EXECUTIVE SUMMARY

The Nature of the Crisis

Figure 4: John Ioannidis

A *reproducibility crisis* afflicts a wide range of scientific and social-scientific disciplines, from epidemiology to social psychology. Improper research techniques, lack of accountability, disciplinary and political groupthink, and a scientific culture biased toward producing positive results together have produced a critical state of affairs. Many supposedly scientific results cannot be reproduced reliably in subsequent investigations, and offer no trustworthy insight into the way the world works.

In 2005, Dr. John Ioannidis argued, shockingly and persuasively, that most published research findings in his own field of medicine were false. Contributing factors included 1) the inherent limitations of statistical tests; 2) the use of small sample sizes; 3) reliance on small numbers of studies; 4) willingness to publish studies reporting small effects; 5) the prevalence of fishing expeditions to generate new hypotheses or explore unlikely correlations; 6) flexibility in research design; 7) intellectual prejudices and conflicts of interest; and 8) competition among researchers to produce positive results, especially in fashionable areas of research. Ioannidis demonstrated that when you accounted for all these factors, a majority of research findings in medicine—*and in many other scientific fields*—were probably wrong.

Ioannidis' alarming article crystallized the scientific community's awareness of the reproducibility crisis. Subsequent evidence confirmed that the crisis of reproducibility had compromised entire disciplines. In 2012 the biotechnology firm Amgen tried to reproduce 53 "landmark" studies in hematology and oncology, but could only replicate six. In that same year the director of the Center for Drug Evaluation and Research at the Food and Drug Administration estimated that up to three-quarters of published biomarker associations could not be replicated. A 2015 article in *Science* that presented the results of 100 replication studies of articles published in prominent psychological journals found that only 36% of the replication studies produced statistically significant results, compared with 97% of the original studies.

Many common forms of improper scientific practice contribute to the crisis of reproducibility. Some researchers look for correlations until they find a spurious "statistically significant" relationship. Many more have a poor understanding of statistical methodology, and thus routinely employ statistics improperly in their research. Researchers may consciously or unconsciously bias their data to produce desired outcomes, or combine data sets in such a way as to invalidate their conclusions. Researchers able to choose between multiple measures of a variable often decide to use the one which provides a statistically significant result. Apparently legitimate procedures all too easily drift across a fuzzy line into illegitimate manipulations of research techniques.

Many aspects of the professional environment in which researchers work enable these distortions of the scientific method. *Uncontrolled researcher freedom* makes it easy for researchers to err in all the ways described above. The fewer the constraints on their research designs, the more opportunities for them to go astray. Lack of constraints allows researchers to alter their methods midway through a study as they pursue publishable, statistically significant results. Researchers often justify midstream alteration of research procedures as "flexibility," but in practice such flexibility frequently justifies researchers' unwillingness to accept a negative outcome. A 2011 article estimated that providing four "degrees of researcher freedom"—four ways to shift the design of the experiment while it is in progress—can lead to a 61% false-positive rate.

The *absence of openness* in much scientific research poses a related problem. Researchers far too rarely share data and methodology once they complete their studies. Scientists ought to be able to check and critique one another's work, but a great deal of research can't be evaluated properly because researchers don't always make their data and study protocols available to the public. Sometimes unreleased data sets simply vanish because computer files are lost or corrupted, or because no provision is made to transfer data to up-to-date systems. In these cases, other researchers lose the ability to examine the data and verify that it has been handled correctly.

Another factor contributing to the reproducibility crisis is the *premium on positive results*. Modern science's professional culture prizes positive results far above negative results, and also far above attempts to reproduce earlier research. Scientists therefore steer away from replication studies, and their negative results go into the file drawer. Recent studies provide evidence that this phenomenon afflicts such diverse fields as climate science, psychology, sociology, and even dentistry.

Groupthink also inhibits attempts to check results, since replication studies can undermine comfortable beliefs. An entire academic discipline can succumb to groupthink and create a professional consensus with a strong tendency to dismiss results that question its foundations. The overwhelming political homogeneity of academics has also created a culture of groupthink that distorts academic research, since researchers may readily accept results that confirm a liberal world-view while rejecting "conservative" conclusions out of hand. Political groupthink particularly affects those fields with obvious policy implications, such as social psychology and climate science.

Just the financial consequences of the reproducibility crisis are enormous. A 2015 study estimated that researchers spent around $28 billion annually in the United States alone on irreproducible preclinical research into new drug treatments. Irreproducible research in several disciplines distorts public policy and public expenditure in areas such as public health, climate science, and marriage and family law. The gravest casualty of all is the authority that science ought to have with the public, but which it has been forfeiting through its embrace of practices that no longer serve to produce reliable knowledge.

Many researchers and interested laymen have already started to improve the practice of science. Scientists, journals, foundations, and the government have all taken concrete steps to alleviate the crisis of reproducibility. But there is still much more to do. The institutions of modern science are

enormous, not all scientists accept the nature and extent of the crisis, and the public has scarcely begun to realize the crisis's gravity. Fixing the crisis of reproducibility will require a great deal of work. A long-term solution will need to address the crisis at every level: technical competence, institutional practices, and professional culture.

The National Association of Scholars proposes the following list of 40 specific reforms that address all levels of the reproducibility crisis. These suggested reforms are not comprehensive—although we believe they are more comprehensive than any previous set of recommendations. Some of these reforms have been proposed before; others are new. Some will elicit broad assent from the scientific community; we expect others to arouse fierce disagreement. Some are meant to provoke constructive critique.

We do not expect every detail of these proposed reforms to be adopted. Yet we believe that any successful reform program must be at least as ambitious as what we present here. If not these changes, then what? We proffer this program of reform to spark an urgently needed national conversation on how precisely to solve the crisis of reproducibility.

Recommendations

STATISTICAL STANDARDS

1. Researchers should avoid regarding the p-value as a dispositive measure of evidence for or against a particular research hypothesis.

2. Researchers should adopt the best existing practice of the most rigorous sciences and define statistical significance as $p < .01$ rather than as $p < .05$.

3. In reporting their results, researchers should consider replacing either-or tests of statistical significance with confidence intervals that provide a range in which a variable's true value most likely falls.

DATA HANDLING

4. Researchers should make their data available for public inspection after publication of their results.

5. Researchers should experiment with born-open data—data archived in an open-access repository at the moment of its creation, and automatically time-stamped.

RESEARCH PRACTICES

6. Researchers should pre-register their research protocols, filing them in advance with an appropriate scientific journal, professional organization, or government agency.

7. Researchers should adopt standardized descriptions of research materials and procedures.

PEDAGOGY

8. Disciplines that rely heavily upon statistics should institute rigorous programs of education that emphasize the ways researchers can misunderstand and misuse statistical concepts and techniques.

9. Disciplines that rely heavily upon statistics should educate researchers in the insights provided by Bayesian approaches.

10. Basic statistics should be integrated into high school and college math and science curricula, and should emphasize the limits to the certainty that statistics can provide.

UNIVERSITY POLICIES

11. Universities judging applications for tenure and promotion should require adherence to best-existing-practice standards for research techniques.

12. Universities should integrate survey-level statistics courses into their core curricula and distribution requirements.

PROFESSIONAL ASSOCIATIONS

13. Each discipline should institutionalize regular evaluations of its intellectual openness by committees of extradisciplinary professionals.

PROFESSIONAL JOURNALS

14. Professional journals should make their peer review processes transparent to outside examination.

15. Some professional journals should experiment with guaranteeing publication for research with pre-registered, peer-reviewed hypotheses and procedures.

16. Every discipline should establish a professional journal devoted to publishing negative results.

SCIENTIFIC INDUSTRY

17. Scientific industry should advocate for practices that minimize irreproducible research, such as Transparency and Openness Promotion (TOP) guidelines for scientific journals.

18. Scientific industry, in conjunction with its academic partners, should formulate standard research protocols that will promote reproducible research.

PRIVATE PHILANTHROPY

19. Private philanthropy should fund scientists' efforts to replicate earlier research.

20. Private philanthropy should fund scientists who work to develop better research methods.

21. Private philanthropy should fund university chairs in "reproducibility studies."

22. Private philanthropy should establish an annual prize, the Michelson-Morley Award, for the most significant negative results in various scientific fields.

23. Private philanthropy should improve science journalism by funding continuing education for journalists about the scientific background to the reproducibility crisis.

GOVERNMENT FUNDING

24. Government agencies should fund scientists' efforts to replicate earlier research.

25. Government agencies should fund scientists who work to develop better research methods.

26. Government agencies should prioritize grant funding for researchers who pre-register their research protocols and who make their data and research protocols publicly available.

27. Government granting agencies should immediately adopt the National Institutes of Health (NIH) standards for funding reproducible research.

28. Government granting agencies should provide funding for programs to broaden statistical literacy in primary, secondary, and post-secondary education.

GOVERNMENT REGULATION

29. Government agencies should insist that all new regulations requiring scientific justification rely solely on research that meets strict reproducibility standards.

30. Government agencies should institute review commissions to determine which existing regulations are based on reproducible research, and to rescind those which are not.

FEDERAL LEGISLATION

31. Congress should pass an expanded Secret Science Reform Act to prevent government agencies from making regulations based on irreproducible research.

32. Congress should require government agencies to adopt strict reproducibility standards by measures that include strengthening the Information Quality Act.

33. Congress should provide funding for programs to broaden statistical literacy in primary, secondary, and post-secondary education.

STATE LEGISLATION

34. State legislatures should reform K-12 curricula to include courses in statistics literacy.

35. State legislatures should use their funding and oversight powers to encourage public university administrations to add statistical literacy requirements.

GOVERNMENT STAFFING

36. Presidents, governors, legislative committees, and individual legislators should employ staff trained in statistics and reproducible research techniques to advise them on scientific issues.

JUDICIARY REFORMS

37. Federal and state courts should adopt a standard approach, which explicitly accounts for the crisis of reproducibility, for the use of science and social science in judicial decision-making.

38. Federal and state courts should adopt a standard approach to overturning precedents based on irreproducible science and social science.

39. A commission of judges should recommend that law schools institute a required course on science and statistics as they pertain to the law.

40. A commission of judges should recommend that each state incorporate a science and statistics course into its continuing legal education requirements for attorneys and judges.