



BY APOORVA MANDAVILLI

“Scientists discover keys to long life,” proclaimed *The Wall Street Journal* headline on 1 July last year. “Who will live to be 100? Genetic test might tell,” said National Public Radio a day later.

These and hundreds of similarly enthusiastic headlines were touting a paper in *Science*¹ in which researchers claimed to have identified a set of genes that could predict human longevity with 77% accuracy — a finding with potentially huge implications for medicine, health policy and the economy.

But even as the popular media was trumpeting the finding, other researchers were taking to the web to criticize the paper’s methodology. “We expect that most of the results of this study will not have the same longevity as its participants,” sniped a blog posted by researchers at the personal genomics company 23andMe, based in Mountain View, California.

Critics were particularly perturbed by the genome-wide association study (GWAS) that the authors had used to identify their longevity genes: the centenarians and the controls in the study had been tested with different kinds of DNA chips, which potentially skewed the results.

“Basically anybody that does a lot of GWAS knows this [pitfall], which is why we all said it so fast,” says David Goldstein, director of Duke University’s Center for Human Genome Variation, who voiced his concerns to a Newsweek

blogger the day the study appeared.

This critical onslaught was striking — but not exceptional. Papers are increasingly being taken apart in blogs, on Twitter and on other social media within hours rather than years, and in public, rather than at small conferences or in private conversation. In December, for example, many scientists blogged immediate criticisms of another widely publicized paper² — this one heralding bacteria that the authors claimed use arsenic rather than phosphorus in their DNA backbone.

A CHORUS OF DISAPPROVAL

To many researchers, such rapid response is all to the good, because it weeds out sloppy work faster. “When some of these things sit around in the scientific literature for a long time, they can do damage: they can influence what people work on, they can influence whole fields,” says Goldstein. This was avoided in the case of the longevity-gene paper, he says. One week after its publication, the authors released a statement saying, in part, “We have been made aware that there is a technical error in the lab test used ... [and] are now closely re-examining the analysis.” Then in November, *Science* issued an ‘Expression of Concern’ about the paper³, in essence questioning the validity of its results.

When asked for a comment by *Nature*, the lead investigator on the paper, Paola Sebastiani, a biostatistician at Boston University

in Massachusetts, said only that she and her co-authors “feel it is premature for us to talk about our experience because this is still an ongoing issue”.

For many researchers, the pace and tone of this online review can be intimidating — and can sometimes feel like an attack. How are authors supposed to respond to critiques coming from all directions? Should they even respond at all? Or should they confine their replies to the conventional, more deliberative realm of conferences and journals? “The speed of communication is ahead of the sheer time needed to think and get in the lab and work,” said Felisa Wolfe-Simon, a post-doctoral fellow at the NASA Astrobiology Institute in Mountain View, California, and the lead author on the arsenic paper. Aptly enough, she circulated that comment as a tweet on Twitter, which is used by many scientists to call attention to longer articles and blog posts.

To bring some order to this chaos, it looks as though a new set of cultural norms will be needed, along with an online infrastructure to support them. The idea of open, online peer review is hardly new.

NATURE.COM
To join the debate about online review, go to:
go.nature.com/b49ej5

Since Internet usage began to swell in the 1990s, enthusiasts have been arguing that online commenting could and

should replace the traditional process of pre-publication peer review that journals carry out to decide whether a paper is worth publishing.

"It makes much more sense in fact to publish everything and filter after the fact," says Cameron Neylon, a senior scientist at the Science & Technology Facilities Council, a UK funding body.

FAST FEEDBACK

In some fields, notably mathematics and physics, this sort of public discourse on a paper has long been the norm, both before and after publication. Most researchers in those fields have been depositing their draft papers in the preprint server arXiv.org for two decades. And when blogging became popular around the turn of the millennium, they were quick to start debating their research in that form.

Scientists in other fields seem less willing to get involved in pre-publication discussion. Biologists, in particular, are notoriously reluctant to publicly discuss their own work or comment on the work of others for fear of being scooped by competitors or of offending future reviewers of their own work. Adding to the disincentive is the knowledge that tenure committees and funding agencies do not explicitly reward online activity.

As a result, several journals — including, in 2005, *Nature* — have tried and mostly failed to interest scientists in various forms of open review. "Most papers sit in a wasteland of silence, attracting no attention whatsoever," says Phil Davis, a communications researcher at Cornell University in Ithaca, New York, and executive editor of *The Scholarly Kitchen*, a blog run by the Society for Scholarly Publishing in Wheat Ridge, Colorado.

Journals have had a little more success with post-publication peer review in the form of comments to the online versions of their papers. But the discussion is hardly vigorous, largely because the journals have usually solicited these post-publication critiques on their own websites, rather than on popular social networking sites.

"Who in their right mind is going to log on to the *PLoS One* site solely to comment on a paper?" asks Jonathan Eisen, academic editor-in-chief of *PLoS Biology*, and a prolific blogger and tweeter. "I guarantee that there are more comments on Twitter about a *PLoS* paper."

The question for researchers is how to deal with this ad-hoc analysis of papers. Unstructured, unruly and often anonymous, online commenting can be exasperating for biologists used to more conventional means of discussion. Like Sebastiani, for example,

Wolfe-Simon initially tried to stay out of the brouhaha over the arsenic paper. "Any discourse will have to be peer reviewed in the same manner as our paper was, and go through a vetting process so that all discussion is properly moderated," she said when the controversy first erupted. She and a co-author later did provide answers to a few of the criticisms on her website.

But Goldstein, who has also had publications on the receiving end of negative online reviews, tries to take the process in his stride. "I think if the work is solid, it holds up over time and this chatter is not going to hurt solid work," he says. Nonetheless, he adds, "there can be a herd mentality to this, which one wants to be really careful of" — especially for examples such as the longevity and arsenic papers, for which neither the rapid spike in fame nor the equally sharp fall into disrepute may be fully justified.

One solution may lie in new ways of capturing, organizing and measuring all these scattered inputs, so that they end up making a coherent contribution to science instead of just fading back into the blogosphere. Perhaps the most successful and interesting experiments of this type can be found at websites such as Faculty of 1000 (F1000) and thirdreviewer.com, and in online reference libraries such as Mendeley, CiteULike and Zotero, which allow users to bookmark and share links to online papers or other interesting sites.

F1000, which was launched in 2002 and evaluates papers from journals across biology, is among the best known of these websites. It now relies on a 'faculty' of more than 10,000 peer-nominated researchers and clinicians who select, evaluate and rate papers with a score of 6 ('recommended'), 8 ('must read') or 10 ('exceptional'). The individual scores are then combined using a formula to generate the paper's F1000 article factor. These scores, in

turn, are making some appearances in tenure packages and grant applications. "It's the only one we've been using in any systematic way," says Liz Allen, who leads post-award evaluation at the Wellcome Trust in London. "It adds another dimension to the citation index."

However, critics note that F1000 rankings tend to correlate closely with traditional citations, which suggest that they add little, if any, extra value. And most papers never attract the attention of the faculty members, so that they are never ranked at all. Even one as talked-about as the longevity paper garnered only a single rating on F1000: a must-read score of 8. For comparison, the currently highest-ranked paper on the site has an aggregate score of 62, and scores of 20 or more are common.

META-TWITTER

Given the vagaries of such measures, there is a growing interest in methods that would aggregate and quantify all of the online responses and evaluations of a paper — producing what Neylon and some others are referring to as 'alt-metrics' — and compare it with more conventional metrics.

"As scholars migrate to newer forms of communication, it becomes very important to measure what they're doing and to compare," says Jason Priem, a second-year graduate student in information science at the University of North Carolina in Chapel Hill, who is focusing his study on alt-metrics.

Neylon is leading a £30,000 (US\$50,000) grant proposal to create and test a working alt-metrics prototype that would rapidly measure a paper's impact by assessing all the activity surrounding it online. In addition, he and many of his colleagues champion a completely online system of pre-publication peer review that would build on the arXiv.org model, and would replace what they see as a flawed process with a more egalitarian and transparent one.

That last step, however, may be a bit harder than most scientists are willing to go — even the ones who energetically blog and tweet their post-publication reviews. Although the latter activity is "a nice secondary mechanism for catching things," says Goldstein, "I think we do not want it to be just a commentary free-for-all as the only arbiter of quality."

"It's exactly like what's said about democracy," he adds. "The peer-review process isn't very good — but there really isn't anything that's better." ■

Apoorva Mandavilli is a writer based in New York.

1. Sebastiani, P. *et al.* *Science* doi:10.1126/science.1190532 (2010).
2. Wolfe-Simon, F. *et al.* *Science* doi:10.1126/science.1197258 (2010).
3. Alberts, B. *Science* **330**, 912 (2010).

No new data!



@ironlisa The speed of communication is ahead of the sheer time needed to think and get in the lab and work.

