Biochemical Pharmacology 97 (2015) 133–136

Contents lists available at ScienceDirect

Biochemical Pharmacology

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

Commentary

Unknown unknowns in biomedical research: does an inability to deal with ambiguity contribute to issues of irreproducibility?

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

[a] Profectus Pharma Consulting Inc., San Jose 95125, United States
[b] Department of Pharmacology, Feinberg School of Medicine, Northwestern University, Chicago 60611, United States

ARTICLE INFO

Article history:
Received 4 July 2015
Accepted 6 July 2015
Available online 31 July 2015

Keywords:
Reproducibility, unknown unknowns
Reagent validation
Peer review

ABSTRACT

The credibility and consequent sustainability of the biomedical research "ecosystem" is in jeopardy, in part due to an inability to reproduce data from the peer-reviewed literature. Despite obvious and relatively inexpensive solutions to improve reproducibility—ensuring that experimental reagents, specifically cancer cell lines and antibodies, are authenticated/validated before use and that best practices in statistical usage are incorporated into the design, analysis, and reporting of experiments—these are routinely ignored, a reflection of hubris and a comfort with the status quo on the part of many investigators. New guidelines for the peer review of publications and grant applications introduced in the past year, while well-intended, lack the necessary consequences, e.g., denial of funding, that would result in sustained improvements when scientific rigor is lacking and/or transparency is, at best, opaque. An additional factor contributing to irreproducibility is a reductionist mindset that prioritizes certainty in research outcomes over the ambiguity intrinsic to biological systems that is often reflected in "unknown unknowns". This has resulted in a tendency towards codifying "rules" that can provide "yes-no" outcomes that represent a poor substitute for the intellectual challenge and skepticism that leads to an awareness and consideration of "unknown unknowns". When acknowledged as potential causes of unexpected experimental outcomes, these can often transition into the "knowns" that facilitate positive, disruptive innovation in biomedical research like the human microbiome. Changes in investigator mindset, both in terms of validating reagents and embracing ambiguity, are necessary to aid in reducing issues with reproducibility.

1. Background

Over the past 5 years, numerous articles, commentaries and position papers in both the mainstream media and the scientific literature have highlighted issues that have contributed to the lack of reproducibility of published biomedical research [1,2,3] as well as its clinical relevance [4,5]. As a consequence, the credibility and value of the biomedical research enterprise has been undermined [6,7,8] leading to concerns regarding its continued level of funding. The scope of the reproducibility issue has been further compounded by the addition of the shortcomings of the translational research process that underpin aspects of the clinical attrition rate in drug discovery [4,9,10] prompted by "*the assumption that translation, rather than fundamental understanding, is the choke point of progress in the application of science to societal problems*" [11].

The current level of concern on the issue of reproducibility is generally considered to have been kick started by Ioannidis' seminal paper, "*Why Most Published Research Findings Are False*" [12] and further elaborated in the commentaries by Begley and others [1,2,3] and by initiatives from the NIH [13,14]. This has led to a focus on three issues intimately related to reproducibility: (i) investigator hubris, incompetence and complacency that also includes bias and fraud [15,16]; (ii) neglect in validating/authenticating experimental reagents despite abundant evidence that these are a major cause of irreproducibility, [1,17,18] and; (iii) the broader topic of inappropriate experimental design, execution, analysis and reporting that is highlighted in the misunderstanding and misuse of statistical procedures [19,20,21].

To address these topics, the NIH issued new guidelines for publication that are intended to improve transparency and

* Corresponding author.
E-mail addresses: kevinmullane@comcast.net (K. Mullane), rivoli1635@comcast.net (M. Williams).

http://dx.doi.org/10.1016/j.bcp.2015.07.002
0006-2952/

accuracy in both the conduct and reporting of research activities [22] and are equally relevant to the grant funding process [23]. These have prompted guidelines that are specifically related to the pharmacological sciences [24,25] that will certainly help in discouraging fraud and will potentially lead to needed improvements in scientific rigor [3,6,23], increasing confidence in the both the reproducibility and relevance of published findings. But this will take time and require not only improvements in the training and mentoring of biomedical scientists [6,14] but fundamental changes in how 21st century biomedical research is funded and incentivized [3,6].

While enhancing the quality of accepted peer-reviewed publications, these guidelines provide little in the way of material consequences beyond manuscript rejection [26,27] that authors can easily circumvent by seeking other, less rigorous publication outlets. This allows the incentives that perpetuate the dysfunctional culture of "*rewarding A, while hoping for B*" [28] to remain in place, continuing to encourage poor behavior in the hypercompetitive scientific environment that represents 21st Century biomedical research [3,6].

Moreover, it is not unusual for researchers to wait until a study is complete to select an appropriate journal for the dissemination of the findings based on the strength of the evidence obtained and its perceived impact within the field. This *post hoc* decision makes journal guidelines that mandate specific aspects of experiment design, execution and analysis unlikely to be perceived as helpful, particularly when such requirements vary significantly between journals suggesting the need for greater consistency in generic guidelines for experimental conduct. In this context, the *Nature Publishing Group*, as part of their continuing efforts to resolve reproducibility issues, have noted that "*funding agencies... [must]... make clear their intentions in promoting rigorous lab standards.. [with]... a concomitant pressure on universities and institutes to demonstrate quality assurance of lab practices and culture*" [29].

Thus funding agencies, by virtue of their central position in the funding process both as arbiters of the validity and relevance of the published outcomes from sponsored research and in grant applications and also in their role as keepers of the taxpayer-funded research purse, are in unique position to enforce change. This can be achieved by providing clear, enforceable consequences rather than devolving this responsibility to the peer reviewed journals, research institutions and learned societies who lack appropriate coercive measures.

Instead, in response to the proposal by Nardone in 2008 of a zero tolerance policy in the context of the need for cell line authentication–"*No cell line authentication, no grant; no cell line authentication, no access to journals as publication outlets*" [26], the NIH noted that "*mandating testing–and specifying particular tests–would conflict with the spirit of the grants program, which encourages individual responsibility for the conduct and direction of sponsored research*"–making it a seminal example of the "*trust me model that is no longer considered appropriate in corporate life nor in government*" [3].

2. Obvious fixes

Of the topics identified, reagent validation and statistical misuse are perhaps the easiest and most straightforward to address, especially the former where the need for relatively inexpensive procedures for cell authentication [30] and antibody validation [18,31] have been generally ignored, in some instances for a decade or more. This probably results from investigator hubris and complacency, that has led to preventable translation failures in cancer drug discovery [32] and avoidable waste that has reached staggering proportions [17,33]. Likewise, in the area of statistics, while much has been written on the intrinsic limitations of the "magical" P value of less than 0.05 ($P < 0.05$) for ascribing significance and, by default, its value to research findings [20,21,34], little attention has been paid to: (i) its contribution to the false discovery rate [21]; (ii) its initial intent being as a measure of whether or not a data set should be taken seriously [35] or is "*worth another look*" [21]; (iii) significance chasing [36]; (iv) the fact that there is often no essential difference between a P value of 0.05 and one of 0.06 [34]–other that what an observer may wish to see–and; (v) the fact that statistical significance does not necessarily imply biological relevance in terms of the effect size [19,20,21,24] leading to the old adage–"*statistically significant but biologically/clinically irrelevant*" [37].

3. Quipping the knowns

Both reagent validation and appropriate statistical usage represent, in the vernacular of Rumsfeld's increasingly appreciated "*knowns*" [38], "*known knowns*". The latter were recently described as "*things an organisation refuses to acknowledge that it knows*" [39] with the organization for the purposes of this Commentary being the biomedical research community.

In addressing the decade plus productivity crisis in pharma that is, putting aside the debatable semantics of the actual number of drug approvals per year, a reflection of the disconnect between research investments and approved drugs [40], various solutions that have been proposed. These involve the use of the algorithms, metrics and jargon favored by management consultants to assess productivity and to aid in decision making in preclinical and translational research [41]. While apparently well intentioned, these solutions tend to adopt perceived best practices from other industries, few if any of which are comparable to the biomedical research endeavor–whether in academia or the biopharmaceuticals industry–that relies on unraveling and utilizing the intrinsic complexity of living biological systems the functional wiring of which–unlike that of computer chips–has yet to be convincingly elucidated and understood [42]–especially when their dysfunctionality is the root cause of human disease.

Notwithstanding, considerable enthusiasm remains in the biomedical research community for codifying simple "rules" that can be used in decision-making in much the same way as the binary statistic of $P < 0.05$ is taken as a definitive "*thumbs up, thumbs down*" to the significance of a research finding.

3.1. The Rule of 5

In drug discovery, probably the most widely known of these rules is Lipinski's "Rule of 5" [43]. Based on an historical data set of drug candidate compounds, this rule focuses on the ability to provide hard data on 5 key physical characteristics of a compound that can be used as a *sine qua non* to identify, improve and advance drug-like compounds to the clinic. Such rules can provide a level of comfort and certainty by reducing the degrees of freedom in the decision making process to a numerical check list and in doing so avoid the ambiguity of guidelines, or the entropy and sordidness of the intellectual thought process. However, these do not necessarily operate in a vacuum and have real world limitations [44] that too often invoke "the exception to the rule".

3.2. The Rule of 3

The most recent iteration of a drug discovery "rule" is the "rule of 3" for phenotypic screening [10]. This is intended to provide a series of criteria for the selection and use of phenotypic screening systems to improve compound identification metrics in drug discovery. These are assessed via the somewhat subjective measures of: (i) Disease relevance of the assay system; (ii) Disease