# Accelerating scientific publication in biology

**Ronald D. Vale**[a,b,1]

[a]*Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158; and* [b]*Howard Hughes Medical Institute, University of California, San Francisco, CA 94158*

Edited by Allan C. Spradling, Carnegie Institution for Science, Baltimore, MD, and approved September 28, 2015 (received for review July 10, 2015)

**Scientific publications enable results and ideas to be transmitted throughout the scientific community. The number and type of journal publications also have become the primary criteria used in evaluating career advancement. Our analysis suggests that publication practices have changed considerably in the life sciences over the past 30 years. More experimental data are now required for publication, and the average time required for graduate students to publish their first paper has increased and is approaching the desirable duration of PhD training. Because publication is generally a requirement for career progression, schemes to reduce the time of graduate student and postdoctoral training may be difficult to implement without also considering new mechanisms for accelerating communication of their work. The increasing time to publication also delays potential catalytic effects that ensue when many scientists have access to new information. The time has come for life scientists, funding agencies, and publishers to discuss how to communicate new findings in a way that best serves the interests of the public and the scientific community.**

scientific publication | arXiv | PhD training | career advancement | journals

Most biologists have become frustrated with the current state of scientific publishing. Attention has been drawn to flaws in using journal impact factors for evaluating scientific merit (1), the hypercompetitive environment created by scientists seeking to publish their work in the top journals (2), and the extensive revisions required by reviewers and editors (3, 4). In this Perspective, I wish to focus on another issue that has received less attention—the increasing amount of data and time required to publish a paper.

As a consumer of scientific literature, I enjoy reading the comprehensive scientific studies that are being published today. However, the foundation of today's data-rich articles is acquired at a cost, which is the time that graduate students and postdoctoral fellows (postdocs) spend in collecting and analyzing data. Indeed, as I will discuss later, the length of time required to produce and then publish a scientific work is likely impacting the duration and quality of PhD and postdoctoral training. Furthermore, as laboratories wait to accumulate more experimental data before they feel that a benchmark for publication is met, crucial results are being sequestered from the scientific community for longer periods of time. In this Perspective, I will argue that creating new outlets for faster and more nimble scientific communication could have positive outcomes on professional training, catalyzing scientific progress, and improving the culture of communication within the life sciences as a whole.

## A Trend Toward Increasing Data Required for Publication

Many senior scientists feel that the amount of data required for publication has increased over their careers (for example, see ref. 4). But is there evidence supporting this claim? Quantifying the amount of experimental data in a publication is nontrivial because data can take many different forms and vary in the amount of time required for acquisition. Furthermore, comparing the amount of data in contemporary versus prior papers is difficult. For example, the time required to obtain certain types of information has decreased; as an extreme example, sequencing an entire genome now requires less time than cloning and sequencing a single gene 40 y ago. However, scientists always push technical limits, and many of the experiments performed today also are difficult and require a long time to master and execute. Thus, I would argue that truly informative experimental data are not vastly easier to obtain now than in the past. Practices in data inclusion, however, may have changed; for example, experiments previously described as "data not shown" would now likely be included in a supplemental figure. Figures also are easier to prepare now with computer programs compared with more cumbersome manual methods in the past.

With the above caveats noted, I sought to compare the amount of experimental information presented in biology papers published in *Cell*, *Nature* (biology only), and *The Journal of Cell Biology* (JCB, operated by editors from the scientific community) from the first 6 mo of 1984 and of 2014. The number of papers published by *Cell* remained approximately the same, decreased slightly for *Nature*, and dropped in half for JCB in 2014 compared with 1984 (Fig. 1*A*). The average number of figures in the print version of papers did not change significantly (Fig. 1*B*), because journal guidelines have remained largely the same between these two time periods. However, during this 30-y span, the number of experimental panels contained within the print version of the paper rose dramatically by two- to four-fold (Fig. 1*C*) (see Fig. S1 for the breakdown of short and long format papers in *Nature* and JCB). Separate labeled panels do not always constitute distinct experiments, and figure labeling styles might have changed in the past 30 y. To examine this point, panels in *Cell* and *Nature* were scored as to whether they contain distinct pieces of data or were derived from the same experiment (see *SI Methods* and Fig. S2). The number of distinct datasets was approximately two-thirds of the number of labeled panels, and this ratio did not change substantially between 1984 and 2014 for either *Cell* or *Nature*. Thus, the fold-increase in panel number seems to reflect a true increase in the amount of data in the print version between 1984 and 2014. The

Author contributions: R.D.V. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]Email: vale@ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1511912112/-/DCSupplemental.

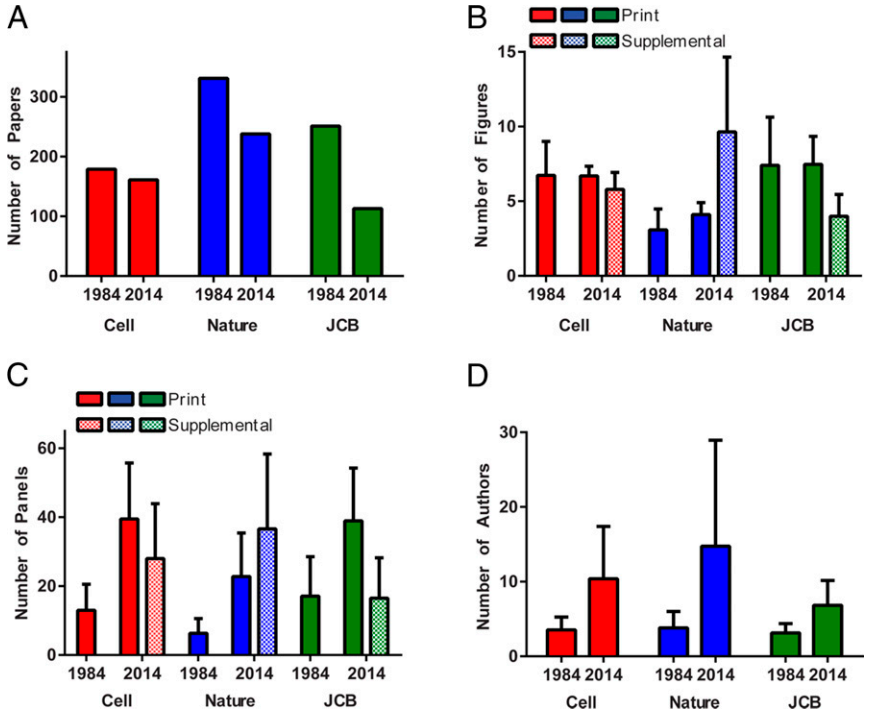


**Fig. 1.** Statistics for papers published in *Cell*, *Nature* (biology papers only), and *The Journal of Cell Biology* (JCB) for the months of January–June in 1984 and 2014. Long and short format papers (articles and letters for *Nature*, and articles and reports/rapid communications for JCB) are grouped together in this figure, but analysis of each category can be found in Fig. S1. (*A*) The total number of papers published during these two 6-mo time periods. (*B*) The average number of figures in the print and online supplement of each paper. For *Nature*, most of the data in this figure are derived from the "Extended Data" section although the "Supplemental Information" section also contributes some data in this analysis. An online supplement did not exist for journals in 1984. (*C*) The number of panels per paper (assigned as a letter in the figure; tables were also scored in this category). (*D*) The average number of authors per paper. The means and SDs are shown in *B*–*D*. See *SI Methods* for details on analysis. See Fig. S2 for an analysis of the pieces of distinct experimental data contained within the panels of the print versions of *Cell* and *Nature*.

increase in the amount of data per paper is even more substantial when supplemental information, which began to appear around 1997, is taken into consideration (Fig. 1 *B* and *C*). In particular, the number of supplemental figures and their panels was comparable with (*Cell*) or exceeded (*Nature*) those that were published in the print version (Fig. 1*C*). Consistent with this trend of more data and the likely use of more diverse and complex techniques, today's papers in *Cell*, *Nature*, and JCB have two- to fourfold more authors than those from 1984 (Fig. 1*D*). However, enlisting more authors is probably not the sole mechanism for acquiring the additional data needed for contemporary papers. As will be discussed later, it also seems to take a longer period to publish a paper now than in the past.

## Factors Driving an Increasing Amount of Data per Publication

What factors have driven the increasing amount of data per publication over the past few decades? One likely factor is supply and demand—more scientists are competing for the same or less real estate (space in top journals, Fig. 1*A*) compared with 30 y ago. Over the past 30 y, the US scientific workforce (e.g., postdoctoral fellows and graduate students) has increased by almost threefold (5, 6), fueled, in part, by the doubling of the NIH budget between 1998 and 2003. In addition to the United States, many other countries recently have expanded their life science research programs. From 1999 to 2005, publications from US laboratories increased only 3.6% annually whereas those from China increased 38.9% (7). Thus, with more scientists desiring high-profile publications for their grants and promotions, the elite journals can set a higher bar for what they accept. A "high impact" result constitutes one important criterion for publication. However, a second and increasingly important benchmark is having a very well-developed or "mature" research story, which effectively translates into more experiments and more data. A whole genome screen followed by a mouse model to understand the physiological functions of one of the gene hits, as well as additional structural work to understand the mechanism, might be what is needed to seal the deal for acceptance. Reviewers, in turn, fall in line with the escalating expectations and continually reset their own benchmarks of "what it takes" to get into a particular journal. With these market forces at work and a positive feedback loop between journal editors and reviewers, the expectations for publication have ratcheted up insidiously over the past few decades.

In addition to the time required to obtain the data for submission, the review process itself typically adds new demands for more data before the work can be officially accepted for publication. If one is fortunate enough to have the paper sent out for review, then three referee reports are commonplace these days. Frequently, each referee requests additional experiments. Many of our own papers have been significantly improved by experiments suggested through peer review. However, many suggested experiments are unnecessary, and sometimes the requested work is so extensive that it constitutes a separate study unto itself. Furthermore, it is not easy to "say no" to referee-suggested experiments or a journal request to curtail the discussion. After all, the journal editor will have another revised paper on his/her desk where all of the referees are completely satisfied. Thus, authors feel as though they are held hostage, fearful that their paper will not be accepted if they do not comply with most, if not all, of the requests.

Although the elite journals are important driving forces in the scientific market place, the trend toward more data is felt throughout the publication ecosystem. One reason is that nonelite journals want to improve their status and, as a consequence, strive to be selective and seek more mature stories. This factor may account for why JCB accepts fewer papers now than it did in the 1980s (Fig. 1*A*). Second, scientists feel pressured to aim high and acquire the data that they think will be needed for publication in an elite journal. But, alas, when it comes time for journal courtship, they find their work editorially rejected not once, but thrice, and then eventually publish their large body of work in a lower tier journal. It is not easy to obtain information on journal rejections from the 1980s although I speculate that the frequency has increased considerably in the past 30 y. Thus, in addition to the time invested in acquiring data, the time spent in finding a home for a paper through sequential journal submissions also significantly delays the transmission of results to the scientific community.

## What Is a Minimal Unit for Publication?

Most scientific papers, now and in the past, usually have one or two key findings. But, with the trend toward publishing more mature scientific stories, it has become harder to publish just a key initial finding or a bold hypothesis.

Let's consider the Watson and Crick publications, perhaps the most famous in modern biology, and imagine how they might fare in today's publishing environment. Many people may be unaware that Watson and Crick published not one paper, but two papers, on DNA in *Nature* in successive months. The first paper, published on April 25, 1953, described a structural model for the DNA double helix (8). Despite having a single figure (a model figure without data), it was listed as an "Article" rather than a "Letter," based upon the magnitude of the idea. The first Watson/Crick paper was accompanied by two other articles on the X-ray diffraction pattern of DNA; the paper by Maurice Wilkins et al. had two figures (9), and the one by Rosalind Franklin and R. G. Gosling displayed a single figure (10). The second Watson/Crick *Nature* paper (also an article published on May 30) was entitled "Genetic Implications of the Structure of Deoxyribonucleic Acid." It described, without any data, a hypothesis for the hydrogen bonding of the "Watson–Crick" base pairs and speculated how the two DNA strands might each provide a template for the replication of genetic information (11). Several months later, Wilkins et al. and Franklin et al. each independently published second *Nature* articles describing more complete analyses of the structure of DNA (12, 13). Thus, the story of DNA, like a Charles Dickens novel, came out in installments. Furthermore, it also should be emphasized that the Watson and Crick model was speculative, particularly with regard to the process of DNA replication. As a result, the revolutionary ideas of Watson and Crick were not instantly accepted, and their implications were not widely understood by the scientific community at the time of publication. Experimental evidence for the unwinding of the DNA strands and semiconservative replication was published in 1958 by Meselson and Stahl (14), and this result placed the Watson and Crick model for replication on a solid footing.

Somewhat tongue-in-cheek, let us imagine a contemporary editorial decision on the 1953 Watson and Crick papers (in reality, these papers were not peer reviewed; see *Nature*'s recollection of the publication process in ref. 15):

*Dear Jim and Francis*:

*Your two papers have now been seen by three referees. Based upon these reviews, I regret to say that we cannot offer publication at this time. Although your model is very appealing, referee 3 finds that it is somewhat speculative and premature for publication. Indeed, your model proposing a semi-conservative replication of DNA raises many obvious questions. As two of the referees point out, it should be possible to determine experimentally whether the two strands can separate and serve as templates. This would address referee 3's concern that strand separation is not feasible thermodynamically. I regret to say that, without such experimental evidence, we will not be able to publish your work in Nature and suggest publication in a more specialized journal. Should you be able to furnish more direct experimental evidence, we would be willing to reconsider such a revised paper. Naturally, we would need to consult our referees once again. Furthermore, because space in our journal is at a premium, if you do decide to resubmit, then we recommend that you combine your two submitted papers into a single and more cohesive article, potentially including the X-ray studies of your colleagues at Cambridge. Thank you again for submitting your papers to Nature. I am sure that this revision will delay your Nobel Prize and the discovery of the genetic code by only one or two years.*

A discovery emerging in closely spaced installments was not unique to DNA. The molecular mechanism underlying familial hypercholesterolemia was unraveled in three key papers by Brown, Goldstein, and coworker between 1973 and 1974, each of which solved a piece of the puzzle (16–18). Similarly, the discoveries of ubiquitination and protein degradation by Hershko, Ciechanover, Rose, and coworkers emerged in three papers in 1979 and 1980 (19–21). Studies on the mechanism of axonal transport by me, Schnapp, Reese, and Sheetz (covering work from 1983 to 1985) were published in five papers in 1985 (22–26). In all of the above examples, the information could have been delayed and compacted into fewer publications, as no doubt would occur today. However, by unfolding these breakthroughs in a series of papers, the progression of results could be quickly disseminated to the scientific community, the value of which will be discussed in the next section.

Today, two opposing factors come into play in deciding when to publish a paper. On one hand, scientists want to get their work published as fast as possible, both for advancing their careers and for claiming priority for their discovery and avoiding getting "scooped." However, publishing in a top journal has become an equally compelling consideration for many scientists, and this latter factor can tip the balance toward delaying submission until more experimental data can be obtained.

## Consequences on the Exchange of Information Within the Scientific Community

The "comprehensive" paper enables authors to build a convincing argument for their hypothesis. Indeed, the Watson/Crick model combined with the Meselson/Stahl experiment would have constituted an amazing paper that would have immediately convinced everyone in the field. However, there is also merit in getting new ideas and key experiments published with reasonable speed, even if they are incomplete. Once in the public domain, the collective power of the scientific enterprise can take effect, and the ideas can be tested and advanced further, not only by the original researchers but also by other investigators as well. Once results are published, other scientists can see connections with their own work, perform new experiments that the original investigators might never do, and also emerge with new ideas. Overall, putting new results and ideas in the public domain is good for science and serves the mission of the funding agencies that seek to advance research overall.

The protracted and uncertain nature of the publication process also may be affecting the exchange of information at scientific meetings. Students and postdocs, although eager to have the chance to present their work, have become increasingly wary about sharing their unpublished data at scientific meetings. As a result, scientific meetings are becoming increasingly filled with recently published or soon-to-be published results, rather than exciting work in progress.

## Consequences for Training

In 1990, the average age at which scientists received their first R01 NIH grant was less than 38 y of age; in 2013, that same milestone was reached at an average age of over 45 y (27). This trend is of great concern for many obvious reasons (2, 27), including the fact that it is making a career in biomedical research less attractive to young people (28). In an attempt to reverse this trend, efforts are now being made to accelerate the career track of young scientists. Many graduate schools require regular thesis committee meetings to promote timely graduation, and a recent Perspective in PNAS suggests limiting funds for graduate training to 5 y (29). Some institutions and granting agencies limit the length of postdoctoral training to 5 y, which is also strongly recommended by a recent National Research Council report (30) and others (29). In addition, new grant schemes, such as the

NIH K99, seek to promote the transition of postdoctoral fellows to junior faculty positions. All of these measures are worthy, but, for them to succeed in reducing training time, they must be accompanied by changes in the publication system. Placing term limits on graduate and postdoc training would be a perfect solution if principal investigators (PIs) were always responsible for keeping their trainees for too long in their laboratories. Although this situation no doubt occurs, graduate students and postdocs also are asking their PIs whether they can stay for a longer period. To understand why, one has to appreciate the connection between publication and career advancement.

Scientific papers are required for obtaining a job, a promotion, or a grant, and thus have become a primary currency for professional advancement. Furthermore, papers in elite journals have become particularly valuable in the career marketplace. Graduate students and postdocs understand the "paper economy," and they want to publish as many papers as possible and ideally publish a paper in *Cell*, *Science*, or *Nature*.

But it seems as though publishing many papers and being published in elite journals is harder now than it was in the past. I examined the publication records for PhD students at the University of California, San Francisco (UCSF) who graduated in the 1980s ($n = 71$) versus those that graduated in the past three years ($n = 104$) (Table 1 and Figs. S3 and S4). The average time for acquiring a PhD increased slightly between the past (5.7 y) and current (6.3 y) student groups; these times to degree are largely consistent with national trends (5, 29). However, even though the contemporary group of graduate students was in school for one-half year longer, they published fewer first/second author papers and published much less frequently in the three most prestigious journals. Consistent with the notion of more data being required for publication, the contemporary students also took an additional 1.3 y, on average, to publish their first first-author paper compared with students from the 1980s. Strikingly, the average time to a first-author publication for the current cohort (6 y for students who publish) is just below the average time of their graduation (6.3 y) and at the desired upper boundary for training in these graduate programs (6 y or less). These general trends also are apparent when comparing the top one-third of students with the best publication records, suggesting that the differences cannot be explained by admitting a pool of less capable students now than in the past (Table 1). UCSF also remains a highly sought-after graduate school, and its reputation has gotten stronger since the 1980s. This type of analysis should be extended to larger numbers of students from many different universities, but these preliminary data suggest that it has become harder for graduate students to publish.

The increasing time to publication poses difficulties in reaching milestones for career advancement. Graduate students often need to apply for a postdoctoral position 9–12 mo before graduation, and thesis committees frequently recommend having a first-author paper accepted for publication before initiating the application process. Postdocs seeking a job or grant support face a similar predicament. For example, let us consider the timing of the highly sought-after NIH K99 Pathway to Independence Award, which provides 1–2 y of postdoctoral training and 3 y of independent support. The postdoc likely requires 2 mo to write a successful grant, and then it can take 9 mo from submission to the time when funding is received. Importantly, a K99 grant will be considered much more competitive if the postdoc has a prior publication; a "manuscript in submission" cannot be listed in an NIH grant application. If it takes a postdoc 3 y to have a paper accepted before submitting a competitive K99 application (often a best case scenario), then a talented young scientist will spend ∼5–6 y in a postdoc before getting a job (three years to publish a paper and an additional year from grant writing to funding, followed by an ∼1- to 2-y training period). In summary, the ability of thesis, grant, and job committees to access a formal and publicly accessible paper could accelerate career transitions toward the end of graduate and postdoctoral training.

Providing young scientists with more opportunities to publish also has other advantages for training. Preparing and publishing a scientific paper is a critical part of the apprenticeship of becoming a scientist. This experience promotes skills not only in writing, but also in organizing experimental data and learning how to convey ideas effectively. The process of completing a scientific paper also teaches young scientists how to be more efficient in planning and executing experiments in their future projects. However, with the increasing time involved in acquiring data and publishing, young scientists get fewer chances to write papers and thus arguably are less well-trained in these skills than trainees in the past (Table 1). Furthermore, if a critical study reaches the point of publication after 4–5 y of work, all too often the PI, who has more experience, takes over the process of writing from a graduate student or postdoc. In such cases, neither the young scientist nor the PI is willing to take chances with the paper being accepted in today's competitive publication environment.

Another value of publishing earlier is that it allows a graduate student or a postdoc to explore more options for using the remaining training period. Rather than myopically focusing on getting their one paper accepted, trainees can decide whether they want to expand their first study, move

**Table 1. Scientific journal publications from UCSF graduate students**

| Graduation year | No. of students | Graduation time, y | Time to first-author paper, y | No. of first-author publications | First- plus second-author publications | First author C/N/S | First plus second author C/N/S |
|---|---|---|---|---|---|---|---|
| 1979–89 | 71 | 5.7 ± 1.0 | 4.7 ± 2.3 | 2.2 ± 1.5 | 2.9 ± 1.8 | 0.52 | 0.80 |
| Top 1979–1989 | 24 | 5.2 ± 0.9 | 3.4 ± 1.1 | 3.1 ± 1.2 | 4.5 ± 1.7 | 1.25 | 1.63 |
| 2012–2014 | 104 | 6.3 ± 0.9 | 6.0 ± 1.9 | 1.4 ± 0.9 | 2.1 ± 1.3 | 0.17 | 0.31 |
| Top 2012–2014 | 34 | 5.9 ± 0.7 | 4.7 ± 1.4 | 2.4 ± 0.8 | 3.5 ± 1.1 | 0.53 | 0.94 |

The publications from PhD students who performed experimental work and graduated in the indicated years of the Biochemistry and Molecular Biology, Biophysics, Genetics, and Neuroscience programs were analyzed. The time periods indicated refer to the year of graduation. A larger time span (1979–1989) was scored compared with the recent time period (2012-2014) because past graduate programs were smaller than they are now. "Top" refers to the top one-third of the students in each group with the best publication records, as assigned qualitatively based upon the combination of criteria described in this table. "C/N/S" refers to papers in *Cell*, *Nature*, and *Science* and represents the average number of publications in these journals per student. Values represent means and SDs. Because coauthorship did not exist in the 1980s, we scored only the order of authorship; thus, a shared first author in the second position was counted as a second authorship in our analysis; an exception to this rule was made if a second position, cofirst author work, was the sole paper from the student's graduate work. For more details of the analysis, see *SI Methods*. Scatter plots for all of the data are shown in Figs. S3 and S4.

on to another research question, or spend some time pursuing additional career training (e.g., teaching).

## Possible Solutions for Accelerating Communication

New journals and publishing platforms have recently introduced several interesting innovations, including providing immediate open access to publications (which *PLOS ONE* is doing on a large scale) and reforming the process and transparency of peer review (e.g., *eLife* and *F1000Research*). The above efforts should be applauded. However, creating more new journals, which are expensive to operate and must struggle to compete for good manuscripts, is unlikely to constitute the transformative solution needed for accelerating scientific communication. A mechanism that has the potential for transformative change must (*i*) operate on a large scale (i.e., hundreds of thousands of papers per year rather than hundreds), (*ii*) succeed in capturing the very best work in the field, (*iii*) be able to launch and coexist with existing journals, and (*iv*) be cost-effective and be possible to implement on a time scale of years rather than decades.

**Lessons from the Physics Community: Should Biologists Adopt an Internet Preprint System?** A mechanism for accelerating scientific communication that meets the above criteria has been developed already by the physical science community. Physicists, mathematicians, and computer scientists typically deposit their scientific manuscripts before journal publication in an open access e-print service called arXiv (pronounced "archive"), which was founded by Paul Ginsparg and is now operated by the Cornell Library. At first created for the high energy physics community, arXiv use has spread over time to other sectors of physics, mathematics, computer science, and quantitative biology. This repository of electronic preprints is searchable, and many physicists have developed a habit of checking for alerts from arXiv first thing in the morning. Generally, although not always, a paper uploaded onto arXiv is then submitted to a journal. Importantly, the public disclosure through arXiv is accepted by the physical science/mathematics community as a priority for a discovery, and an arXiv posting is acceptable as a reference in a journal, book, or grant application. After the original paper is posted in arXiv, new versions can be uploaded: for example, after a paper has been revised through the journal review process or in response to other comments received by the community. However, earlier versions of the paper are retained, and the nature of the changes is indicated in revised uploads.

ArXiv evolved from a common practice in the physics community, beginning several decades ago, of mailing unpublished manuscripts to colleagues in the field. This practice also was more common in the early years of molecular biology, a famous example being Watson and Crick obtaining a preprint from Linus Pauling that proposed the erroneous triple helix model of DNA. As technology evolved, mail turned to email, and physicists sent their manuscripts to colleagues by this electronic route. With the development of the internet, physicists rallied around the formation of a preprint server, and arXiv was established in 1991. From its inception through January 2015, one million papers have been submitted to arXiv. In 2013 alone, arXiv papers were downloaded 67 million times. Differing from the bulk of work in biology, arXiv contains many purely theoretical papers. However, landmark experimental studies also are routinely disseminated first on arXiv, a recent example being the discovery of the Higgs boson.

Would a centralized, open access, and widely used preprint repository be sensible for biologists, as it has been for physicists? Harold Varmus advocated for such a system (termed E-biomed) in 1999 when he was director of the NIH (31), and others have more recently echoed benefits (32). Currently, there are a few preprint servers specifically for biology, including bioRxiv.org (launched in 2013 by the nonprofit Cold Spring Harbor Press) as well as *PeerJ* and *F1000Research*, for-profit companies that also offer platforms for peer review. However, preprints in biology have not achieved a critical mass for takeoff. Last year, for example, bioRxiv received 888 preprints compared with 97,517 for arXiv, even though many more papers are published in the life sciences.

Having never used a preprint server myself, I tried the experiment of submitting this Perspective to bioRxiv and PNAS on the same day (July 10, 2015); after initial screening, the article was posted as a PDF on bioRxiv on July 11 (33). Fig. 2 shows the number of views of the bioRxiv article and social media exchanges ("tweets") from the time of preprint posting until the receipt of two peer reviews and an editorial decision from PNAS (August 21). The data show that the preprint reached a large audience (views of the abstract were over twice that of the whole article) and also reveal how social media can drive viewership. Importantly, even before the receipt of two anonymous referee reports, I received extensive feedback on the article through comments posted on bioRxiv, direct emails from readers, and numerous personal discussions. Such feedback helped me to formulate a set of the pros, cons, and uncertainties surrounding preprints, as discussed below (for a more extensive discussion of these issues, see *SI Q&A Regarding Preprints*).

### The Pros: Fast, Free, and Feasible.

*i*) Submission to a preprint repository allows a paper to be seen and evaluated by colleagues and search/grant committees immediately after its completion. This open availability of the study could enable trainees to apply for postdoctoral positions, grants, or jobs earlier than waiting for the final journal publication. It also allows independent investigators to transmit their latest work in a reliable

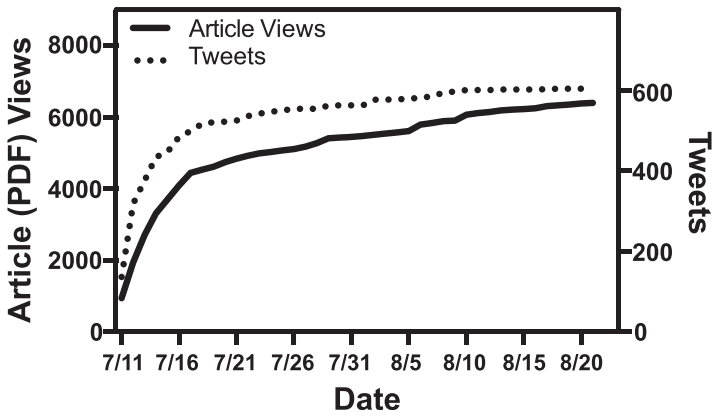


**Fig. 2.** Cumulative article (PDF) views and Tweets for the original version of this Perspective after its posting on bioRxiv (33). The data show the viewership and social media exchanges from the time of its posting (July 11, 2015) until the time when two peer reviews and a favorable editorial decision were transmitted to the author by PNAS (August, 21, 2015). Abstract views were more than twice the number of the PDF views. Data were provided by bioRxiv.

manner to grant review committees, without an unknown delay imposed by the journal publication process. A recent study of several journals found an average delay of ∼7 mo from acceptance to publication (34), but some journals take longer (34) and this time does not take into account journal rejections and the increasingly prevalent need to "shop" for a journal that will publish the work.

*ii*) A primary objective of a preprint repository is to transmit scientific results more rapidly to the scientific community, which should appeal to funding agencies whose main objective is to catalyze new discoveries overall. Furthermore, authors can receive faster and broader feedback on their work than occurs through peer review, as I have discussed as a case in point with this article (Fig. 2; also see an experience from a junior faculty member in *SI Q&A Regarding Preprints*).

*iii*) If widely adopted, a preprint repository (which acts as an umbrella to collect all scientific work and is not associated with any specific journal) could have the welcoming effect of having colleagues read and evaluate scientific work before it has been branded with a journal name. For grants, jobs, and awards, physicists will read and evaluate science posted on arXiv. The life science community needs to return to a culture of evaluating scientific merit from reading manuscripts, rather than basing judgment on where papers are published.

*iv*) A preprint repository is good value in terms of impact and information transferred per dollar spent. Compared with operating a journal, the cost of running arXiv is low, with most of its operating costs covered from modest subscription payments from 175 institutions and a matching grant from the Simons Foundation. Unlike a journal, submissions to arXiv (and currently bioRxiv) are free.

*v*) Future innovations and experiments in peer-to-peer commentary and evaluation could be built around an open preprint server. Indeed, such communications might provide additional information and thus aid journal-based peer review.

*vi*) A preprint server for biology represents a feasible action item because the physicists/mathematicians have proof-of-principle that this system works and arXiv has coexisted with journals, with each providing different services in science communication (*SI Q&A Regarding Preprints*).

**The Cons: Lack of Peer Review and Information Overload.**

*i*) The lack of peer review might invite lower quality or irreproducible data to be disseminated. Although a risk (*SI Q&A Regarding Preprints*), several factors mitigate such concerns. First, arXiv and bioRxiv each have an initial screening mechanism that helps to eliminate overtly "unscientific" articles. Second, the major factor for ensuring quality is that the reputation of the investigator is at stake, and achieving a good reputation within the community is a primary motivating factor for scientists. Indeed, a preprint submission is immediately visible to the entire community whereas a journal submission is seen confidentially by only a couple of referees. Thus, posting of a poor quality paper on a preprint server will be widely visible and reflect poorly on the investigator and his/her laboratory. Scientists take pride in their work and will be guided by their own internal standards in deciding when their work is ready to be released to the community. Third, the paper can receive input (as this article has) from more than two or three referees, which could help authors correct flawed experiments/statements and help produce a better final product published in the journal. Fourth, peer review by journals, although helpful, is certainly not a fool-proof mechanism for identifying problems or eliminating scientific irreproducibility, especially because the referees' first task is to assess whether the work is "exciting enough" rather than "accurate enough." If a recent fictitious method for preparing pluripotent stem cells (35) had first surfaced as a preprint, many scientists would have likely noted its flaws well before journal publication. Thus, the buyer always must beware and exercise appropriate judgment for scientific quality, regardless of whether a study appears in an elite journal or an electronic preprint server. In addition, one could imagine an option of incorporating author-initiated peer evaluations as part of a preprint, which most scientists do informally before submitting their work to a journal and is not unlike the mechanism by which National Academy of Science members submit papers to PNAS.

*ii*) Preprints could expand the problem of information overload in biology by opening the door to less interesting reports that are not being published by journals. Although this consequence could ensue, certain "unpublishable" studies, such as a negative result or whether a prior finding can be reproduced, might provide useful information to some scientists. Furthermore, scientists are already living in a world of information overload. Rather than suppressing preprints, the answer may lie in better search filters, such as key words, colleagues of interest, social media cues, and potentially even other measures of validation [such as whether the work was supported by a grant from NIH, the National Science Foundation (NSF), or other major agencies].

**Uncertainties: Culture, Priority, and Government and Journal Support.** If the pros seem attractive and the cons manageable, then why are preprints not being used by biologists? One reason is that most biologists simply don't know about preprint servers. But there are other reasons as well. Many believe that biology has a different culture from physics, which will make it impossible for the success of arXiv to be extended into biology. "Culture" refers to the moral fabric of the community—how credit for a discovery is assigned, how information is shared, and how a scientist's work is evaluated. Currently, many issues regarding preprints, which are clear for physicists, are clouded by uncertainty in the biology community (*SI Q&A Regarding Preprints*). In the fast moving world of experimental biology, will a preprint publication result in an increased risk of losing credit and getting scooped? Will a preprint put a journal submission at risk for automatic rejection? Will a preprint be recognized by grant agencies, thesis committees, etc.? These uncertainties create considerable barriers to use of preprints in the biology community. The following leadership and policy changes could eliminate these barriers:

*i*) Preprints become accepted as evidence for establishing priority of a discovery, as is true in physics.

*ii*) Preprints become accepted as evidence of productivity in grant applications. Currently, NIH allows listing only of accepted peer-reviewed papers in a grant. However, grant reviewers are "peer reviewers" and should be able to judge the quality of a scientist's most recent work in the form of a preprint.

*iii*) Preprints become accepted by life science journals. Currently, many journals (*Science*, *Nature*, *eLIFE*, PNAS, and others) allow prior preprint submissions; however, some

journals still have ambiguous policies, which constitutes an overall deterrent.

**Help from the Journals: Creating a New "Key Finding" Format.** A preprint server provides a solution for improving the ease and speed of communicating a paper, but it does not necessarily address the escalating amount of data needed for publications in journals (Fig. 1). Here, journals themselves could take the lead. Many journals now have "short" communications (e.g., *Nature Letters*, *Science Reports*, *The Journal of Cell Biology Reports*, *Current Biology Reports*). However, their guidelines have primarily curtailed the number of words, because researchers have found creative ways of stuffing more and more data into the allowable number of figures and supplemental online material (noting the obvious element of irony, please see Fig. S1 for the amount of data included in *Nature Letters* and JCB *Reports*). It is worthwhile considering introducing a new journal format whose focus is on limiting data more than text. One could imagine a format limited to eight panels arranged in up to four figures and with no supplemental data. One of the figures could be identified as the "Key Finding," with a text box describing why it contains the cornerstone result of the paper. Is it possible to convey good science in such a restricted format? It was possible 30 or more years ago (this idea is effectively the *Nature Letter* or *Science Report* of the past) so it should be now. Creating a new format has the potential of permeating throughout the publishing world, like cover art, commentaries, etc., provided that it is popular among authors and readers.

## Conclusions

We may be approaching a breaking point in the publication process in the life sciences. The analysis of graduate students presented here suggests that the average time to first-author publication has ratcheted upwards and is now approaching the length of PhD training. Furthermore, the strong desire of investigators and their trainees to publish in high profile journals, the requirements of US graduate programs (implicit or explicit) for PhD candidates to publish a first-author paper, the inability to include not-yet-accepted manuscripts in grant applications, and the hopes of federal agencies to shorten PhD/postdoc training are all coming into conflict with the ground realities of the present day scientific communication system. In addition to scientific training, important elements of scientific culture also stand to gain from improving the practices and timing of publication, including better evaluation practices for promotion and regaining an open atmosphere of communicating unpublished results at scientific meetings.

Changing the status quo seems daunting if not impossible, particularly to many young scientists who feel frustrated by the present publication system. It is easy to assign the fault to the journals, but such blame is misplaced and diverts attention from where the lion's share of the responsibility lies—in our own life sciences community. As scientists, we need to define our culture and take ownership in developing a system for communicating research results that best suits our needs as well as the needs of the public. We have not done so, at least not yet. Optimistically, change can happen if our community sets its mind to the task, recognizing that universal consensus may not be achievable and that certain subfields of biology will likely embrace new ideas more readily than others. Young scientists, who have grown up in a culture of sharing information on the internet, also may embrace a new opportunity if it is presented to them.

As is often the case, it is easier to articulate the problem than derive an effective solution. One idea discussed here for accelerating publication in the life sciences is the widespread adoption of electronic preprints. Mechanisms for submitting preprints already exist; however, with everyone standing at the shore and very few people willing to jump in, the water looks cold and uninviting. Thus, a challenge for this idea becomes changing behavior on a massive scale, which first requires removing barriers and providing better incentives for preprint publishing; only then can the experiment be done properly of establishing whether preprints serve the needs of biologists. Others may feel that reform of the existing journal system (better and more transparent reviewing and better evaluation metrics) might suffice without resorting to a preprint server or other new model. But will these reforms work without implementing new incentives for currently overwhelmed scientific referees and will they be sufficient to truly change the "daily lives" of graduate students and postdoctoral fellows? Others feel that journals and preprints are both arcane and that developing an entirely new system is needed. To discuss and debate these issues, it may be an opportune time to hold a meeting of major stakeholders (junior and senior scientists, funding agencies, scientific societies, philanthropists, and journal editors) specifically to discuss the issue of how to accelerate the communication of scientific results in biology. The most important stakeholder in this discussion is the National Institutes of Health, which has already greatly influenced publication practices by requiring its grantees to abide by public access policies. Because NIH is deeply interested in (*i*) promoting public good by catalyzing research discoveries, a process that is facilitated by rapid access to scientific results, and (*ii*) advancing the career paths of its trainees, the topic of accelerating scientific communication should be of great interest to them. Indeed, everyone will likely step into the water together with new prepublications and/or publication practices if NIH determines that it serves the greater good of the scientific community and the nation's research agenda. Through thoughtful discussion, engagement, and action, our system of scientific communication can be guided to meet the current needs, challenges, and exciting opportunities in the life sciences.

**ACKNOWLEDGMENTS.** I thank Walter Huynh, Courtney Schroeder, and Phoebe Grigg for their considerable help with the analyses of publications presented in this paper. I also thank Ron Germain, Satyajit Mayor, Richard Sever, and Harold Varmus for their detailed comments on the initial manuscript, and the many other individuals who commented on the article after it appeared on bioRxiv.

**1** Bertuzzi S, Drubin DG (2013) No shortcuts for research assessment. *Mol Biol Cell* 24(10):1505–1506.
**2** Alberts B, Kirschner MW, Tilghman S, Varmus H (2014) Rescuing US biomedical research from its systemic flaws. *Proc Natl Acad Sci USA* 111(16):5773–5777.
**3** Raff M, Johnson A, Walter P (2008) Painful publishing. *Science* 321(5885):36.
**4** Snyder SH (2013) Science interminable: Blame Ben? *Proc Natl Acad Sci USA* 110(7):2428–2429.
**5** National Institutes of Health Advisory Committee to the Director (2012) *Biomedical Research Workforce Working Group Report* (National Institutes of Health, Bethesda).
**6** Bourne HR (2013) The writing on the wall. *eLife* 2:e00642.
**7** Sachs F (2007) Is the NIH budget saturated? Available at www.the-scientist.com/?articles.view/articleNo/25416/title/Is-the-NIH-budget-saturated-/. Accessed October 10, 2015.
**8** Watson JD, Crick FH (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171(4356):737–738.
**9** Wilkins MH, Stokes AR, Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. *Nature* 171(4356):738–740.
**10** Franklin RE, Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature* 171(4356):740–741.
**11** Watson JD, Crick FH (1953b) Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171(4361):964–967.
**12** Wilkins MHF, Seeds WE, Stokes AR, Wilson HR (1953b) Helical structure of crystalline deoxypentose nucleic acid. *Nature* 172(4382):759–762.
**13** Franklin RE, Gosling RG (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* 172(4369):156–157.
**14** Meselson M, Stahl FW (1958) The replication of DNA in Escherichia coli. *Proc Natl Acad Sci USA* 44(7):671–682.
**15** Brümmer J (2003) How genius can smooth the road to publication. *Nature* 426(6963):119, discussion 119.
**16** Brown MS, Dana SE, Goldstein JL (1973) Regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity in human fibroblasts by lipoproteins. *Proc Natl Acad Sci USA* 70(7):2162–2166.
**17** Brown MS, Goldstein JL (1974) Familial hypercholesterolemia: Defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity. *Proc Natl Acad Sci USA* 71(3):788–792.
**18** Goldstein JL, Brown MS (1973) Familial hypercholesterolemia: Identification of a defect in the regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity associated with

overproduction of cholesterol. *Proc Natl Acad Sci USA* 70(10): 2804–2808.
**19** Hershko A, Ciechanover A, Rose IA (1979) Resolution of the ATP-dependent proteolytic system from reticulocytes: A component that interacts with ATP. *Proc Natl Acad Sci USA* 76(7):3107–3110.
**20** Ciechanover A, Heller H, Elias S, Haas AL, Hershko A (1980) ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci USA* 77(3): 1365–1368.
**21** Hershko A, Ciechanover A, Heller H, Haas AL, Rose IA (1980) Proposed role of ATP in protein breakdown: Conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc Natl Acad Sci USA* 77(4):1783–1786.
**22** Vale RD, Schnapp BJ, Reese TS, Sheetz MP (1985) Movement of organelles along filaments dissociated from the axoplasm of the squid giant axon. *Cell* 40(2):449–454.
**23** Schnapp BJ, Vale RD, Sheetz MP, Reese TS (1985) Single microtubules from squid axoplasm support bidirectional movement of organelles. *Cell* 40(2):455–462.
**24** Vale RD, Schnapp BJ, Reese TS, Sheetz MP (1985) Organelle, bead, and microtubule translocations promoted by soluble factors from the squid giant axon. *Cell* 40(3):559–569.
**25** Vale RD, Reese TS, Sheetz MP (1985) Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* 42(1):39–50.
**26** Vale RD, et al. (1985) Different axoplasmic proteins generate movement in opposite directions along microtubules in vitro. *Cell* 43:623–632.
**27** Daniels RJ (2015) A generation at risk: Young investigators and the future of the biomedical workforce. *Proc Natl Acad Sci USA* 112(2):313–318.
**28** Polka JK, Krukenberg KA (2014) Making science a desirable career. *Science* 346(6215):1422.
**29** Pickett CL, Corb BW, Matthews CR, Sundquist WI, Berg JM (2015) Toward a sustainable biomedical research enterprise: Finding consensus and implementing recommendations. *Proc Natl Acad Sci USA* 112(35):10832–10836.
**30** Committee to Review the State of Postdoctoral Experiences in Science and Engineering (2014) *The Postdoctoral Research Experience Revisited* (National Academies Press, Washington, DC).
**31** Varmus H (1999) *E-BIOMED: A Proposal for Electronic Publications in the Biomedical Sciences*. Available at www.nih.gov/about/director/pubmedcentral/ebiomedarch.htm. Accessed July 6, 2015.
**32** Desjardins-Proulx P, et al. (2013) The case for open preprints in biology. *PLoS Biol* 11(5):e1001563.
**33** Vale RD (2015) Accelerating scientific publication in biology. bioRxiv. Available at dx.doi.org/10.1101/022368. Accessed August 22, 2015.
**34** Royles SJ (2015) Waiting to happen: Publication lag times in cell biology journals. Available at https://quantixed.wordpress.com/2015/03/09/waiting-to-happen-publication-lag-times-in-cell-biology-journals/. Accessed July 6, 2015.
**35** Anonymous (2014) STAP retracted. *Nature* 511(7507):5–6.