EDITORIAL

Reproducibility - again

Ole H. Petersen (Orcid ID: 0000-0002-6998-0380)
School of Biosciences, Cardiff University, CF10 3AX, Wales, UK
Email: PetersenOH@cardiff.ac.uk

There has been, and continues to be, widespread concern about the so-called reproducibility crisis in science. Countless editorials and opinion pieces in journals and newspapers continue to discuss this issue. It would appear that many science administrators, journalists and politicians, including some scientists, believe that most of the findings reported in scientific publications cannot be reproduced and therefore are useless. Some even assert that the apparently growing crisis is due to widespread fraud, fuelled by problems with current research and publication practices.

I was faced with the problem of irreproducible findings very early in my career. I started unsupervised experimental work on the electrophysiology of gland cells as an undergraduate medical student at the University of Copenhagen in the mid-1960s. Together with another medical student from my class, we discovered that key experimental findings, by a senior and eminent physiologist, which had been published extensively and prominently, could not be reproduced and were simply wrong. Inevitably, senior colleagues found it difficult to accept that two undergraduate students could have disproved work published by a senior highly regarded professor and it proved very difficult to get the correct data published. However, after a considerable delay, it did happen and the literature was permanently corrected (Petersen, 1992). I am convinced that there was no fraud involved in the generation of the experimental results that turned out to be wrong, but there were definitely serious technical flaws and inadequate protocols. These were never exposed in detail as many journal editors, then and now, discourage explicit criticism of papers they have previously published.

During a long career, I have on five separate occasions had to deal with publications, of immediate importance for my work, that turned out to contain irreproducible and manifestly misleading data. In order to be able to move on, it was in each case necessary for me to publish papers correcting the literature. In all these instances, the principal problems were poor technique, sub-optimal protocols and inadequate preparations. My experience is not unique and not new. Alan Hodgkin (1976), for example, mentions instances of irreproducible and wrong results from other laboratories that initially confused his work.

Currently, much attention is given, by both funding agencies and journal editors, to the problem of inadequate statistical analysis, but in the irreproducibility cases I had to deal with, statistics was not the problem. In some cases, the statistical analysis was excellent but, unfortunately, the primary data were not! In my own field, cellular electrophysiology and imaging, I am convinced by individual clear and continuous records with a firm baseline and then an intervention that produces a clear change that is reversible. This is of course the classical approach to electrophysiological work, best exemplified by the famous papers from Hodgkin and Huxley published in *J Physiol* in the 1950s (for example, Hodgkin and Huxley,1952). In these papers, there are no statistics. Typical, and very convincing, traces are shown and the number of experiments is recorded. That is all! That said, there are of course many areas where statistics, and sometimes sophisticated statistics requiring considerable expertise, is required.

The personal experiences, with previously published irreproducible data described above, may seem to fit into the general narrative about the problematic nature of the published scientific record. However, in order to put my own negative experiences into a proper perspective, one should consider that my five papers correcting wrong and irreproducible data only constitute a tiny proportion of my original research papers. Overwhelmingly, the work of others has been a good basis for my own work. Whereas many conclusions and model concepts, including some of my own(!), have turned out to be wrong, the actual data have been reliable.

Reproducibility is not precisely defined and it is abundantly clear that the criteria used vary enormously even between scientists. Personally, I don’t think it is

particularly helpful to agonise over more or less sophisticated definitions. The broad issue is whether scientific research published in proper peer-reviewed journals is, on the whole, helpful for scientific progress or is mostly wasteful because it is deeply flawed. A recent very thoughtful paper (Faneli, 2018) questions, on the basis of the evidence available, whether there really is a reproducibility crisis and whether it is helpful to act as if such a crisis exists. Faneli (2018) concludes that the crisis narrative is not supported by recent meta-research studies and that the problems that do exist are not distorting the majority of the literature. Furthermore, there is no evidence that the problem is growing.

In my experience, the self-correcting nature of science does work, but often rather slowly and often in a manner that fails to explain fully how wrong results came about, thereby delaying a full understanding of how to proceed on a safer basis. The self-correcting nature of science is, in my opinion, its most important and positive feature. The scientific journals play a crucial role here. While clearly having an obligation to review carefully and critically manuscripts that deal with corrections of previously published papers, journal editors must be careful not to fall into the trap of trying to protect the reputation of previous authors by demanding excessive quantitative information, thereby unnecessarily delaying or even potentially preventing corrections of the published record. My personal experience may be relevant here. The five papers with irreproducible data of interest to my own research, mentioned above, were published in five different journals. One of them was in *J Physiol*. In this case the self-correcting nature of science worked as it should. After the usual careful evaluation, including the normal revision process, my correcting paper was promptly accepted and published in *J Physiol*. I had a similar good experience with *Nature*. Unfortunately, in the three other cases, my correcting papers were held up for an unreasonably long time and in two of these instances the papers with the correct data were finally not accepted in the journal that had published the wrong results, but had to published in other journals.

Even if Fanelli (2018) is right in denying the existence of a reproducibility crisis, one should not be complacent and we should all do our best to reduce the number of publications with irreproducible data. Clarity of presentation, quality of recordings and images, appropriateness of experimental protocols and internal consistency are

all crucial features that can and should be checked by referees and editors. There clearly are cases where lack of transparency, stemming from poor reporting, has ‘swept problems under the carpet’ and contributed to the difficulties of others to check and, if necessary, correct wrong results. It was for that reason that the 2010 ARRIVE (**A**nimal **R**esearch: **R**eporting of **I**n **V**ivo **E**xperiments) reporting guidelines, to which *J Physiol* adheres (see Instructions to authors), were written. A 20-item checklist covers the key information that is needed in a scientific paper. However, the problem with checklists is that they may create the impression that all the issues highlighted are equally important, which can distort efforts and therefore be unhelpful. I therefore agreed to be member of an international working group, including editors and funders, aiming at improving the ARRIVE guidelines. One of the aims of this on-going work is to “organise the items in the ARRIVE guidelines into tiers reflecting different levels of priority” (Percie du Sert *et al.* 2018).

The increasing number of publications that we all have to deal with poses significant challenges and in this context it is more important than ever to promote efforts to improve the clarity of scientific papers. Improved guidelines may be part of this process. However, it is not helpful to promote the probably incorrect assumption of a growing reproducibility crisis. Some suggestions I have heard from some influential individuals, indicates to me that certain ‘cures’ (for example, regular audits of ALL recorded material) could be worse than the ‘disease’ by drastically reducing the essential playful creativity needed for real progress.

While scientific workshops, symposia and congresses are important instruments in the continuing process of correcting and expanding our knowledge, the scientific journals are unquestionably the most important institutions in these processes. From my personal experiences as author, referee and editor, I conclude that the Scientific Society Journals perform their duties better, and more efficiently, than other organisations. *J Physiol* has an outstanding record of publishing important papers with data that have stood the test of time. Our journal has, unlike many others, had very few retractions and the cited half-life of our papers is longer than for other prominent publications. *J Physiol* also has a long-standing and continuing tradition of publishing very detailed Methods sections, which is of vital importance for reproducibility.

References

Petersen OH (1992) Stimulus-secretion coupling: cytoplasmic calcium signals and the control of ion channels in exocrine acinar cells. *J Physiol* **448**, 1-51.

Hodgkin AL (1976) Chance and design in electrophysiology: an informal account of certain experiments on nerve carried out between 1934 and 1952. *J Physiol* **263**, 1-21.

Hodgkin AL and Huxley AF (1952) Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *J Physiol* **116**, 449-472.

Faneli D (2018). Is science really facing a reproducibility crisis, and do we need it to? *Proc Natl Acad Sci USA* **115**, 2628-2631.

Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Altman DG, Avey MT, Baker M, Browne W, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P, Howells DW, Karp NA, MacCallum CJ, Macleod M, Petersen O, Rawle F, Reynolds P, Rooney K, Sena ES, Silberberg SD, Steckler T, Wurbel H and Holgate ST (2018) Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Science* **2**, e000002.