

Preprints: a tool and a vehicle towards greater reproducibility in the life sciences

Iratxe Puebla

ASAPbio, Cambridge, United Kingdom
ORCID: 0000-0003-1258-0746
iratxe.puebla@asapbio.org

Abstract

We have seen a number of initiatives arise in recent years aiming to tackle concerns around the reproducibility of published findings. Researchers in the life sciences now have a range of tools at their disposal to boost the reproducibility of their science and preprints have emerged as an instrumental element within this toolkit. Preprints broaden the “when, by whom, and how” of the review and feedback on research compared to the journal publication process, help address publication bias, and can play an important role as a vehicle towards open science practices. Preprints hold further untapped potential to close the gap between discovery and dissemination, and to accelerate the path to a more reproducible research ecosystem.

There has been much debate about the reproducibility of research findings. Questions about whether published results could stand the test of time have been raised by scholars for decades, particularly those in the social sciences and medicine. Researchers working in clinical fields were well aware of the risks that publication bias and selective reporting of clinical trial data could pose for patient treatment, and these concerns led to the adoption of a clinical trial registration policy by leading medical journals in 2004 (1) in attempts to address the non-reporting of clinical trials findings. Just the following year a seminal publication by John Ioannidis suggested that a majority of the published findings are false (2), and adding to the doom and gloom, a project looking to directly replicate earlier published preclinical cancer studies reported a number of hurdles and difficulty in replicating the results (3).

So what’s new? The debate around reproducibility has gradually expanded from this initial focus on medicine and social sciences to a broader range of disciplines and, in a promising trend, the discussion has started to move from an articulation of the problem to more energized exchanges about potential solutions to tackle the reproducibility challenges.

Preprints as an important element in the reproducibility toolkit

In the same manner as research studies can address different research questions and follow different designs, there is no single golden equation to ensure reproducibility. We need a multipronged approach that takes into account the needs of specific disciplines and is inclusive toward the different types of outputs produced during scholarly research. Researchers therefore need a range of tools at their disposal, and the tools and steps needed to ensure reproducibility will vary from one study to another.

In the life sciences, the last decade has seen a number of initiatives added to this reproducibility toolkit: a drive towards data and code sharing, the start of Registered Reports, platforms for the sharing of protocols (4), the development of a myriad of reporting guidelines [see TOP guidelines (5) and those in the EQUATOR Network library (6)], and the emergence of preprints as a mode for research dissemination.

Preprints are not only a tool for increased transparency by themselves, but also a vehicle to synergize increased research reproducibility in combination with the other initiatives in this space. We will explore some of the main ways in

which preprints contribute to research reproducibility.

Preprints expand the review of the work and allow it to happen earlier in the research process

Many of the approaches proposed in the reproducibility space have a focus on bringing additional scrutiny and openness to the research process, at stages earlier than that of the traditional journal publication point.

Preprints provide a means of bringing this focus to an earlier stage of the research process for any type of paper. Authors retain control to disseminate their work when and as they are ready, and thus a preprint allows them to share their work earlier than they would have done via the journal publication path. Sharing their work at this earlier stage allows researchers to gain feedback from the community which they can then incorporate as they pursue further experiments or analyses, and polish the manuscript before an eventual journal submission (see Figure 1). Even if the authors choose to post their preprint in parallel to journal submission, they can still receive community comments in addition to those supplied via the journal process, either privately or via comments posted on the preprint record. Any comments on the preprint record will be openly available for readers, who also benefit from that feedback as they read the preprint, independently of when and in which journal the paper is eventually published.

In addition to enabling the possibility for feedback while the research is still taking place, sharing a preprint allows anyone in the scientific community to review and comment on the work. A journal's traditional review process generally involves one editor and two or three reviewers, meaning that a piece of research that may have taken years to conceptualize and carry out is then *post hoc* reviewed by a very small group of experts. This framework places a huge burden on the reviewers involved, a burden that has exacerbated over time as science has become increasingly complex and interdisciplinary. Is it

reasonable to expect a small group of two or three experts to be able to appraise all aspects of a multi-group interdisciplinary research, from its theoretical basis, to the data, code, analytical approach, and interpretation of what the findings mean for different fields? While research has evolved to often take place in a collaborative inter-connected, and interdisciplinary environment, the way peer review is run at most journals has hardly changed (the only exception being that it now takes place via online platforms instead of having to send printed copies of manuscripts by post).

In comparison, sharing research via a preprint opens the review of the work to anyone in the research community, from scholars working in different disciplines who can comment on interdisciplinary research, to early-career researchers who are underrepresented (and often uncredited when they participate) as part of the journal peer review process. Having more eyes on a paper early on provides a greater chance for potential flaws to be identified, allowing the authors to address the errors and prevent them from reaching the manuscript version that they submit for journal publication.

Preprints also allow authors to easily update their paper and post new versions. Authors retain control of when they'd like to post a new version and this is done in a process much more fluid than that required for corrections to the record after journal publication. Addressing errors in journal articles requires a time-consuming process of correspondence with editors and further review rounds, and it can take months or years to complete the necessary correction. In addition, retractions and editorial Expressions of Concern carry significant stigma for the authors, which can be a deterrent to proactively address potential issues in a publication. Preprints on the other hand, allow authors to easily make corrections to the paper where this is needed, in a matter of days and at a stage where making changes carries much lower reputational risk.

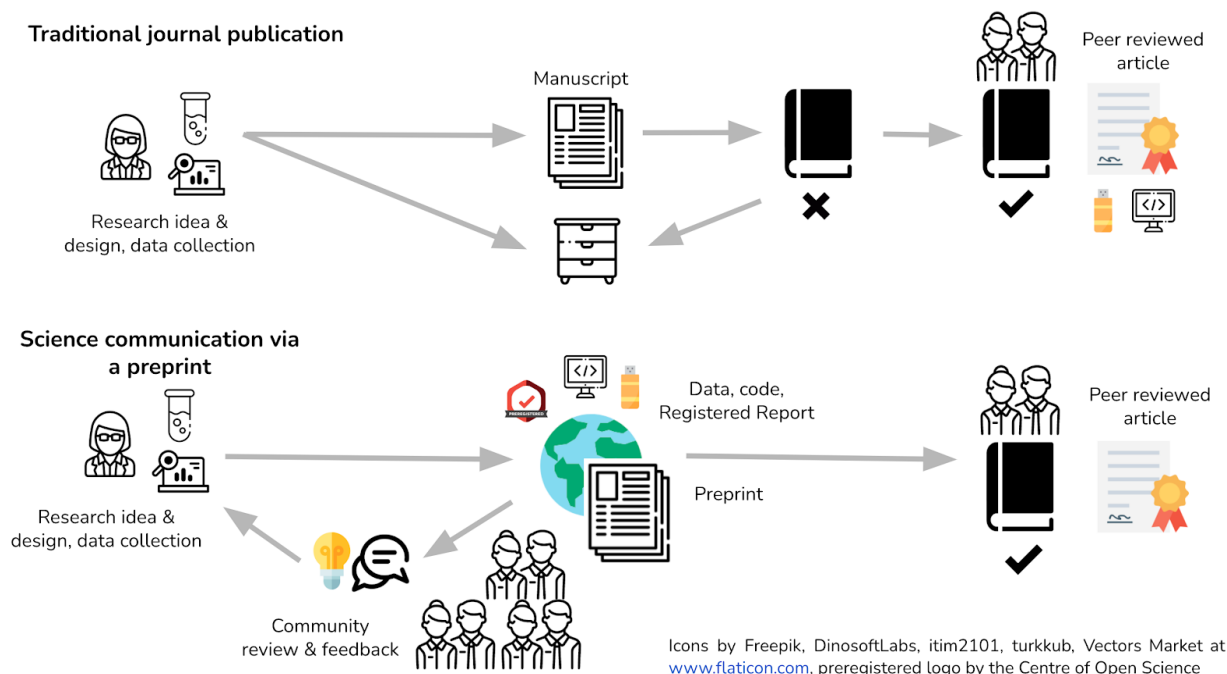


Figure 1. Science communication with or without a preprint. The top diagram represents science communication via the traditional journal publication path, the bottom diagram represents science communication via a process that involves the posting of a preprint. The preprint allows increased scrutiny of different research outputs as well as broader community feedback earlier in the process, increasing reproducibility.

Preprints help tackle publication bias

One important element in the concerns around reproducibility relates to publication bias, that is, the fact that studies reporting positive and significant results reach publication in a higher proportion to those that report negative or null findings. Publication bias can have major implications for research, from wasted time and investment, to the more dramatic risks for patients if important evidence needed to inform treatment choice is not made available [examples of this come from trial data on antidepressant drugs (7) or the anti-arrhythmic drug lorainide (8)]. Publication bias can be the result of different factors, either related to the researchers themselves, or to the journal publication process.

Researchers may sometimes be reluctant to publish results that did not support their original hypothesis, or may de-prioritize pursuing publication of results which did not reach the level of significance. This is generally not a deliberate attempt to conceal results but rather a pragmatic choice as researchers must prioritize

their time and the effort they'll invest in publications. A common example is that of papers in the pipeline that do not get completed as the student or postdoc working on the experiments moves to another position, and the collaborators are not in a position to pursue the project further due to time or funding constraints. All this leads to the so-called 'file-drawer' effect, as the results remain in the lab and do not make it to the body of literature. Preprints allow much more flexibility in format and length compared to a submission to a journal, reducing the burden on researchers to prepare the paper for dissemination, even if the results are preliminary and no follow-up experiments can be done, thus tackling the 'file-drawer' effect.

Another driver behind publication bias is that manuscripts that report negative or null results can find it harder to meet the favor of editors, as many journals select articles for publication based on the perceived impact of the reported findings. In an interview a few years ago, the plant biologist Ottoline Leyser rightly pointed

out the pitfalls of a culture around impactful or ground-breaking research: "Ground-breaking is what you do when you start a building. You go into a field and you dig a hole in the ground. If you're only rewarded for ground-breaking research, there's going to be a lot of fields with a small hole in, and no buildings" (9). Research relies on an iterative process where findings are validated by others and established as evidence over time; for this process to work efficiently researchers need to have access to all of the results and not only to a certain selected group. By sharing their work as a preprint, authors can report their findings independent of the direction of the results and of considerations around novelty or impact. The preprint provides a permanent citable record which allows other researchers to build on the findings or to include them in evidence synthesis exercises such as systematic reviews and meta-analyses, helping mitigate the potential negative effects of publication bias.

Preprints as a vehicle for open science

In order to allow collaboration and earlier scrutiny on research work, there have been calls for the sharing of all the different outputs associated with an individual research project: data, code, protocols, and materials. Having these outputs available allows others to scrutinize and give feedback on the work, again at stages earlier to and decoupled from the final journal publication.

Preprints can include links and/or references to the records for the dataset, code, or protocol deposited at dedicated repositories. Preprints are also compatible with Registered Reports, a publication model where the research rationale and methodological plan is reviewed before the study even starts (10), if the plan is deemed rigorous, the study is completed and the results submitted for publication. Some journals publish both the initial plan and the final study, others will only consider the latter for publication. Registered Reports (the initial plan, the final paper, or both) can be posted as a preprint and thus made available independent to or in addition to publication at a journal. This interconnection of outputs can in turn also act as

a driver for further adoption of preprints: a survey of researchers across a range of disciplines reported that having links to data or to pre-analysis plans were among the top factors researchers linked to the credibility of preprints (11).

But there is more to be done to fully materialize the benefits that preprints bring to reproducibility. At present, only a minority of preprint servers require data sharing or data sharing statements as part of the submission process (12). One of the benefits for authors lies in the ease of preprint submission and thus there is a need to balance the burden that data requirements would place on authors with the benefits for openness and reproducibility for research overall. As an intermediate step, preprint servers could consider the implementation of simple statements that designate whether or not data are available and if so where, in the form that OSF Preprints have adopted, as an initial step to facilitate visibility and the evaluation of data-sharing practice in the context of preprints (12).

Preprints are often posted in parallel to the submission of the manuscript to a journal; decoupling these two steps so that preprints are shared earlier (as we have seen during the initial stages of the Covid-19 pandemic) would provide greater avenues for collaboration between groups, and boost the benefits of early review and iteration discussed above. We envisage a future where preprints can be produced and updated at various stages within the research process, allowing synergies between preprints and the sharing of data, code, and other outputs, and the dissemination of different milestones in a research project as it progresses. Closing the now years-long gap between discovery and dissemination will hasten the scientific dialogue that ultimately is needed for a reproducible research ecosystem.

Competing interests

The author is an employee of ASAPbio, a non-profit organization promoting the productive use of preprints.

Acknowledgements

I thank Jessica Polka for helpful suggestions and feedback on the manuscript.

References

1. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2004 Sep 16;351(12):1250–1.
2. Ioannidis JPA. Why Most Published Research Findings Are False. *PLOS Med*. 2005 Aug 30;2(8):e124.
3. Nosek BA, Errington TM. Making sense of replications. *eLife*. 2017 Jan 19;6:e23383.
4. Bring structure to your research [Internet]. protocols.io. [cited 2020 Dec 3]. Available from: <https://www.protocols.io>
5. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. *Science*. 2015 Jun 26;348(6242):1422–5.
6. Library | The EQUATOR Network [Internet]. [cited 2020 Dec 3]. Available from: <https://www.equator-network.org/library/>
7. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. *N Engl J Med*. 2008 Jan 17;358(3):252–60.
8. Therapeutic fashion and publication bias: the case of anti-arrhythmic drugs in heart attack. [Internet]. The James Lind Library. [cited 2020 Dec 3]. Available from: <https://www.jameslindlibrary.org/articles/therapeutic-fashion-and-publication-bias-the-case-of-anti-arrhythmic-drugs-in-heart-attack/>
9. BBC Radio 4 - The Life Scientific, Ottoline Leyser on how plants decide what to do [Internet]. BBC. [cited 2020 Dec 3]. Available from: <https://www.bbc.co.uk/programmes/b08q5wxx>
10. D. Chambers C, Feredoes E, D. Muthukumaraswamy S, J. Etchells P, I Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University; Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neurosci*. 2014;1(1):4–17.
11. Soderberg CK, Errington TM, Nosek BA. Credibility of preprints: an interdisciplinary survey of researchers. *R Soc Open Sci*. 7(10):201520.
12. Beck J, Ferguson CA, Funk K, Hanson B, Harrison M, Ide-Smith M, et al. Building trust in preprints: recommendations for servers and other stakeholders [Internet]. OSF Preprints; 2020 Jul [cited 2020 Nov 6]. Available from: <https://osf.io/8dn4w/>