

---

## PubPeer: Scientific Assessment Without Metrics

Boris Barbour and Brandon M. Stell

“She has a *Nature* paper!” How many times have you heard that sentence? And how many times was it accompanied by a discussion of what was in the paper? Maybe not so often, which is symptomatic of the problems in scientific research today. Worse, the obsession with the journal, at the expense of the result, is a pragmatic adaptation to the incentive structure of research today. Success in the competitions for jobs and grants will follow that publication, independently of its substance (or lack of it).

Publishing in *Nature* or one of several “glamour” journals is supposed to be synonymous with high-quality research of great significance. It represents a metric—a shorthand representation of the quality of research output that is often given numerical form as the “impact factor,” described elsewhere in this volume. The lure of metrics is incredibly strong, because they offer an apparently efficient and objective basis for allocating resources such as grants and jobs, for which competition is fierce.

But every metric distracts our severely limited attention from the substance of the research. And, in doing so, metrics allow career success (or survival) to become dissociated from research quality. Defenders of the most sought-after journals will stress that stiff competition, stringent criteria for general interest, and attentive refereeing ensure that only the highest quality papers are accepted. Yet there are numerous examples of papers of dubious quality and with overblown interpretations being accepted in even the best journals. In many ways this is hardly surprising, because the true scientific impact of a paper is extremely difficult to judge when it is new.

The “top” journals distort science in other, more subtle, but equally pernicious ways. Their criteria for acceptance require papers to present “revolutionary concepts.” Of course revolutions sometimes happen, but the reality is nearly always more prosaic. This means that aspiring authors feel pressured to adapt their presentations to fit these criteria,

quite often exaggerating interpretations and misleading readers. Even worse, our whole approach to research has become governed by the desire to strike gold, even if most risky bets simply don't pay off and authors are obliged to spin their often useless results in some misleading way. All manner of solid, incremental, and useful work is deemed of too low impact to stand a chance of reaching a top journal and is therefore neglected by the grant evaluation committees and hiring committees in charge of identifying and nurturing the most promising research. We thus find ourselves in a situation where all of our research is constrained by the need to create whatever the high-impact journals consider to be high-impact stories. We are caught in a self-perpetuating game where the score is a publication that suits one of the top journals.

It is often stated that "science is self-correcting." If it is, then maybe the occasional excess won't be too damaging? However, as we have argued in our blog post "A crisis of trust,"<sup>1</sup> we believe that a combination of the pressures to publish high-impact research, the acute competition in research careers, and multiple levels of conflicts of interest have created an environment today that is hostile to any form of self-correction.

The traditional form of correction is the publication of a new paper, presenting contradictory findings and explaining how the original study was in error. But such studies are vanishingly rare today. Probably the major discouragement is editorial policy, because although the top journals often seem happy to accept somewhat shaky work raising an interesting possibility, they are almost never keen to publish a paper ruling out that sexy conclusion they previously published. There is a conflict of interest here, because the journal editors obviously prefer to avoid even implicit criticism of their evaluation process and brand.

A further obstacle is financial. Obtaining funding is extremely competitive and most funding bodies seek to identify the "revolutionary" work that will be publishable in the top journals. Proposing to redo somebody's experiments will in most cases be laughed out of the committee. A rare but hopeful counter-example was the large-scale project to replicate one hundred psychology experiments under the auspices of the Open Science Foundation, funded by the Laura and John Arnold Foundation.<sup>2</sup> Famously, fewer than one-half of the experiments replicated. Replications often require human resources as well as financial resources, but if it is neither possible to hire somebody to do the work nor to offer them a realistic chance of publishing in a top journal, it becomes doubly difficult to motivate people to carry out the replication. A final obstacle is social. Replicating work is almost uniformly seen as a sign of mistrust (it often is),

and this can rapidly lead to poor relations with the original researchers and be perceived as aggressive by the community. Replicators risk exposing themselves to reprisals via the many anonymous decisions governing their careers (reviews of papers, grants, job applications, promotions).

An alternative and more lightweight method of correcting science is through correspondence and commenting. Indeed, issues in papers often don't require additional experiments or complete new manuscripts to be identified; such issues could include errors of experimental design, analysis, or interpretation. However, current journal systems for correspondence and for commenting have proved almost totally ineffectual. In the case of correspondence, many journals simply do not have a correspondence section. For those that do, it remains extremely constrained in terms of the number of items a journal will publish and the allowed size of those items. Less apparent, but equally inhibiting, editorial policies repress or dissuade most if not all attempts to publish correspondence. At times, the discouragement and obstacles faced by correspondents are almost comical. Although there should be no space limitation for commenting on journal websites (at least for those that offer the facility), many of the same discouragements operate. In sum, there is a widespread and self-fulfilling perception that journals do not welcome correspondence or comments that criticize their publications. The result is that correspondence and commenting on journal websites make no significant contribution to the correction of science.

Some journals have shown recent signs that they are becoming aware of this problem, but, as set out in our blog "Nature editors: all hat and no cattle,"<sup>3</sup> even thought leaders such as *Nature* are struggling to develop a coherent "correction" strategy within the constraints of their business model and presumptions of excellence. At the time of writing, the journal *Circulation Research* has formulated a promising policy, in which they state that refutations will be considered to have the same impact as the original work they address. This would remove one huge barrier at glamour journals, which almost invariably have judged refutations to fall below their general interest thresholds.

Thus, the present combination of career incentives and journal editorial policies has led to a situation where most researchers actively avoid criticizing directly or indirectly any published work. The self-correction of science is sick, although there are now some signs of progress.

It was in this context that PubPeer was developed and launched almost five years ago. The goal of PubPeer is to offer a web platform (<https://pubpeer.com>) for centralized discussion of publications, in order to create

a community of “post-publication peer reviewers” and offer a new mechanism of scientific assessment. The website is independent of the journals, removing many of the psychological barriers and conflicts of interests inhibiting discussion and criticism of papers. Shortly after our launch, we introduced a key feature—strong user anonymity—and from that point on our comment volume climbed rapidly. It has become clear that a large number of substantive and critical comments had never before found an outlet. Specifically, a very large number of comments (concerning thousands of papers) have been posted that highlight manifest misconduct in the preparation of figure data. Typical issues include fabrication of western blot data through duplication of bands. Through observing these numerous cases, we have come to believe that the suppression of self-corrective mechanisms in science has enabled a surprisingly large number of researchers to build very successful careers based upon the most dubious of research practices.

Anonymity is PubPeer’s most controversial aspect, but we remain convinced that its benefits strongly outweigh its disadvantages. (Our analysis of this is laid out in our blog “Vigilant Scientists.”)<sup>4</sup> Although anonymity apparently introduces the possibility for people to be defamed with impunity, there are in fact layered safeguards to minimize this. First and foremost, our guidelines make clear that comments must be based upon public facts and/or logic, such that other readers can verify each comment’s conclusions for themselves; hearsay is forbidden. These guidelines are enforced by moderation and reporting facilities for all posted comments. Of course, there are additional guidelines regarding politeness in posts, avoiding ad hominem arguments and discouraging speculation. These guidelines, combined with the constraints of the scientific discourse, which is inherently fact based, have, we believe, enabled a very high accuracy of user comments and allowed the site to avoid most of the problems associated with anonymous commenting elsewhere on the internet. In addition, authors are alerted to comments and encouraged to respond to any questions or criticisms by defending and explaining their work. Many of the issues raised on PubPeer would be solved instantly by posting the original data, yet most authors have not chosen this course of action.

Most of the arguments against PubPeer, and more specifically its anonymity, have focused on the belief that it will enable unfounded denigration of researchers. However, the site’s critics are rarely able to produce even a single example of a career that has been unjustly harmed by criticism on PubPeer, despite the large number of comments now in the PubPeer database. For sure there are researchers who commented on PubPeer

whose careers have been harmed, but in all of the high-profile cases that we are aware of, the criticisms have been found to be accurate and justified. In truth, only criticisms that are valid are damaging.

Against this perhaps surprisingly slight “defamatory” risk, we balance the benefit of the strong encouragement of post-publication peer review by anonymity on PubPeer. The kind of information posted to PubPeer can benefit readers in a number of ways. Comments typically highlight research that is unreliable in some way. Knowledge of this can prevent researchers wasting their time and (taxpayers’) money attempting to build upon some apparently exciting but flawed breakthrough. Furthermore, since public policy and medical guidelines are, at least in theory, based upon evidence, often in the form of published research, early warning of potential problems can guard against erroneous policy and even save lives. The guiding policy at PubPeer can be summarized as putting first the readers and users of published research, even if that can sometimes be uncomfortable for the authors of the publications discussed. As we stated on our blog: “...a few ruffled academic feathers pale into insignificance when patients’ lives, taxpayer billions, and young researchers’ careers are at stake. We also suspect that the researchers’ employers—those same taxpayers and patients—would share this point of view.” We also believe that defenders of authors’ rights not to be criticized have forgotten that publication is a choice, freely made by the authors. To put things bluntly, authors who don’t wish their work to be criticized or questioned are always free not to publish.

At the time of writing, PubPeer is approaching five years of operation. It is clearly here to stay and is exercising a growing influence on research, both through a number of spectacular scandals that have been brought to light via the site, but also via a cumulative effect on journals, institutions, and researchers themselves, as the site becomes integrated into research, editorial, and administrative workflows. Although no journals or databases currently integrate PubPeer comments, users can install browser plugins that provide as-you-browse alerts to papers with comments.

With those five years of experience, what has emerged is that most comments are indeed critical. Clearly the most effective motivation to comment is when a reader perceives a problem in a paper. In a way this is unsurprising: firstly, science proceeds by falsification; secondly, most papers are of necessity written to present the work in the most positive light possible. Decline and decay is the fate for many papers.

Post-publication peer review, which PubPeer facilitates, is compatible and synergistic with two other key trends refocusing our attention on the

substance of research. The first is the use of preprints; servers such as the ArXiv and bioRxiv offer instant access to papers as soon as they are ready, eliminating a delay of many months and even years. Arguments about low-quality papers no longer being filtered by peer review can be countered by the extremely successful example of the ArXiv and by the flood of journals without meaningful quality controls. We recall moreover that even the top journals are clearly susceptible to allowing “impact” to compensate for rigor in the papers they publish. PubPeer accepts comments on preprints for both ArXiv and bioRxiv and would implement others if and when they come online (we are however personally in favor of a few, centralized preprint servers).<sup>5</sup>

The second trend is toward full data access. Several publishing groups (including *PLoS* and *NPG* at the time of writing) require authors to share their data upon request or even before publication the full set of their original data. Although the systems and processes of this are by no means reliable or routine, the direction is clear. Access to the data will enable much more meaningful post-publication review and discussion. Authors and journals will no longer be able to suppress checks and reanalyses by controlling (usually refusing) access to the data. One consequence of this will be that many of the kinds of issues that appear on PubPeer and currently lead to somewhat ungracious and inconclusive discussions will in future be resolved directly, by access to the original data that is today withheld. In some respects, we are in a transition period.

What should the future of scientific research look like? What we would like to see and what we are working toward is an environment that is much more accepting of discussion, criticism, reanalysis, and replication. This in turn will focus attention on the substance of publications (whatever their format), with the consequence that authors will be under pressure to publish good quality work. This contrasts with the current situation where authors can largely escape criticism once a “definitive” publication has appeared. Intensive post-publication review also places pressure on authors to provide “after-sales service.” Those that publish low-quality work will have to spend more time explaining and defending their work, which is as it should be.

Will we need journals in the future? Preprint servers could certainly satisfy the requirements of publication at a tiny fraction of the cost. We are quite unconvinced by journal claims of a high added value for their reviewing, formatting, and indexing.<sup>4</sup> Many publishers also try to argue that preprints would produce a flood of low-quality work and that researchers do not have time to trawl through millions of preprints to

find those that are of high quality and of interest to them. Yet it is unlikely that the current system prevents publication of low-quality work (a bad paper eventually gets published in one of the more than ten thousand currently existing journals). Furthermore, physicists all use the ArXiv, which divides papers into broad subjects (an approach replicated by bioRxiv), and don't seem to have any greater difficulty in keeping abreast of new work than researchers in other fields. Once the community grows to a critical mass, post-publication review on sites such as PubPeer could replace (or complement) some if not all of the evaluation that is currently carried out with pre-publication review, although the volume of the latter obviously greatly exceeds post-publication review at this time.

Finally, we return to the issue of metrics, be they explicit (e.g., number of citations) or implicit (the journal). As we have argued above, we believe that all metrics are dangerous attempts to measure the unmeasurable, encourage gaming of the metric, and distract attention from the substance of the science. We consider it unlikely that any metric will prove beneficial to the management of science.

## Notes

1. The PubPeer Foundation. 2015. "A crisis of trust." Accessed July 6, 2016. <https://blog.pubpeer.com/publications/EE46E14F6AA97049928835DBC6B908>.
2. Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science." *Science* August 28, 2015: Vol. 349, Issue 6251, doi:10.1126/science.aac4716.
3. The PubPeer Foundation. 2016. "Nature editors: All hat and no cattle." <https://blog.pubpeer.com/publications/AE11BE44CF3C40A558F3B453BF53C7>.
4. The PubPeer Foundation. 2016. "Vigilant Scientists." Accessed July 6, 2016. <https://pubpeer.com/publications/B6CF3DB974A8ECC64B1A0303BBCD6F>.
5. Cell Press. 2016. "Let's talk about preprint servers." Accessed July 6, 2016. <http://crosstalk.cell.com/blog/lets-talk-about-preprint-servers>.

