



PLOS BLOGS

# **Absolutely Maybe**

Browse all PLOS Blogs

# Reasons to Worry Less About the Explosion of Preprints

May 25, 2020 / By Hilda Bastian / Science Communication

59





It looks as though preprints are here to stay in biomedicine, and I think that's great. But I've been hearing variants of this cry for weeks now: *The plague brought a plague of preprints! They're a menace!* 

It's easy to sympathize with this impulse. We've heard it all before, though, about other communication innovations before we got our heads around them. I think it's probably safe to say we hear it after every new form of communication emerges. Consider this quote – you could insert preprints into the space, and it would fit right in with current conversations:



Is there anywhere on earth exempt from these swarms of new []?

Who said it, and what's the missing word that caused this frustration? It was this guy, the famous Dutch philosopher, Erasmus:



And the missing word is "books". Yes – *books*! Gutenberg's printing press had ushered in the age of mass book writing and reading. Here's more context for what he wrote – **in 1508**:

It is the innumerable crowd of printers that now throws all into confusion, especially in Germany... [Y]ou may print anything. Is there anywhere on earth exempt from these swarms of new books? Even if, taken out one at a time, they offered something worth knowing, the very mass of them would be an impediment to learning... They fly out in swarms, some of them with no author's names...

Erasmus, Adages, 1508

We developed ways to manage books, though, and we would later come to manage the field of biomedical journals. It's always a work in progress, as new problems arise and we need to keep improving our methods of filtering it all. But that fear of uncontrolled flow of information, the concern that the masses can't handle it, that there are too many people who can't be trusted with it – that never really goes away, does it? Although you would be hard put to find someone trying to make this case about books now!

What Erasmus couldn't envisage were the ways we'd start to develop means of wrangling the swarms. Classification systems would emerge, simple at first, but increasingly elaborate. And that later generations would just find it normal.



Woodcut of Leiden University Library in 1610, by Johannes Woudanus (**via Wikimedia Commons**)

Along the way, an awesome profession would emerge, dedicated to libraries and the science of organizing and retrieving information. **Melvil Dewey**'s classification system probably made his the most recognizable name in library science history. As I wrote **here**, though, librarianship became a female-dominated profession in the **19th century** – because it was "a new and fast-growing field in need of low-paid but educated recruits". In the US, the first woman entered the profession in 1852 – and by 1910, 79% of librarians were women. Only teaching was more female. We owe them so much, and we need them for coming to terms



We need those information scientists to study how we can do the most thorough and most efficient searches for preprints now. That's not as straightforward as you might think. Firstly, there are a lot of preprint servers now – the jargon for online preprint databases. Secondly, just because something's in a database, doesn't mean you'll find it. That's partly a function of what terms you use, but it also depends on the database's search engine. And thirdly, as the quantity grows, you have to trade off between finding every possibly relevant preprint and having a manageable number of hits to sort through.

Daniel Garisto published a good backgrounder **late last year** on the history of preprints, and the beginnings of their adoption in biomedicine. And I wrote a post about the pros and cons of preprints in biomed **back in 2016**. I don't think anyone, though, had "the worst pandemic in 100 years will massively expand the use of preprints overnight" on their bingo cards. But we already have at least one preprint and at least one journal article about it! Ironically, it was a journal publication about preprints that appeared first.

It was by Maimuna Majumder and Kenneth Mandl, **in March**, analyzing media and other interest in preprints versus journal articles about the reproduction number for the new coronavirus. They concluded that because of the speed of release of preprints, they were driving the discourse, not journal articles. Decision-making can be informed quickly, they point out, but it can go badly wrong, too, as when a preprint had to be retracted after an outcry, because it "erroneously claimed that COVID-19 contained HIV 'insertions'".



The preprint about pandemic preprints just appeared **on 23 May**. It's by Nicholas Fraser and colleagues who are all involved with preprint servers in one capacity or another. They estimate that within the first 4 months from when the first person with Covid-19 was identified, there were 16,000 scientific publications, of which 6,000 were preprints. There were already 166 by the end of January. Although it's relatively new, the majority of Covid-19 preprints have appeared on medRxiv.

Some of them have been invaluable, and the speed of access to the information has been part of what made them valuable. It's not just the retracted one that has caused alarm, though. In fact, this post arguing that we should worry less about preprints, is actually the third of a trilogy that began because of an influential preprint that I think is disturbing and unreliable. And that preprint itself appeared on the back of a hugely controversial previous preprint by that author.

The saga begins with the preprint of the **Santa Clara seroprevalence study**, a highly contested piece of research, one of the authors of which is John Ioannidis. James Heathers wrote an informative and entertaining **recap** of the preprint and reaction to it, as well as the arguments about preprints themselves it provoked. He wrote, "the idea of releasing work previous to 'formal' publication isn't the problem — it's us". I agree. He argues that a critical part of the preprint process is responding to criticisms. Heathers also criticized the authors' media campaign: "The preprint-followed-by-immediate-formal-demand-for-attention is a disgusting new normal". Really, though, isn't that how scientific conferences have often

#### functioned too, based only on some peer review of abstracts?

The first Covid-19 preprint I **criticized** was a review Ioannidis did of the infection fatality rate (IFR) based on seroprevalence studies, including his own. His preprint was posted **on 19 May**. I think its problems were grievous, including a biased sample of studies and methods of data analysis, data errors, and an unsubstantiated claim about its comparability with influenza. **Part 2** of this trilogy looked at 2 other studies of IFRs posted as preprints on 18 and 19 May. One of those was another review, that I think also had grievous problems. Basically, I didn't think 2 of those 3 preprints rose to a high enough standard for scientific publication.

The thing is, we really don't have grounds for confidence that a journal would have ensured these problems were solved before publication. For example, back in November 2019, I wrote a post **here on this blog** tearing into another of Ioannidis' studies, which had some very similar problems to his IFR review preprint.

In the last few days, a Covid-19 study published at arguably the world's most influential journal by 53 authors turned out to have errors that are **real shockers**. By the end of March, there had been another high-profile issue, as Adam Marcus and Ivan Oransky **reported**, when it turned out a key influence on the US government's disastrous early decision about Covid-19 tests had been a single unreplicated paper that had since been retracted from a journal. As if that's not enough, Marcus and Oransky write, the paper that set off the whole hydroxychloroquine caper was published by a journal the day after it was submitted – not a whole lot of quality control going on there, that's for sure.



At the heart of the concern about preprints, is a concern about research that hasn't had enough peer review. I often write about **the evidence around peer review**. And while peer review can obviously make an important difference to a manuscript, the only kind of peer review that we can say with certainty makes a difference is **peer review by statisticians**. Now that so many journals are publishing peer reviews, it's pretty obvious why: a lot of peer review reports are skimpy – even when there are glaring problems with the manuscript.

For pre-publication peer review to really work in the way many people think it does, it would need a lot more than just a few peer reviewers for studies that could be influential. The benefits editorial peer review sometimes provides, though, isn't likely to outweigh the harm of the delays and drain on people's time of the system we have now. Authors are often going through the process at more than one journal before their manuscript is published – **one estimate** is that there's 15 million hours spent each year on *redundant* peer reviews.

And then there's the enormous time suck of re-formatting a manuscript for each new journal's submission requirements. Insane. May 2020 was also a reminder that preprints have the potential to free science from that ball and chain. Many journals will accept preprints as the submission now. But on **13 May**, *eLife* unveiled another innovation, combining that benefit, with portable peer review. Now you can request peer review of your preprint from them. If you're lucky and get chosen, you can have that peer review posted if you like, and use it for another journal if *eLife* doesn't accept your manuscript. Scale is obviously a serious constraint here. But it's a great reminder of why preprints are a reason to be cheerful. Disruption isn't comfortable, and it takes a lot of adjustment. But it can,

sometimes, lead to valuable transformations.

~~~~

Disclosures: My **scientific publications** include journal articles and preprints at bioRxiv and medRxiv, and I will be posting more preprints soon. All my preprints have been, or will be, submitted to journals. I have been writing about, working with, and studying the management of the growth of clinical trials and systematic reviews **since 2010**. It's a focus of my doctoral thesis on the ways shifting evidence affect the reliability of systematic reviews, including some analysis of search strategies (submitted April 2020). From 2011 to 2018, I worked on projects for PubMed, at the National Center for Biotechnology Information (NCBI) (part of the US National Library of Medicine at the National Institutes of Health).

The cartoons are my own **(CC BY-NC-ND license)**.(More cartoons at **Statistically Funny** and on **Tumblr**.)

The portrait of Erasmus is by Hans Holbein (the Younger), the Louvre **via Wikimedia Commons**.

Woodcut of Leiden University Library in 1610, by Johannes Woudanus in 1649 (**via Wikimedia Commons**).

The photo of unnamed librarians in the Webster Public Circulating Library in New York City in the late 19th or early 20th century, photographer not known, also comes **via Wikimedia Commons**.

LEAVE A REPLY

Your email address will not be published. Required fields are marked \*

COMMENT

#### NAME\*

#### EMAIL\*

#### WEBSITE

ORCID

Add your ORCID here. (e.g. 0000-0002-7299-680X)

Save my name and email for the next time I comment.



**RELATED POSTS** 

#### LISTICLES

### 5 Things We Learned About Peer Review in 2019

December 31, 2019 / By Hilda Bastian

For something so fundamental to the practice of science, it's perplexing that it took so long for serious research into...

Read more

SCIENCE COMMUNICATION

Who You Gonna Search? Engine News for Systematic Reviewers October 31, 2019 / By Hilda Bastian

Isn't it almost always the way? I had barely finished my recent post about alternatives to Google Scholar, it seems to...

Read more

### SCIENCE COMMUNICATION

## **Google Scholar Risks and Alternatives**

September 27, 2019 / By Hilda Bastian

I wasn't there. But it sounds like a ripper of a talk. "There seems to be a narrowing of our...

Read more

PLOS is a nonprofit 501(c)(3) corporation, #C2354500, and is based in San Francisco, California, US

#### Home

Resources

**Research Communities** 

Open Science

Publish with PLOS

About PLOS

Media

Contact

Pay Invoice

Careers

Advertise

**Financial Overview** 

Privacy Policy

Terms of Use

Text & Data Mining

**News & Updates** 

Email address

I have read and agree to the terms of the PLOS Privacy Policy and hereby consent to send my personal information to PLOS.

Submit