A failure to reproduce: How bad biomedical science is holding us back [version 1; peer review: 1 approved with reservations, 2 not approved]

Hussein Jaafar¹, Robert M. Maweni²

¹Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
²Croydon University Hospital, London, UK

Abstract
Irreproducibility is a common problem in the biomedical sciences. Numerous studies have revealed the systemic and chronic nature of the problem, yet not enough is being done to combat it. The financial cost is estimated to be 28 billion dollars in the United States alone. Combine this financial cost with the time spent on irreproducible studies and the net effect is staggering. The factors for this lack of reproducibility are however identifiable and concrete steps can be taken to improve the situation. This article describes some of the factors leading to irreproducibility in the biomedical sciences and how stakeholders at every level of the field can act to reverse them.

Keywords
Reproducibility , Biomedical Science , Irreproducibility , publishing , funding , standards and practices , institutions

Corresponding author: Hussein Jaafar (hussein.jaafar@kcl.ac.uk)
Competing interests: No competing interests were disclosed.
Grant information: The author(s) declared that no grants were involved in supporting this work.
Copyright: © 2016 Jaafar H and Maweni RM. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
How to cite this article: Jaafar H and Maweni RM. A failure to reproduce: How bad biomedical science is holding us back [version 1; peer review: 1 approved with reservations, 2 not approved] F1000Research 2016, 5:415 (https://doi.org/10.12688/f1000research.8370.1)
A tale of two papers

In 2005 a paper was published by Dr. John Ioannidis entitled “Why most published research findings are false” (Ioannidis, 2005b). It was essentially Dr. Ioannidis’s claim that “false findings may be the majority or even the vast majority of published research claims”. The strange thing about this paper was that it wasn’t exaggerating.

In his paper Dr. Ioannidis used a mathematical proof that assumed modest levels of researcher bias. This bias could be human error, bad methodology or any number of other factors. He argued that a sufficiently motivated researcher that wishes to prove a theory correct, can do so most of the time regardless of whether the theory is actually correct. The rates of “wrongness” his model predicted in various fields of medical research corresponded to the observed rates at which findings were later refuted. And these rates of “wrongness” were not insignificant. 80 percent of non-randomized studies, 25 percent of “gold-standard” randomized trials and even as much as 10 percent of the “platinum-standard” large randomized trials turn out to be irreproducible.

A second paper by Dr. Ioannidis was published that same year. In this paper Dr. Ioannidis looked at 49 of the most cited articles in the most citied journals (Ioannidis, 2005a). 45 of these papers claimed to have described effective interventions for various diseases ranging from heart attacks to cancer. Of these 45, seven were contradicted by subsequent studies, seven others had their reported effects diminished by subsequent studies and 11 were largely unchallenged. Only 20/45 (44%) of the field guiding papers had been replicated successfully. And in a finding that shows how these irreproducible papers are impacting the field, Dr. Ioannidis found that even when a research paper is later soundly refuted its findings can persist, with researchers continuing to cite them as correct for years afterwards.

The counterargument is of course that despite all this the system clearly does work on the whole. Even if mistakes are being made and inefficiency is rampant, if something is wrong it will eventually be found out and corrected. Take for example the now infamous recent controversy regarding stimulus-triggered acquisition of pluripotency cells or STAP cells. Published in Nature, these two papers were considered a massive breakthrough in stem cell research (Obokata et al., 2014a; Obokata et al., 2014b). However very quickly problems emerged with the data presented in the paper. An investigation by the hosting institute found the lead investigator guilty of misconduct and the papers were subsequently retracted (Editorial, 2014). This is one example showing how the checks and balances in place within biomedical research can work and work well. Despite these measures the price of irreproducible research remains a substantial one.

The price we pay

A recent study estimated the cost of irreproducible pre-clinical research at 28 billion dollars in the US alone (Freedman et al., 2015). The study estimated the overall rate of irreproducibility at 53%, but warned that the true rate could be anywhere between 18% and 89%. While the exact figures are certainly debatable the clear message is that this is a significant issue even if we assume the rate of irreproducibility is at the lower end of their scale. Even big pharma companies have noted the lack of reproducibility coming from academia. They report that their attempts to replicate the conclusions of peer-reviewed papers fail at rates upwards of 75% (Prinz et al., 2011).

Going hand in hand with the financial costs we cannot forget about the time invested on these irreproducible studies. One could argue that this is indeed the more damaging factor given that it slows the development of potentially lifesaving treatments and interventions that would significantly improve quality of life for large proportions of society. This is of course an even more difficult if not impossible metric to measure accurately but it logically must be impacted upon. For example consider this: a researcher has a hypothesis and carries out 3 experiments to test this hypothesis. The first 2 experiments are successful and seem to confirm the researcher’s hypothesis. The findings from these first two experiments are thus published. This naturally leads to a third experiment which seems to strongly disprove the hypothesis and so the researcher abandons this line of work to move onto something else. The researcher is unlikely to publish the findings of the third experiment but the 2 other papers will remain published. This may lead other research groups to continue this work from the first 2 papers, perhaps even carrying out the same failed experiments, unaware that it had already been tried and rejected. Again it is difficult to know how often something like this happens simply because we don’t know how often researchers are leaving their negative results unpublished. However even a cursory read through some biomedical journals will reveal that papers with negative results are few and far between.

Yet another group that believes this is a major issue in need of addressing is the Global Biological Standards Institute or GBSI. The GBSI carried out a study in which they interviewed 60 individuals throughout the life science community including biopharmaceutical R&D executives, academic & industry researchers, editors of peer-reviewed journals, leaders of scientific professional societies, experts in quality management, experts in standards, academic research leaders and many more disciplines (GBSI, 2013). In the extensive interviews with these professionals a systemic and pervasive problem with reproducibility was reported. Over 80% of academic research leaders they interviewed had some experience with irreproducibility. The reasons for this irreproducibility included inconsistencies between labs, non-standardised reagents, variability in assays, cell lines, experimental bias, differences in statistical methods, lack of appropriate controls and several others.

Why we falter

There is a perception amongst the general public that scientists are a group of meticulous, highly organised, extremely intelligent section of society (Castell et al., 2014). This perception is certainly not without a basis in reality but fails to appreciate the human aspect of scientists and our work. Mistakes happen, negligence occurs. Politics, money, bureaucracy and rivalries all get in the way of scientific research (GBSI, 2013; Wilmshurst, 2007). This all happens on a pretty regular basis according to the GBSI report with one researcher quoted as saying “We’ve had to address issues with replicating published work many times. Sometimes it’s a technical issue, sometimes a reagent issue, sometimes it’s that the technique was not being used appropriately” (GBSI, 2013). Within the
biomedical research community these obstacles are a disliked but tolerated part of doing science. They are unfortunately sometimes considered part of the job and just how the system works as is demonstrated by their prevalence (Tavare, 2012; Wilmshurst, 2007).

This is a dangerous and irresponsible attitude to allow to continue within our community. It is precisely because of our line of work that we should seek to uphold to highest professional and academic standards. Our work which can quite literally be the difference between someone living or dying, or having to suffer from debilitating illness on a daily basis. Just because our impact is delayed by its long journey from bench to bedside does not make it any less crucial to people’s lives.

Yes, there are reasons for things being the way they are. The publishing process is outdated. There’s never enough funding to go around and what little there is goes often has strict criteria attached to it (Editors, 2011). When applying for positions in academia, publications are king with quantity being above quality or accuracy in many cases and these publications are being dominated by a select few in the field (Ioannidis et al., 2014). It is therefore unfortunately often necessary for scientists to play the game and submit to the demands of the system. Corners are cut, statistics are “reinterpreted” and results exaggerated all in the hopes of getting past the journals review panel, or having a grant proposal approved (Baker, 2016; GBSL, 2013). None of this is maliciously done of course, at least not usually, but malicious or not the negative effects are the same. Whatever the reason it’s still bad science.

The solutions

So what is the solution? First we must appreciate that the problems leading to this situation are clearly multifaceted and require concerted efforts from groups and individuals at all levels throughout the community. We will discuss 4 of the main areas that could be improved upon.

Publishing/Journals

One change that could have the most wide reaching effect would be a greater emphasis on the importance of replicability in studies. Just as the number of citations a researcher has today is considered an important metric for the quality of their work, the number of a researcher’s papers that have been reproduced by other groups should be strongly considered too. Standardised metrics such as this will help place greater importance on the quality of a piece of work, rather than just the exciting nature of its claims. This can help us to reduce the pressure to publish in the highest impact factor journals. With standardised metrics the prestige of a journal won’t necessarily matter as long as the papers quality metrics are solid.

For this to work however journals need to be more accepting of papers whose sole purpose is to reproduce or confirm another group’s work instead of favouring papers that report new discoveries or interventions. In addition journals could allow researchers to pre-register their planned experiments with them in exchange for a potential fast track to publication if they then carry out those experiments, even if they give negative results. This would go a long way to improving transparency and encouraging researchers to not discard unfavourable experiments. It would also help avoid situations like the one we discussed above where researchers continued to work on the basis of experiments they were unaware had already been disproven. Some examples of initiatives moving towards these goals include the All Trials campaign and the registered reports approach (AllTrials, 2016; Chambers, 2016).

Standards and practises

Next there needs to be an expanded development and adoption of standards and best practices. Standards can be physical standards like standardised assays and cell lines or they can be documental standards such as protocols and practises. To do this guidelines for best practises and standards need to be easily accessible and widely available, which is one of the aims of the EQUATOR network (EQUATOR, 2016). Improvement in standards is arguably the most important factor for increasing reliability but given its wide ranging, complicated and technical nature we will leave discussion of it for others to pursue.

Funding bodies

Similar to journals, funding bodies could introduce mechanisms to reward reproducibility in a researcher’s work. They should look at an investigator’s past record with producing reproducible work (new journal metrics would complement this) and also look at his/her current grant application to see if it’s set up to produce reproducible results. The NIH for example has introduced new guidelines beginning in January 2016 to improve reproducibility in grant in applications (NIH, 2016). The 4 main areas the guidelines seek to address include the scientific premise of the proposed research, rigorous experimental design for robust and unbiased results, the consideration of relevant biological variables and the authentication of key biological and/or chemical resources proposed in a grant application.

Investigators/Institutions

Finally and perhaps ultimately, the responsibility of producing high quality reproducible work is down to the principle Investigators, their lab group and the institution that hosts them. These are the people that can have the most immediate impact through self-correction and adherence to the best standards and practises whenever and wherever possible.

The bottom line

This article is in no way all-encompassing regarding the issue of reproducibility. Several problems and solutions exist which we have not discussed in detail, not least of which the troubling subject of researcher bias, for example. The references below do however discuss some of these in greater detail. Many will argue that implementing the above proposals will of course require additional work and possibly some significant upfront costs but we would counter that the longer term impact to biomedical research would be immense and one we cannot afford to miss out on.
Author contributions
HJ conceived of the article topic. HJ and RM contributed to researching, writing and referencing the article. HJ drafted the manuscript which RM reviewed and agreed to the final content therein.

Competing interests
No competing interests were disclosed.

Grant information
The author(s) declared that no grants were involved in supporting this work.

Acknowledgments
Thank you to Matt, Laura and Zein for reading and advising me in the course of writing this article.

References

AllTrials: All Trials Registered. All Results Reported. AllTrials, 2016.

Reference Source


PubMed Abstract | Publisher Full Text


Reference Source


Reference Source


PubMed Abstract | Publisher Full Text

Editors T: Dr. No Money: The Broken Science Funding System. Sci Am. 2011.

Reference Source


Reference Source


PubMed Abstract | Publisher Full Text | Free Full Text


Reference Source


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


Reference Source


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text
Open Peer Review

Current Peer Review Status: ✗ ✗ ✗

Version 1

Reviewer Report 15 July 2016

https://doi.org/10.5256/f1000research.9003.r13423

© 2016 Chue Hong N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Neil Chue Hong

Software Sustainability Institute, University of Edinburgh, Edinburgh, UK

I carry out this review following the guidelines set in "An Open Science Peer Review Oath".

There have been increasing scrutiny of the many fields of research, with reproducible research being used as one of the key drivers for many different concerns: trust, economic efficiency, reliability, transferability of research research.

This opinion article by Jaafar and Maweni presents an overview of previous research, current status of and potential solutions for irreproducibility in the biomedical sciences.

Overall, I thought the structure and style of the paper was suited to an opinion article, but felt that in its current version it lacked the "authors' perspective on a topical issue that has not yet been covered in the same way in the existing literature". Basically, it does not cover more ground or add the unique perspective of the authors to make it stand out from existing articles in this area.

Whilst the authors have done a good job of collecting together an up-to-date reference list of studies in the area, it's not clear to me that the opinions they express in the article move the discussion on further from comment pieces such as Begley or work by journals. Many of their arguments are at a high level e.g. advocating for adopting of standards is not backed up with an opinion on which types of standards will lead to improvements in which particular types of reproducibility.

I think that the area where the authors could most readily formulate, justify and argue an opinion would be in the area of investigators and institutions - what are their roles, and how do they interact. Do the authors perceive the drivers of each to be complementary or divisive? What will this mean for the future of biomedical research?

To do this, I think they need to extend their literature search to cover more on the related work to do with trends in the way research is being carried out in the biomedical sciences, and related work on the influences and incentives for principal investigators, researchers and heads of research of at institutions carrying out the research, and then provide an opinion on whether these can be used to address the
issues of irreproducibility in the biomedical sciences that they summarise.

In addition, I think it would be useful in an opinion piece to give some element of personal perspective - as researchers working in the field of biomedical sciences themselves, the authors could give an opinion on how it affects them in their day to day work, and what solutions they feel would work if applied in their own institutions.

I am categorising this as "Not approved" as even though my criticisms of the piece could be addressed with specific, major revisions, I believe that this opinion article needs substantial new material to be added for it to be a useful addition to the literature in this area and therefore does not meet the criteria for an F1000 opinion article. A revision of the paper, concentrating on extending the second half would be the most obvious suggestion to address this. As an alternative suggestion to the authors, it may be that this article would be better presented as a review article if taken forward in its current form.

References

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 09 May 2016
https://doi.org/10.5256/f1000research.9003.r13428

© 2016 Gorgolewski K. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Krzysztof Gorgolewski
Department of Psychology, Stanford University, Stanford, CA, USA

This is a very well written paper discussing the important and timely topic of reproducibility. As an opinion piece it is factually accurate, but despite promising introduction to the topic, it provides very little in terms of practical solutions. Listed proposal lack detail (How is the replication going to be defined in the new metric? Who decides if experiment A replicates experiment B? Why would commercial publishers bother implementing and maintaining such metric?) and poor understanding of the complex system of incentives in the academic world (Why the NIH does not have a stronger position on replicability and data sharing? Why there aren’t enough good quality standards in life sciences? Why PIs chase high impact factor journals instead of replicating other people work?).
Unless the second part of the paper undergoes a major revision and provides practical, detailed and feasible solutions I doubt this publication will have a non-negligible influence on the reproducibility crisis.

Other comments:
- You say "When applying for positions in academia, publications are king with quantity being above quality (…)" - all I keep hearing is that it takes one Nature paper to get a tenure. Do you have any evidence of the "more papers is better than good papers in terms of landing a job" claim?
- You also mention the it is hard to assess how many null results are not published. However, in the context of meta-analysis there are techniques to assess publication bias from the expected shape of the effect size distribution. Isn't there any piece of meta-research assessing what percentage of experiments should by chance yield a null result? It would be definitely worth looking for.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

Gary G Borisy  
Department of Microbiology, The Forsyth Institute, Cambridge, MA, USA

Jaafar and Maweni are entitled to their opinion but I don't believe they have contributed in a sufficiently significant way to warrant being indexed. They provide no original analysis of their own; they provide a brief overview of contributions by other authors and they provide only a brief, cursory statement of "solutions". The proposed solutions in part reiterate suggestions of others but a more serious problem is that they miss the mark in how basic research is actually done. Productive researchers rarely replicate previous work explicitly. They build on previous work. To successfully build essentially validates the previous work but the point of research is to extend into the unknown, not to merely replicate. From this point of view, the emphasis on 'reproducibility' in their solutions does not capture the heart of the matter.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

---

Hussein Jaafar, King's College London, London, UK

Thank you for your review Dr. Borisy. We would like to respond to your critiques.
“They provide no original analysis of their own; they provide a brief overview of contributions by other authors and they provide only a brief, cursory statement of "solutions".

This critique is understandable however this is an opinion article and as such is it intended to be brief and not all encompassing, as a full review might be. According to F1000Research guidelines "Opinion Articles give the authors’ perspective on a topical issue, providing a balanced view of different opinions in the field". We believe this article achieves this. As such your comment about not providing an original analysis is a little perplexing. This is not a piece of research nor does it proclaim to be. The purpose of an opinion article as we see it is to gather relevant information and present that information in a logical and educational format, along with, of course, our opinions on the matter.

“The proposed solutions in part reiterate suggestions of others but a more serious problem is that they miss the mark in how basic research is actually done. Productive researchers rarely replicate previous work explicitly. They build on previous work. To successfully build essentially validates the previous work but the point of research is to extend into the unknown, not to merely replicate.”

It is true that basic research builds on previous work and success would validate that previous work. But the point of this article is to highlight the fact that in the process of building on that previous work much time and money is wasted on research that is simply incorrect. We are not suggesting that the point of research is to merely replicate. We are saying that it would be of great benefit to science if we could make our experiments more replicable from the onset and subsequently emphasize and merit researchers whose work can be consistently replicated (either by building on it or by explicitly replicating it).

It’s important for us to highlight that when we use the word “reproducible” we are clearly emphasising the implementation of procedures and systems which would encourage reproducibility at the experimental design stage. We are not exclusively speaking about reproducing already completed studies. We are talking about making experiments inherently more likely to be successfully reproduced from the get go, thus saving a lot of time and money down the road.

**Competing Interests:** Article Author

---

**Comments on this article**

**Version 1**

Reader Comment 06 May 2016

**Omar González-Santiago**, Universidad Autónoma de Nuevo León, Mexico
Standing on the shoulders of giants should be the premise in scientific research but where are the real giants, the real data. Any scientific research should be tested with acid and resist the replication process. Who can assure that data could not be fraudulent?. The scientific community must ensure that any scientific knowledge is actually a giant and not a Lilliputian. I agree with your opinions

**Competing Interests:** None

Reader Comment 04 Apr 2016

**Shaun Lehmann,** Australian National University, Australia

I am glad that you are highlighting this issue as it is of real importance.

I am currently working in phylogenetic method development and I have had quite a lot of difficulty in feeling confident about reproducing published phylogenetic work (which needless to say is a very common part of a variety of biological analyses).

The primary issue I have found is that while authors routinely supply GenBank accession details for sequences they have used, the alignment itself is almost never supplied. This is a serious issue. There is no guarantee that reproduction of the alignment and alignment curation process (according to an all to brief description of the alignment process) will result in an identical alignment to the one used in the paper, and no way to check without the alignment itself. If one is stringent in how they define reproducibility, this is simply not reproducible work.

As can be seen, the above issue exists even before one delves into the complexities of phylogenetic methods themselves and their own peculiarities of reproducibility. Very troublesome indeed.

A rudimentary search of recent phylogenetic analyses in Google Scholar will provide you with dozens of examples. To the author: please feel free to use this as a specific case study if necessary.

**Competing Interests:** None

Author Response 30 Mar 2016

**Hussein Jaafar,** King's College London, London, UK

A spelling mistake in the abstract has been noted and will be amended after peer review.

**Competing Interests:** Author of article
- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com