PERSPECTIVE

The research data reproducibility problem solicits a 21st century solution

Bhupinder Bhullar

Science Publication Network, Birsigstrasse 10, CH-4054 Basel, Switzerland

Abstract: Reproducibility is a hallmark of scientific efforts. Estimates indicate that lack of reproducibility of data ranges from 50% to 90% among published research reports. The inability to reproduce major findings of published data confounds new discoveries, and importantly, result in wastage of limited resources in the futile effort to build on these published reports. This poses a challenge to the research community to change the way we approach reproducibility by developing new tools to help progress the reliability of methods and materials we use in our trade.

Keywords: data reproducibility, scientific method

Correspondence to: Bhupinder Bhullar, Science Publication Network, Birsigstrasse 10, CH-4054 Basel, Switzerland; Email: Bhupinder.bhullar@gmail.com

Received: July 15, 2016; **Accepted:** September 13, 2016; **Published Online:** October 31, 2016

Citation: Bhullar B, 2016, The research data reproducibility problem solicits a 21st century solution. *Advances in Precision Medicine*, vol.1(2): x–x. http://dx.doi.org/10.18063/APM.2016.02.006.

Introduction

Sir Isaac Newton once stated, "If I have seen further, it is by standing on the shoulders of giants"[1]. He was alluding to the fundamental principles underlying the discipline of science — which is that we build on previous discoveries. Replication and corroboration of research findings are a significant element to this process. The importance of replicating research results cannot be understated — because the failure to corroborate findings has negative consequences on society, sometimes with direct effect on human health and welfare. Systematic analyses by several independent groups indicate that lack of data reproducibility amidst published papers is a bigger problem than most scientists may be willing to admit. This is a challenge for the research community to collectively respond by developing new processes, standards, and metrics that are aligned with the way scientists work.

Real fraudulent intentions in publications are rare, but this assertion is one of the first that comes to mind when reproducibility problems are presented. The reason for this is largely set in by sensational media headlines of scientific misconduct that attract public attention. Empirical data from analyses of publications since 1992 suggests that real misconduct in scholarly works is less than 1% of all publications[2].

Several teams have attempted systematic investigations on the lack of reproducibility and reported that a large percentage of papers in high impact, peer-reviewed journals cannot be replicated. The lack of reproducibility was assessed by the inability to replicate the major assertion of the publication. Ioannidis *et al*. reported the inability to reproduce 90% of primary data from 18 microarray studies[3]; Prinz *et al*. reported a 65% lack of reproducibility in 67 publications in oncology, cardiovascular and women's health[4]; and Begley and Ellis, found that 89% of oncology papers they internally attempted to verify were not reproducible in their major claims[5]. The lack of reproducibility, at a ballpark frequency of 65% to 89%, is much higher than the rate of fraudulent work at ~1% of publications[2]. These results may indicate that there are problems in the way research is corroborated between research labs, as opposed to outright fraud.

Low Reproducibility: Reasons and Impacts

Here, I primarily focus on non-reproducible work that occurs due to routine issues in laboratory processes, such as reagent and measurement issues, which might address low reproducibility between research groups. We researched into the causes of reproducibility by interviewing scientists in Switzerland. Researchers in biology (n=60, from PhD level students to principal scientists), at four institutions (University of Geneva, University of Basel, University of Zurich, and École Polytechnique Fédérale de Lausanne), were asked questions about situations where they were unable to reproduce published methods and what might be the causes behind this. Of these interviewees, 100% stated that they were unable to reproduce published methods at one point or another. This survey was not meant to be exhaustive, but a starting point to design a study for further investigation. We were able to categorize their interview responses into the following arguments for reasons of low reproducibility of published methods which I report here as food for thought:

- Poor quality of reagents, statistical analysis, or data measurement methods,
- Imprecise reporting on the methodological protocols,
- Inability for direct communication among authors,
- Complex nature of modern scientific research (technically demanding),
- A lack of accountability for researchers,
- A publish-or-perish culture in academia that rewards quantity over quality of research,
- Funding bodies and academic journals that value "novelty" over replication.

All of the researchers we surveyed experienced frustration in not being to replicate experimental outcomes. Our survey data exposed that scientists would be interested in resources or means to boost the reliability of published data. But what is the real impact of this lack of data reproducibility? We lack systematic data on the economic and societal impact on non-reproducible work, because currently these discrepancies are not widely tracked. In comparison, investigations into papers that were retracted because of fraud give a sample accounting of cost and impact of non-reproducible work.

Stern *et al*. looked into financial costs associated with published papers that were retracted between 1992 and 2012[2]. The team looked for grant numbers associated with the publication and traced back the amounts of the grant awarded by US funding agencies. The total awarded grant was divided by all of the publications that were associated with the grant number. Using a sample set of 43 retracted publications, where complete data was available, Stern *et al*. calculated a median financial cost of $239,381 per article (minimum: $7061.95 and maximum: $3,608,713.94)[2].

Steen investigated 180 retracted medical papers which were primary patient studies conducted between 2000 and 2010[6]. These papers had been cited 5000 times. The 180 primary clinical studies and 851 "secondary" studies (secondary defined as citing the original retracted paper), collectively treated 98,501 patients. Steen also states that "papers retracted for fraud (n=70) treated more patients per study ($p<0.01$) than papers retracted for error (n=110)"[6]. Publications in high impact medical journals are likely to be viewed by many clinicians and used to establish treatment guideline for patients. For example, a published clinical study on hepatocellular carcinoma treatment using chemoembolization and radio frequency ablation, was retracted by JAMA in 2009 after a post-publication investigation revealed a flaw in the experiment design[7]. As not all clinical trial studies are published, this estimate of approximately 100,000 patients being put at risk is conservative because only individuals enrolled in published studies were tallied.

Because non-reproducible elements of publications are more numerous than outright fraud, the real cost on society is also significant. The obvious negative impact is the loss of research time and money invested by the laboratory attempting to replicate the benchmark data. This effect is compounded globally because there is currently no central platform where reproducibility of results can be communicated and shared. Over time, different labs around the world will pick up the same publication and attempt to repeat the experiments. The unknown effect of non-reproducible data, and perhaps the main indirect cost to society, is the hindered rate at which real innovation take place — such as the discovery of life-saving treatments for diseases. Without the ability to replicate benchmark data from publications, novel projects may not even get started.

Federal research funding agencies are paying attention and are implementing pilot projects to teach better methods to design and execute preclinical studies. For example, the National Institutes of Health Data Reproducibility office has developed training modules for scientists to improve experimental design[8]. In

addition, the recommendations from funding agencies and journals are to encourage the use of more appropriate statistics, meta-analysis, and transparency with regards to publication of methods. However, I believe that these efforts do not fully address the data reproducibility issues that are faced by the millions of research scientists around the world. The competitive economics of our current science culture favour novelty over reproducibility. This is suggested by the high level of retractions that are directly correlated with the impact factor of journals[9]. The metrics that measure scientific contributions of individuals are skewed toward rewarding "high impact" publications (both with grant money and career progression for the investigators). Journals normally do not publish reproducibility reports, and a typical journal contribution has a high overhead cost to the authors that limits this reporting.

Possible Solutions

I believe that to date, the research community lacks practical solutions to help address the roots of the reproducibility problem adequately. As the Switzerland survey of scientists indicated, the lack of transparency in research methods is a main hurdle to reproducibility. In addition, the variable nature of reagents (e.g., changing cell lines, impurities in chemical samples) and methods to measure and record data, impact the likelihood of repeatability. To facilitate an exchange of information on these parameters, the research community needs an independent online platform that allows scientists to publish focused reproducibility reports.

The "Reproducibility Reporting" platform should address many of the shortcomings of current publication and communication pipelines, while simultaneously taking advantage of the digital tools available today to enhance engagement and communication channels between researchers. For instance, allowing rapid publication of focussed reports will decrease the time to communicate and broadcast the results from reproducibility testing. Although the platform enables rapid reporting, the process cannot skimp on assessing the quality of the "Reproducibility Report". Guidelines for publishing discrepancy data need to clearly articulate the standards for publication, such as minimal statistical testing and carefully evaluating deviations in materials used and methods employed. One can imagine that as communities build on the platform, discussion parameters will become standardized over time to facilitate easier discourse between researchers.

Sharing the raw data of the reproducibility test will also enable insight into the discrepancies. Labs attempting similar methods can examine deviations in measurements or in the performance of reagents, such as cell lines, more easily. Open Data standards are emerging, which are currently used for government data, but can be adopted to publish raw science research data[10]. One advantage of employing Open Data standards is that the raw data will have a publication record and the original authors are credited for their work through citation. An added advantage is that researchers can potentially repurpose the data for a different analysis.

The current social media platforms, where data reproducibility is discussed, allow users to anonymously post content. This has resulted in posts that attempt to discredit publications and hurt the authors' credibility, without fully exploring reasons for the underlying lack of data reproducibility. The result is that the scientific community turns defensive and discussions of reproducibility problems have a cultural taboo association. The ability to facilitate moderated, constructive dialogue between peers on the "Reproducibility Reports" platform will circumvent the negative aspects of unregulated social media platform.

The straightforward approach to constructive dialogue is an empowered and vibrant peer review system. The standard journal peer review system uses a secret review by 2 to 3 reviewers. However, current technology allows crowd-sourcing the review amongst expert who may be more versed in that area. Crowd-sourcing takes the diversity of opinion of a larger number of independent experts who specialize in the field and can draw on local knowledge, and aggregates this information. Leveraging the wisdom-of-the-crowd may enable finding the "diamond in the rough" solutions or identify critical errors in the design of the experiment.

Anonymity of the reviewer may need to be preserved as part of the reviewing process, but because all users are validated by the platform, abuse of the platform by anonymity is not possible. Exploitation of reviewing privileges (such as conflict of interest) will be flagged through crowd-sourcing, because in small social networks, such as research communities, one can more readily identify misuse. To empower peer-review, the reviewers' comments should be visible to the community. This will increase debate and clarification by allowing the community to directly

comment on the review itself. But how could one incentivize credible and careful peer review by the community? I would like to propose an "Open Review" protocol. Scientific peer review of journal articles in submission is currently a requirement of academic scholarship. The peer reviewer is not paid, receives no credit for the review, and remains anonymous. Often the peer reviewer may have ideas that can advance the investigators' work. Under "Open Review" guidelines, the peer reviewer is further encouraged to improve the work under review bypublishing the review as an "opinion" article. By allowing the peer review to be published, the reviewer can protect their contribution under "creative" license and their work can be cited.

By adhering to "Open Access" guidelines, the global research communities can freely engage with the content created on this platform. Open Access will enable a broader reach, in particular, to researchers in developing countries, which should increase the diversity of dialogue on the platform. It is tempting to propose that such a platform, which enables faster elucidation of what works and what does not work, will enable global communities to advance their research programs rapidly. This is because the major output from a reproducibility platform will likely be a "standards database" for reagents and measurements that work.

Conclusion

It is important to emphasize that the reproducibility reports platform is complementary to the current research journals already established. Origination, novelty and innovation of scientific endeavours are still the engine that advances society. Communicating the reproducible nature of these original articles will stem the losses of valuable resources and allow researchers to focus on what works. Ideas published in research papers influence subsequent research, but vetted reproducible data will enable more researchers to stand on the shoulders of giants.

Conflict of Interest and Funding

There is no conflict of interest.

References

1. Wikipedia, n.d., *Standing on the shoulders of giants*, accessed June 25, 2016, <https://en.wikipedia.org/wiki/Standing_on_the_shoulders_of_giants>
2. Stern A M, Casadevall A, Steen R G, *et al*. 2014, Research: financial costs and personal consequences of research misconduct resulting in retracted publications. *eLife*, vol.3: e02956.
3. Ioannidis J P A, Allison D B, Ball C A, *et al*. 2009, Repeatability of published microarray gene expression analyses. *Nature Genetics*, vol.41(2): 149–155.
4. Prinz F, Schlange T, and Asadullah K, 2011, Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, vol.10(9): 712.
5. Begley C G and Ellis L M, 2012, Drug development: raise standards for preclinical cancer research. *Nature*, vol.483(7391): 531–533.
6. Steen R G, 2011, Retractions in the medical literature: how many patients are put at risk by flawed research? *Journal of Medical Ethics*, vol.37(11): 688–692.
7. DeAngelis C D and Fontanarosa P B, 2009, Retraction: Cheng B-Q, *et al*., Chemoembolization combined with radiofrequency ablation for patients with hepatocellular carcinoma larger than 3 cm: a randomized controlled trial. *JAMA*. 2008; 299(14): 1669–1677. *The Journal of the American Medical Association*, vol.301(18): 1931.
8. National Institutes of Health, n.d., *NIH Research Training — Rigor & Reproducibility*, accessed June 29, 2016, <http://www.nih.gov/science/reproducibility/training.htm>
9. Fang F C, Casadevall A and Morrison R P, 2011, Retracted science and the retraction index. *Infection and Immunity*, vol.79(10): 3855–3859.
10. *The Open Data Substrate*, n.d., accessed June 29, 2016, <http://open-data-standards.github.io/>