Biochemical Pharmacology 138 (2017) 7–18

Contents lists available at ScienceDirect

Biochemical Pharmacology

journal homepage: www.elsevier.com/locate/biochempharm

Commentary

Enhancing reproducibility: Failures from Reproducibility Initiatives underline core challenges

Kevin Mullane [a], Michael Williams [b,*]

[a] Gladstone Institutes, San Francisco, CA, United States
[b] Department of Pharmacology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

ARTICLE INFO

Article history:
Received 12 March 2017
Accepted 5 April 2017
Available online 8 April 2017

Keywords:
Reproducibility
Peer review
Scientific self-replication

ABSTRACT

Efforts to address reproducibility concerns in biomedical research include: initiatives to improve journal publication standards and peer review; increased attention to publishing methodological details that enable experiments to be reconstructed; guidelines on standards for study design, implementation, analysis and execution; meta-analyses of multiple studies within a field to synthesize a common conclusion and; the formation of consortia to adopt uniform protocols and internally reproduce data. Another approach to addressing reproducibility are *Reproducibility Initiatives* (RIs), well-intended, high-profile, systematically peer-vetted initiatives that are intended to replace the traditional process of scientific self-correction. Outcomes from the RIs reported to date have questioned the usefulness of this approach, particularly when the RI outcome differs from other independent self-correction studies that have reproduced the original finding. As a failed RI attempt is a single outcome distinct from the original study, it cannot provide any definitive conclusions necessitating additional studies that the RI approach has neither the ability nor intent of conducting making it a questionable replacement for self-correction. A failed RI attempt also has the potential to damage the reputation of the author of the original finding. Reproduction is frequently confused with replication, an issue that is more than semantic with the former denoting "similarity" and the latter an "exact copy" – an impossible outcome in research because of known and unknown technical, environmental and motivational differences between the original and reproduction studies. To date, the RI framework has negatively impacted efforts to improve reproducibility, confounding attempts to determine whether a research finding is real.

Contents

* Corresponding author.
E-mail address: rivoli1635@comcast.net (M. Williams).

http://dx.doi.org/10.1016/j.bcp.2017.04.008
0006-2952/

Falsehood flies, and truth comes limping after it, so that when men come to be undeceived, it is too late; the jest is over, and the tale hath had its effect

[Swift, 1710]

1. Introduction

Advances in biomedical research have historically occurred in the context of sharing experimental findings, via the peer reviewed literature, that add to, advance and/or challenge a canon of established research. Traditionally, the key elements in validating such findings involve the interdependent activities of peer review and scientific self-replication.

Concerns regarding reproducibility trigger emotive responses. While many researchers acknowledge that it is a problem and admit to instances of being unable to reproduce another researcher's work (and sometimes their own), few are willing to analyze their own practices in detail. This may relate, in part, to the fact that whether a study can be repeated or not, itself a binary event, is often interpreted in the context of whether the result is real or true [1]. By inference, this can impugn the integrity and truthfulness of the investigator. Moreover, even the taint of fraud can damage promising careers, so reproduction has become something to avoid, perpetuating the current cultural mores that encourage the avoidance of both personal and institutional responsibility.

2. Experimental validation – an historical perspective

The scientific revolution that took place in Europe in the 16th–18th centuries which was closely linked to the intellectual Age of Enlightenment (also known as the Age of Reason) led to the formation of scientific societies including the Royal Society of London and the Académie Royale des Sciences in Paris. This concomitantly prompted the creation of the first scientific journals [2] along with the concept of experimental validation, espoused by the 17th century scientist philosopher, Robert Boyle, a concept based on the idea that an experiment had to be witnessed to give it legitimacy. As a result, the Royal Society, along with other learned societies, became the venue for public demonstrations where witnesses to an experiment were required to be both knowledgeable and also of good moral character. Despite the transparency of this process, Boyle is also credited with noting that "you will find. . .many of the Experiments publish'd by Authors, or related to you by the persons you converse with, false or unsuccessful" [3].

This led the Irish satirist and author of *Gulliver's Travels,* Jonathan Swift, to guide his hero to the mythical island of Balnibarbi where in 1708 Gulliver visited the Grand Academy in Lagado, an event that provided Swift with the opportunity to ridicule the profligacy of much of the science performed under the auspices of the Royal Society. Swift's satire used as examples of 18th century scientific endeavors the extraction of sunbeams from cucumbers, reducing human excrement to its original food sources, and determining the color of paint solely by feel and smell, all of which were gathered under the unbridled optimism of the immediate rewards that would occur [4]. Although Gulliver bore witness to much of the science ongoing at the mythical Grand Academy, the experiments conducted rarely panned out.

The naïve and subjective optimism of the members of the Grand Academy continues to manifest itself today in the way that the mainstream media hypes any new research discovery as a cure for some previously intractable disease, steadfastly ignoring a long list of failures and the dashed hopes of patients arising from previous pronouncements [5,124]. Swift also took the opportunity to satirize the continuous clamor for funding – an "Encouragement to Ingenuity" – still very much in evidence 300 years later – where one Lagado practitioner claimed that his discoveries "might be still improved, and much expedited, if the Publick would raise a fund" [4].

3. Reproducibility in 21st century biomedical research

While irreproducible studies have been a fact of life since the first scientific journals were published, it is only recently that the extent of this problem has been highlighted, where poor investigative practices, sloppy procedures, rampant bias in data selection, and/or an ignorance of validation techniques and statistical methodologies have led to numerous instances of research findings not being reproduced [6–10]. Indeed, it has been estimated that some 75–90% of published research findings cannot be reproduced [9], leading to an estimated $28bn per year - nearly half of the annual preclinical research budget in the US - being wasted on attempts to reproduce published research studies [11]. Aside from the waste, patient lives have also been placed at risk, when studies with drug candidates that had been incorrectly optimized using cell lines that were not what they were thought to be led to clinical trials that had to be terminated when the preclinical findings could not be reproduced [12]. As a result, scientific journals and the NIH, with their value and credibility at stake, have laid out new standards [13,14] and implemented computerized tools to help detect fraudulent activities such as plagiarism [15] and image manipulation [16], but the system remains based inherently on trust. As a result, it has been suggested that the biomedical research enterprise in its current form is in need of radical reform [9,17,18].