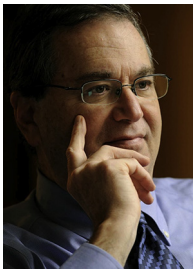
Commentary

MOLECULAR METABOLISM

Irreproducibility of published bioscience research: Diagnosis, pathogenesis and therapy



Jeffrey S. Flier was named the 21st Dean of the Faculty of Medicine at Harvard University on July 11, 2007, at which point he became the Caroline Shields Walker Professor of Medicine at Harvard Medical School. Flier is an endocrinologist and an authority on the molecular causes of obesity and diabetes. Prior to becoming dean, he served from 2002 to 2007 as Harvard Medical School Faculty Dean for Academic Programs and Chief Academic Officer for Beth Israel Deaconess Medical Center (BIDMC), a Harvard teaching affiliate. Born in New York City, Dr. Flier received a BS from City College of New York in 1968, and an MD from Mount Sinai School of Medicine in 1972, graduating with the Elster Award for Highest Academic Standing. Following residency training in internal medicine at Mount Sinai Hospital from 1972 to 1974, Flier moved to the National Institutes of Health as a Clinical Associate. In 1978, he joined the Faculty of Medicine at Harvard Medical School, serving as Chief of the Diabetes Unit at Beth Israel Hospital until 1990, when he was named chief of the hospital's Endocrine Division. Dr. Flier is one of the country's leading investigators in the areas of obesity and diabetes and has authored over 200 scholarly papers and reviews. His research has produced major insights into the molecular mechanism of insulin action, the molecular mechanisms of insulin resistance in human disease, and the molecular pathophysiology of obesity. An elected member of the Institute of Medicine and a fellow of the American Academy of Arts and Sciences, Flier's honors include the Eli Lilly Award of the American Diabetes Association, the Berson Lecture of the American Physiological Society, and an Honorary Doctorate from the University of Athens. He was the 2003 recipient of the Edwin B. Astwood Lecture Award from the Endocrine Society, and in 2005 he received the Banting Medal from the American Diabetes Association, its highest scientific honor. In 2008, Dr. Flier was awarded the Albert Renold Award from the American Diabetes Association for outstanding achievements in the training of diabetes research scientists and the facilitation of diabetes research. In 2010, Flier was awarded an Honorary Doctor of Science Degree from the University of Edinburgh. In 2011, Dr. Flier received the Rolf Luft Award for Metabolic Research from the Karolinska Institute in Sweden. In July of 2016, after nine years as Dean of Harvard Medical School, Dr. Flier stepped down from that position, rejoining the HMS faculty, based in the Neurobiology Department, as the George Higginson Professor of Physiology and Medicine, and Harvard University Distinguished Service Professor. Dr. Flier is married to Eleftheria Maratos-Flier, MD, who is also on the faculty of Harvard Medical School and with whom he has collaborated on research in the area of neuroendocrine control of body weight. They have two daughters, Lydia and Sarah, and live in Newton, Mass.

Jeffrey S. Flier [a]

During a 40 year career as a biomedical researcher and academic leader, my primary professional goal has been to discover and disseminate new knowledge relevant to biology and health, with my own efforts focused on metabolic physiology and disease. I have done this during a period of dynamic growth of the bioscience enterprise, which has produced remarkable discoveries to illuminate our understanding of human biology and disease while creating numerous benefits for the health and welfare of society. The bioscience research ecosystem that supports this effort is large and complex. Physicians and PhD scientists in academia and industry conduct basic and clinical research, spending over 100 billion dollars yearly in the US alone [1]. The results of this research eventually appear in over one million scientific papers per year [2], in more than 5000 journals [3] of varying focus, standards, and impact. These publications are the bricks on which the edifice of scientific progress is built. The worldwide scientific community reads, discusses, assesses, and wherever possible builds upon the results reported in these papers. The overall arc of progress from this activity is evident, and there is little doubt that the future will continue to bring discoveries of profoundly important impact.

But the direction of scientific progress is not exclusively forward. Much research is exploratory in nature, and tentative conclusions are both expected and beneficial. Research publications will contain errors, despite procedures designed to avoid them. Fortunately, a fundamental attribute of science is its capacity for "self-correction", through published ideas and claims being reviewed and tested by others [4]. Although scientists should and most often do seek to publish reliable results, to expect a standard of certainty before publication, and/or to excessively stigmatize or penalize claims later found honestly to be in error, would diminish progress by replacing a spirit of scientific excitement and daring with professional fear of error. The key question, therefore, is to define an optimal balance, surely weighted in the direction of reliability, but appropriately tolerant of tentative conclusions and honest errors, while continuously seeking to reduce the latter.

But today we face claims, from a variety of sources, that published bioscience research is far less reproducible than anyone previously imagined [5–7]. If the most extreme of these claims are true, they challenge the integrity of the research enterprise, threatening the public support and funding that sustain it. Indeed, we might be required to seriously reconsider our approach to conducting and publishing research. Consequently, the bioscience research community, and those committed to its success, must take these claims seriously.

I write from a perspective that begins with forty years in the trenches of metabolic research. I have published many papers, and have played diverse roles in scientific publishing, both as peer reviewer and editor. As Dean of Harvard Medical School, I oversaw both academic appointment and promotion processes that emphasize the evaluation of published work, and the investigation of research misconduct and fraud by our faculty. I have both extolled and defended biomedical research to the general public. Research reproducibility is of paramount importance in each of these realms.

1. REPRODUCIBILITY IN DISTINCT DOMAINS OF BIOSCIENCE RESEARCH

Bioscience research covers a broad spectrum from basic science in a variety of disciplines, to translational research that links basic science

[a]Department of Medicine and Neurobiology, Harvard Medical School, 220 Longwood Ave, Goldenson 542, Boston, MA, 02115, USA

E-mail: Jeffrey_flier@hms.harvard.edu.

Available online 21 November 2016

http://dx.doi.org/10.1016/j.molmet.2016.11.006

and animal research to therapeutic implications, to clinical research involving human subjects. The need for reproducibility applies to all such research, though some issues are specific to specific domains. The dominant focus of this paper is on reproducibility in basic and translational research, but I will reference, where appropriate, specific issues related to clinical research reproducibility.

2. WHAT DOES RESEARCH REPRODUCIBILITY MEAN?

To understand this problem, we must first define it. Unfortunately, the definition and criteria for considering research reproducible are far from unambiguous [8,9]. In the most restrictive and precise use of the term, experimental results are replicated when an independent group performs the same experiment, under the same conditions and using the same reagents, then finding the same results. Most published research is never replicated in this manner, and I don't believe that a healthy scientific ecosystem requires it to be. Certainly, the academic and financial incentives to routinely replicate the work of others do not exist. Rather than being replicated, published results are typically built upon and extended by subsequent research, perhaps carried out under somewhat different conditions from the original work, or with different reagents, but with results that are seen as consistent with and supportive of the earlier claims. Thus, though the first results were not formally replicated, they provided a sound foundation for subsequent work, thereby advancing our understanding.

In some fields, replication may be accomplished by reanalyzing a published data set and reaching the same conclusions, rather than generating a new data set [9].

Problems arise when new research or a new analysis of prior research is, or at least seems to be, inconsistent with previously published findings. This may have many explanations, related to differences in design or execution of either the initial or the subsequent work [10,11]. One often cannot conclude whether one of the studies is at fault, or whether they are just different, and the answer often requires additional experimentation. In cases where a new result appears inconsistent with a prior report, if the new study employed experimental conditions and reagents distinct from the first, it should not be taken as a failure to replicate it. It is possible that the experiment would have been successfully replicated had its precise conditions been employed. On the other hand, a failure to replicate under slightly different conditions may suggest that the initial finding is at least less generalizable than initially claimed.

Reproducing research employing molecules, cells, and model organisms is practically less daunting than attempting to reproduce clinical research, especially complex and expensive clinical studies involving hundreds or thousands of patients. This fact, plus the immediate importance of human studies, renders it especially important that clinical research experiments be well designed, pre-registered, have data available at publication for analysis by others, and be committed to publishing results whether positive or negative in relationship to the initial hypothesis. Of course, similar arguments can be made for optimal design of pre-clinical research as well, though there are many practical obstacles to all of these practices becoming routine.

There are at least two ends of the spectrum of reproducibility. At one end, a huge number of the one million papers published each year are minimally read or referenced [12]. Like a tree falling unobserved in the forest, whether such papers are reproducible is unknown. There are undoubtedly some gems buried within that largely unexamined and untested pool of papers, just as there may be many among the unread and unreferenced papers whose truth would be challenged if more closely scrutinized. But either way, these papers now have limited impact on the world of science. Perhaps in the future, natural language processing and machine learning algorithms will permit the under-examined literature to be more effectively explored.

At the other end of the spectrum are papers of high impact with results of potential importance and great interest to the community. Sometimes, and thankfully quite rarely, the conclusions of such papers are rapidly shown to be false. In such cases, other scientists have been actively involved in the same research question, and at the time of publication already have data bearing on the new papers' claims. They view the success of their own research as requiring the new claims to be rapidly verified or rejected. If their work fails to replicate it, they would be motivated to quickly publish the negative results. Two prominent examples of high profile studies whose major findings were quite rapidly shown to be false, were the claims about the putative beta cell stimulatory hormone betatrophin [13–15] and STAP cells as a facile approach to creating totipotential stem cells [16–18]. In such cases, though the corrections and/or retractions may be quickly published, it is important to understand why the erroneous findings came to be believed by the scientific teams and then published in high impact journals to public acclaim. Such outcomes can result from honest errors and/or incompetence, or more nefarious causes involving research misconduct or even outright fraud.

Although high impact papers whose results are rapidly retracted garner major attention, we should perhaps be more concerned about another situation that appears to be more common: papers of potential importance where many in the community have difficulty building on or reproducing the results, but despite much informal discussion of these failures, whether at meetings or at the water cooler, papers documenting this fact do not get published.

3. HOW COMMON IS IRREPRODUCIBLE RESEARCH?

What fraction of the published bioscience research is not reproducible? The most straightforward answer is that we really don't know. The answer would require examining data from a sufficiently large and representative sample of studies where replication was attempted, with all of the caveats about what true replication means. The empirical data from which to draw such conclusions are generally not available.

Some papers addressing this topic make claims about the high rate of irreproducibility of the published literature [5]. These are theoretical examinations, based on sampling of the literature and assessing such factors as appropriateness of sample sizes, statistical deficiencies, experimental and publishing bias, and other factors, and these have led to dire conclusions about the truthfulness of the larger literature. In another approach, several articles have reported that pharmaceutical labs are unable to reproduce the core findings of a large fraction of academic papers published in high impact journals [6,7]. These pharma groups are highly motivated to verify the results of publications reporting new therapeutic targets, and they have the resources and highly trained scientists to do the work. On the other hand, despite their claims of failures to reproduce, these papers present no actual data on which to evaluate their claims [6,7]. Such concerns appear to be quite prevalent in the biopharma community, are clearly worrisome to the academic community, and very likely reflect real problems. But as for all scientific claims, their conclusions cannot be accepted as true without the opportunity to examine the underlying data. It would be very helpful to the scientific enterprise if

