Reproducibility and replicability of science and thoracic surgery

Jennifer S. Lawton, MD

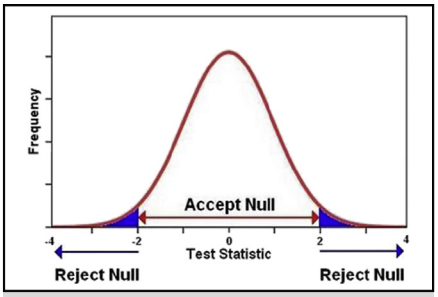
From the Division of Cardiac Surgery, Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Md.

Received for publication July 9, 2016; revisions received Aug 23, 2016; accepted for publication Aug 25, 2016.

Address for reprints: Jennifer S. Lawton, MD, Johns Hopkins University School of Medicine, 1800 Orleans St, Zayed 7101, Baltimore, MD 21287 (E-mail: jlawton4@jhmi.edu).

J Thorac Cardiovasc Surg 2016;■:1-3
0022-5223/$36.00

http://dx.doi.org/10.1016/j.jtcvs.2016.08.044



All ''good'' science is hypothesis driven and begins with a null hypothesis. Taken from Google images at: http://m9.i.pbase.com/o9/10/152510/1/152440759.QXJVvzP1.NIPCC_Null_Hypothesis.PNG. Accessed July 7, 2016.

Central Message

The concept of reproducibility in science has led to multiple changes in the global scientific community and beyond. This commentary details some of the changes and how they affect thoracic surgeons.

Feature Editor's Note—*There are abundant examples of irreproducible scientific research that is slowly but surely surfacing to the consciousness of cardiothoracic surgeons. The editorial staff of the journal* Nature *conducted a survey of more than 1500 scientists to estimate the reproducibility of peer-reviewed published results. More than two thirds of the sampled scientists tried and failed to reproduce another scientist's experiments, and more than half failed to reproduce their own experiments. Among these 1500 scientists were medical researchers who had similar problems with the reproducibility of published research.*[1]

In 2012, scientists at Amgen attempted to reproduce benchmark studies of cancer research, collaborating closely with the original investigators. They were able to reproduce only 6 of 53 of these high-profile publications.[2] *In a highly unusual move, this pharmaceutical company released their internal findings to highlight the lack of reproducibility in published reports.*

Bayer Pharmaceutical's internal efforts to validate new drug target claims found that in-house experimental data do not match literature claims in 65% of drug target-validation projects.[3]

There is an increasing rate of failure of Phase III trials to reproduce findings of positive Phase II trials,[4] *suggesting that irreproducible results exist not only at the basic science levels but also at advanced stages of clinical drug testing for Food and Drug Administration drug approval.*[4] *Irreproducible results published in the scientific literature are associated with astronomical costs and wasted resources.*

In this issue of the Journal, *Dr Jennifer Lawton outlines the causes of and possible solutions to this prevalent, challenging, and ethically disturbing problem. Her insights strike at the heart of cardiothoracic surgeons' ability to accept and use published reports. Every surgeon who ever went to the literature to gain insight into a difficult clinical problem needs to read and understand Dr Lawton's article.*

Victor A. Ferraris, MD, PhD

The notion and demonstration of a lack of reproducibility (the ability to repeatedly obtain the same results from data) or replicability (the ability of other investigators to observe the same result under identical conditions) in science have raised ethical questions regarding research and led to multiple changes and initiatives. The inability to reproduce clinical research results that influence patient care has obvious ethical implications. Likewise, erroneous basic science research that provides a foundation for further work can be devastating to medical advances and new discovery. When results cannot be duplicated by other laboratories, negative motives are inferred; however, a variety of factors may be involved, including the dynamic nature of experiments in complex organisms, improper study design and statistical analysis, lack of quality control for biological reagents, bias against negative results, and pressure to publish.[5]

A variety of changes in response to reproducibility concerns have been introduced, including specific checklists for scientific journal manuscript submission and review, new requirements for applications for research funding, the creation of social media sites for dissemination of knowledge (National Institutes of Health [NIH] initiative PubMed Commons) and postpublication review of research (PubPeer, Nature's Protocol Exchange), and the formation of new nonprofit companies (Center for Open Science).[6-9]

Changes pertinent to the practicing thoracic surgeon will be explored in this editorial.

Plans for initiatives to improve the reproducibility in science were introduced by the NIH in 2014.[10] Some of the proposed reasons why these initiatives are necessary include the following: the poor training of researchers on experimental design and ethics of research, decreased emphasis on technical details of experiments, and the lack of reporting of the basic elements of experimental design (eg, blinding, sample size calculation).[10] In addition, preclinical research (vs clinical) is suggested to be more susceptible to reproducibility issues.

The NIH Initiative of Rigor and Transparency now (as of January 2016) applies to all submitted research grants and mentored career development awards. This initiative is postponed for individual fellowship, institutional career development, and institutional training grant applications.[11] Research grants or Mentored Career Development Applications will be evaluated on 3 new areas that will affect grant scoring: Scientific Premise, Scientific Rigor, and Consideration of Relevant Biological Variables. In addition, a Plan for Resource Authentication will be reviewed after scoring. *Scientific premise* is established by the justification of why the proposed project is logical and necessary based on the key data presented (data that may or may not come from the applicant's own research) and should be included in the *Significance* section of research grants. *Scientific rigor* is included in the *Approach* section for research grants and pertains to the statistical procedures, data analysis, interpretation, and reporting of results. *Consideration of Relevant Biological Variables (Sex)* is addressed in the *Approach* section and pertains to the use of both sexes of subjects (human or vertebrate animals). Justification must be provided if both sexes are not used or included. *Plan for Resource Authentication* should be documented *in an attachment*, and review will be after grant scoring. This plan should detail important biological and chemical resources that may differ between laboratories or over time and that could influence data and are integral to the proposed research.

How do these new efforts to improve the reproducibility of science specifically affect the surgeon investigator? NIH initiatives now require the necessary documentation of rigor and transparency in all submitted grant applications (without an allowance for additional pages), and these will influence grant scoring, mandatory journal statistical peer review will increase time to publication, and journal checklists that require responses to questions regarding blinding and calculation of sample size will require additional time and careful responses.

These initiatives will not seem threatening to conscientious and careful investigators. Most scientists would readily embrace the proposed changes as appropriate and expected ethical scientific behavior. Likewise, the majority of scientists report only their best data after exhaustive and thorough confirmation. Detailed descriptions of methods will facilitate the training of new laboratory personnel and the writing of methods sections for articles and grant proposals.

Some researchers may hesitate to explicitly detail their experimental methods for fear that other investigators will steal their ideas or models, particularly given the current state of competition for reduced grant funding. There also may be some self-aggrandizing or narcissistic researchers who may take great pleasure in the belief that their own laboratories are the only ones in the world that can perform a particular assay or technically challenging model successfully.

In my own basic science laboratory, we have been dismayed when we have been unable to reproduce the work of others. I learned years ago during my 2-year laboratory fellowship in general surgery residency that when a new member joins the laboratory, he or she must reproduce basic experiments that have already been done. I have continued this in my own laboratory. Many would argue that this redundancy is inappropriate because of unnecessary animal sacrifice or, in the case of clinical research, that reproducing large clinical trials would be impossible. In addition, how will such replication of research be justified financially or ethically?

In addition, who has the time to repeat *all* of their experiments? Current significant funding challenges make additional experiments difficult, and research (as we are told more and more these days at academic institutions) is not a money-making proposition. New faculty investigators are heavily pressured to produce new, novel, and innovative preliminary data to be deemed ''worthy'' of funding and must do this on a ''tenure clock.'' Likewise, established investigators must demonstrate continued productivity to obtain continuous funding. Articles may take 6 to 9 months to reach publication, therefore reducing potential time to consider careful replication of findings.

The innumerable examples in the literature and media of the inability to reproduce published work and reports of the submission of deliberately false or fictitious work also have placed pressure on journal editors, unprepared peer reviewers, and grant peer reviewers.[8,12,13] Journals that are constantly pressured to increase impact factor are less likely to publish negative data that are deemed less innovative or exciting to their readers (and less likely to be cited by others). In addition, most peer reviewers are busy surgeons, many of whom are being told to make more relative value units, who donate their valuable (uncompensated) time to review articles.