory;
 - ◇ a discussion of the expected constraints on generality, such as which methodological features the authors think could be varied without affecting the result and which must remain constant;
 - ◇ reporting of precision or statistical power; and
 - ◇ discussion of the uncertainty of the measurements, results, and inferences.
- Funding agencies and organizations should consider investing in research and development of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion. Concurrently, investments would be helpful in outreach to inform and train researchers on best practices and how to use these tools.
- Journals should consider ways to ensure computational reproducibility for publications that make claims based on computations, to the extent ethically and legally possible.
- The National Science Foundation should take steps to facilitate the transparent sharing and availability of digital artifacts, such as data and code, for NSF-funded studies—including developing a set of criteria for trusted open repositories to be used by the scientific community for objects of the scholarly record, and endorsing or considering the creation of code and data repositories for long-term archiving and preservation of digital artifacts that support claims made in the scholarly record based on NSF-funded research, among other actions.

Additional recommendations, along with detail on those included above, can be found in the report.

CONFIDENCE IN SCIENCE

Replicability and reproducibility, useful as they are in building confidence in scientific knowledge, are not the only ways to gain confidence in scientific results. Multiple channels of evidence from a variety of studies provide a robust means for gaining confidence in scientific knowledge over time. Research synthesis and meta-analysis, for example, are valuable methods for assessing the reliability and validity of bodies of research. A goal of science is to understand the overall effect from a set of scientific studies, not to strictly determine whether any one study has replicated any other.

The committee was asked to consider if lack of replication and reproducibility impacts the public's perception of science. The committee was not aware of data that would directly answer that question, and coverage of the issue in public media remains low. Regardless, the report notes that scientists and journalists bear responsibility

for misrepresentation in the public’s eye when they overstate or otherwise misrepresent the implications of scientific research. The report offers the following recommendations:

- Scientists should take care to avoid overstating the implications of their research, exercised also in their review of press releases, especially when the results bear directly on matters of keen public interest and possible action.
- Journalists should report on scientific results with as much context and nuance as the medium allows. In covering issues related to replicability and reproducibility, journalists should help their audiences understand the differences between non-reproducibility and non-replicability due to fraudulent conduct of science, and instances in which the failure to reproduce or replicate may be due to evolving best practices in methods or inherent uncertainty in science.

COMMITTEE ON REPRODUCIBILITY AND REPLICABILITY IN SCIENCE

HARVEY V. FINEBERG (NAM), (*Chair*), Gordon and Betty Moore Foundation; **DAVID B. ALLISON** (NAM), School of Public Health-Bloomington Director, Indiana University; **LORENA A. BARBA**, School of Engineering and Applied Science, George Washington University; **DIANNE CHONG** (NAE), Boeing Research and Technology (retired); **JULIANA FREIRE**, Tandon School of Engineering, New York University; **GERALD GABRIELSE** (NAS), Department of Physics, Northwestern University; **CONSTANTINE GATSONIS**, Center for Statistical Sciences, Brown University; **EDWARD HALL**, Department of Philosophy, Harvard University; **THOMAS H. JORDAN** (NAS), Department of Earth Sciences, University of Southern California; **DIETRAM A. SCHEUFELE**, Madison and Morgridge Institute for Research, University of Wisconsin–Madison; **VICTORIA STODDEN**, Institute for Data Sciences and Engineering, University of Illinois at Urbana–Champaign; **TIMOTHY D. WILSON**, Department of Psychology, University of Virginia; **WENDY WOOD**, Department of Psychology, University of Southern California and INSEAD-Sorbonne University; **JENNIFER HEIMBERG**, *Study Director*; **THOMAS ARRISON**, *Program Director*; **MICHAEL COHEN**, *Senior Program Officer*; **MICHELLE SCHWALBE**, *Director*; **TINA WINTERS**, *Associate Program Officer*; **THELMA COX**, *Program Coordinator*.

For More Information . . . This Consensus Study Report Highlights was prepared by the Board on Behavioral, Cognitive, and Sensory Sciences based on the Consensus Study Report *Reproducibility and Replicability in Science* (2019). The study was sponsored by the Alfred P. Sloan Foundation and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project. Copies of the Consensus Study Report are available from the National Academies Press, (800) 624-6242; <http://www.nas.edu/ReproducibilityinScience>.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people’s lives worldwide.

www.national-academies.org