

Transcript

Title: Preprints and the changing landscape of biomedical publishing

Speaker: Jonny Coates, ASAPbio

Introduction

Good afternoon everyone. Welcome to the Jülich Open Science Speaker Series. Today we are very happy to have Dr. Jonny Coates. Dr. Jonny Coates obtained his PhD in immunology and then became interested in meta-research. And he has since become one of the foremost supporters and proponents of preprints. So he has a lot to tell you about preprints today. So Jonny, the floor is yours.

Lecture

Thank you. Yeah, it was a strange career transition. I now do not do any immunology. All I do now is chat about how amazing preprints are to anyone who will ever listen to me. And even if you won't listen to me, I'll still talk about it.

So I'm going to give you an overview of what preprints are, where they sit in the sort of current landscape of publishing. And I'm going to do that in the framework of a different model for scientific publishing, one that I think preprints fit very well into and is where we should at least try and go. There are some pitfalls to that, which hopefully I'll remember to talk about. And then I'm going to end on where preprints sit in terms of the open access arguments.

I am the Associate Director of ASAPBio, which is a non-profit focused on the productive use of preprints in the life sciences. So preprints are used across a bunch of different scientific disciplines, more so in some than others, are focused very much within biomedical science.

You can find these slides, or a version of these slides anyway, through this link here or through the QR code. If you do have any questions, you can find me on Twitter, Blue Sky. You can find ASAPBio on the same places or on Mastodon, or you can send me an email and I will happily have more chats with anyone who wants to have them about anything to do with preprints. So I've already told you what ASAPBio is, so we can skip that one.

So one of the big problems with academia, never mind publishing, is that in the age of the Internet, academia is one of the few things that just seems to have never really quite caught up to that. We use the Internet, but there's been all these technological advances, and yet when we look at things like journal publication, it actually takes longer than it ever has.

So you're seeing here some data showing the time it takes to get your first author paper. And in the 80s, this was just under five years on average, whereas about 10 years ago, this had ridden to about six years on average. And obviously, this is just increasing as we go on, so that average will now be higher still.

And this is one of my favorite plots I think I ever show anyone. This is data looking at various different journals. In here, you're seeing Cell, Nature, and the Journal of Cell Biology. And again, it's looking at the 80s compared to about 10 years ago. But what you're seeing is a brilliant example of partly why this graph on the left exists.

In the 80s, there were way fewer requirements in terms of the number of panels, and therefore the amount of data, the number of experiments needed in a paper to get published. Not only that, but you can see in the 1980s, or 1984 in particular, the supplemental idea just didn't exist. So the thing that we've seen between 1984 and what you're seeing here in 2014 is that not only has the number of panels in a paper more than doubled in pretty much every case, but actually the amount of supplemental material, so the number of panels and supplementary material, has also pretty much doubled, if not more so in that time.

And so the amount of work has really increased about fourfold, if not more, just to get a paper published in the same journals. And on one hand, that's great, because that reflects an advancement in the kind of things we can do and the kind of questions we can ask.

A single-cell sequencing paper is going to have a lot of panels, because it generates a lot of data. But on the other hand, this also reflects the fact that people are demanding more than they used to. And altogether, this slide is quite problematic, because we don't want to delay the time it takes to get papers out there, because then that impacts people's careers, but it also impacts the speed at which we share scientific findings and knowledge.

And the more data that's required, the more we have to ask of that system in terms of reviewer time, the time it takes to do revisions for papers, the money that is involved in all these things. And again, it comes back to careers, because if you then require a lot of panels and a lot of experiments to publish in a high-impact, in air quotes, journal, then that restricts that kind of science to the labs that can actually afford to do that.

And so, it's no longer an equitable system. And I do recommend go read those papers, because they are very interesting. And so, really what we need to do is to untangle this idea of publishing, because at the moment, publishing consists of these broadly, these three areas.

We've got dissemination, peer review, and then this curation and prestige element. And right now, they are tied up into journals. Journals do all those three things all at once. And this actively motivates delaying of sharing for the idea of creating a complete story. And that's needed for a lot of reasons in the current system. It's needed for jobs and for funding. It's also how we look at each other and we assess each other as researchers. And this is, we think, the big problem in publishing, one of the big problems in publishing.

How do we detangle that? Well, one way of doing that is the published review curate model. So, in this idea, you detangle dissemination from the peer review and the curation and prestige, and really you could detangle them further. What we argue is that dissemination should be pre-print driven. You post your pre-print, and I'll get onto what pre-prints are if you don't know what they are soon. And then you go through post-publication peer review, and then you can go through a separate step all at the same time to the curation step. I would argue we should draw prestige from this. Prestige should be a different thing entirely, but that's not the subject of this talk.

In the current system, you submit your manuscript to a journal. It may then go out for review. If it doesn't go out for review, you submit to another journal. Assuming it does go out for review, the reviewers will either reject or recommend changes. It's very rare they just accept.

You go through rounds of revision, and this can be rounds, rather than just a single round. Eventually, what happens is either that journal will accept or reject, and you might have to go through more rounds of revision. But eventually, you get to a journal, and all the reviewers are happy, and they will then publish your work.

This takes months to years. I always like to give the example, when I was an undergraduate, I was in a lecture once, and the lecturer was talking about a bit of work he published, and it took him 10 years from start to actually publish that one paper, which I always thought was insane. It was like half my lifespan at that point.

Preprints, on the other hand, so still the idea of the scientific manuscript, but you submit this to a preprint server, and on average, within 48 hours, that work is then public. It is then open for community feedback and discussion, and you can then still, at the same time, submit to the system and get it published.

Or, and this shows you how quickly things develop, because this slide is out of date, you can then take that preprint straight to a journal and potentially have it published. [eLife](#)'s model is kind of that. Now, I say this is 48 hours on average, even at the height of the pandemic, it was still about 48 hours turnaround time, which was amazing effort on their part. Journals cannot do anything that quickly, unless it becomes rather questionable practices.

If you're not familiar, what are preprints? Preprints are not peer-reviewed by the server before posting, and the typical definition you might see for a preprint is it is a manuscript shared prior to journal-organized peer review. They are posted rapidly. Preprints are versioned, and that means they are very easy to update or correct, and everyone here will have published a paper and then found out there's like 20 typos in there, now you can't do anything about it.

They are compatible with most journals now. There's only a few journals that really don't like preprints, and they are very much in the minority. You can always check if you need to figure out if your journal accepts preprints, but like I said, the vast majority accept preprints.

Preprints are citable, and so you can use those as you would a paper. They do receive a DOI, so they're very stable, permanent objects. And citations to preprints can be pooled across various different tools, so Google Scholar is a good example that does this. And a lot of funders now also consider preprints as evidence of productivity when you're applying for grants and fellowships.

Preprints in the life sciences actually go back way before any other preprint system, really. Physics is the one often people will call out as saying, they've been doing preprints since the 90s, and they got there first. Well, actually, life sciences have been doing preprints since the 1960s, so we got there first.

But those experiments kind of ended very quickly, and this is for a lot of reasons. Primarily, which some people might find surprising, is that it was society publishers who ended those experiments. They at the time really did not like preprints as an idea. And so, it wasn't really until about 2013 when preprints in the life sciences, I would say, became established. And there's a lot of factors around why that happened and worked at the time.

But in 2013, bioRxiv was launched. They're actually celebrating their 10th anniversary this month. And since then, we've seen preprints increase in the life sciences. This slide is also slightly outdated, so there's over 600,000 life science preprints now. And you can really see around 2020 and the few years nobody really likes to talk about, the explosion of preprints that happened in that period.

There's also been a lot of big sort of events around preprints, so preprints being on PubMed. Europe PMC, in particular, is a really, really good source now for preprint information. That's where we get this data from. And then, probably the most recent would be eLife switched their model up to focus entirely on preprints. They've been doing that a little bit for a few years before they took the full leap.

And so, preprints are really, really growing. Now, this looks great, but actually, this is still only about 8% to 10% of the total biomedical literature is preprinted first. So, it's still a very small number. Now, I think that's good, in that it means a lot of the problems with publishing are avoided at the moment. And it means we can also experiment, as I'll talk about a little bit, and do a lot of cool things that we can't do with journals and with the current system. But that is also quite a small number, and it would be great if it was bigger.

One of the benefits of preprints is that they reduce barriers to sharing. I said on the first slide that a lot of the problems at the moment have all come back down to equity, and access is a big part of that. So, what you're seeing here is the percentage of preprints published by country. And 65% of preprints are currently published from high-income countries, versus 25% published from low-income countries.

And this, from conversations I've had, certainly would reflect that the APC model has failed, because low-income countries cannot afford to pay those costs. They also can't afford the livelihood subscription costs. But what this is showing you is that preprints are a way around that. Although only 25% of preprints have been published from low-income countries, they are still available for everyone to read for free.

That research is not hidden, it's not lost, it's out there, it's available, and we can all build on it and use it. And this is a really important step into bringing more quality into academia, and also into the publishing system. And COVID is a really good talking point for preprints, because it highlights a lot of the benefits of preprints.

When COVID hit, preprints were already starting to increase, but the pandemic really accelerated that. And what you're seeing here are just the number of preprints by date for COVID research. And you can see it really does explode pretty quickly. I can't remember if the data is in here or not, but about 25% of the entire COVID-19 literature in the first 10 months was a preprint.

We actually looked at COVID preprints, and they were shorter than non-COVID preprints, they were updated more frequently, which is another really good benefit of preprints. Obviously, that loss reference counts. They did actually look different compared to non-COVID preprints. And this was a great test as to the benefits of preprints and what preprints can really do for not just science, but society.

And I say that because preprints were used to shape a lot of policy around COVID. I guess the best example I would give is that the UK lockdowns were based largely on preprint data. And this is something we did not see prior to the pandemic. We looked at various different policy documents from different bodies around the world, and preprints were not really mentioned until the pandemic hit. And then everyone was suddenly using them.

The World Health Organization used them. The CDC in the US and the EU both referred to preprints. So there really was a cultural shift. Yeah, so this is the 25% data. So not only was there a lot of preprints being posted about COVID, but they also received a huge amount of attention. So focusing on the graph you see on the right, they were very highly cited.

They got an enormous attention on Twitter. They also got a lot of attention in news articles. And again, that was something we hadn't seen really prior to that. So journalists, policymakers, and the general public now were all kind of turning their attention towards preprints in a way they hadn't before. Scientists, like I said, had been using them, but they were now certainly a lot more aware of them and using them a lot more.

One of the things that comes up a lot, particularly during the pandemic, but even since and before, is that the idea that because a preprint is not peer-reviewed, it must be low quality. That is abstractly false. And coming back to that, it's only about 8% to 10% of the literature. The benefit of that is that your preprint is probably going to get more attention than it would if everything in the world was ever preprinted.

And so I don't really understand why anyone would think a preprint would be low quality, because you putting out there for everyone to read and everyone can read it and they will comment on it. There are plenty of examples where people have picked up shoddy, very bad science. I think there's an example in here. And we've done that because it's been a preprint, because it's been open to read and it's accessible. We've looked at this question and we've shown very much similar data to the example here.

This was a study that looked at the confidence intervals, comparing the preprint to its published peer-reviewed version. And what this effectively is showing you is that there's basically no change in the confidence intervals. They were smaller by about 7% on average. And this is a very good example of how this entire body of literature looks.

Any study that compares a preprint to its peer-reviewed version basically tells you the same thing, which is that as a preprint undergoes peer review to be published, very little changes. Generally, the data gets strengthened a little bit, but the key conclusions in particular do not change, which I don't think really is a surprise to anyone.

I don't think even if it's a preprint or not, the key conclusions generally don't change with peer review. But this really highlights the problem with the amount of emphasis I think that we put on peer review. But also shows that preprints are actually good quality.

There are a lot of benefits and a lot of concerns around preprinting. That was one big one that we come across. Premature media coverage is another big concern that we often come across, as is things like the idea of being scooped and just general uncertainty. These are mostly myths.

So the idea of scooping, for example, is a myth that just will not go away. I've just written a piece about that. And we're really trying to tackle a lot of these to try and reduce the barriers that people have towards preprinting.

And here's my example. So I said that public feedback was really important for not only correcting the size of literature, but it's important to maintain the quality. This is a really good example of during the pandemic where a preprint appeared and it got a lot of attention because it frankly wasn't very good. And it had problems.

The attention very quickly translated into withdrawal. So preprints are never removed. They just get this little sort of tagline saying it's been withdrawn. This happened within about 48 hours, which is incredibly quick. And if you compare that to any kind of publisher, that's just not how they work.

We actually had a meeting yesterday talking about correcting the scientific literature and some of the problems with publishers and their reluctance to do that. Because the difference here is that preprints don't really have any conflict of interest against removing work.

Whereas a publisher, that can reflect badly on their brand. And as one of the publishers yesterday highlighted, they get a lot of threats towards legal action and lawsuits when they try and work. So this is a really, not only is this an important aspect of how we do science, but it's a really helpful benefit of preprints and this kind of speed and correcting things.

The alternative, of course, is that also if the authors have feedback that is negative, they can then update their preprint and so address those issues. And bioRxiv in particular did a great job of actually integrating with a lot of feedback providers. You're seeing an example here of eLife, but if you've not used bioRxiv, they have this cool little toolbar on the top and it's got links from community comments, anything on Twitter, blogs, and then TRiP here is their review service feedback.

When you read a preprint, if it has been reviewed, everything is all in one place, which is a really good signal for trustworthiness, I guess, of the work. And this is important because in the traditional system we're currently in, peer reviewers often feel like they act as gatekeepers. Peer reviewer number two is a popular meme, and that's not necessary because peer reviewers can be collaborators.

This is data from Review Commons, slightly old data, but Review Commons is a service that will review preprints prior to journal submission, and then various different journals will allow you to submit your preprint and the reviews at the same time to speed the whole process up.

And this is a survey of authors who had gone through that process, and generally what you're seeing is that they did find the reviews were more collegial, and this is also anecdotally backed up by pretty much everyone you speak to, and that the reviews were more reasonable.

And also that the process is, so what you don't see here really is that everyone felt the process was faster, although from the other side, so from the sort of journal side, we find that generally people think the process is sped up through these kind of ways of doing things.

And preprint review really opens up what we can do with not only preprints, not only review, but also just that wider publishing landscape because we can do a lot of experimenting. eLife is basically doing a massive experiment at the moment in how they do things, but all the things like Review Commons, peer community, and all these review services are experiments.

And this slide is just kind of highlighting some of the different things we can do here, but this is another huge benefit of preprints that goes way beyond preprint review. We can experiment with format, we can experiment with how often you update a preprint and what it means to publish, because one of the projects we have at the moment is looking at those preprints that where people post early data and then update their work as they get more and more data in, which is how science works and is one way of changing and getting away from that traditional journal article output.

Looking back to the preprint feedback, the whole ecosystem has really exploded in recent years. It was quite simple a few years ago, there were not a lot of services, you had things like pre-review and preprints, but certainly since the pandemic, the whole system has really, really grown a lot.

And how does this fit into our model? I said dissemination was posting a preprint. We now got a system where you post a preprint and then it can undergo peer review, that can then also be separate from this creation step, or it can be part of it by sending those things together to a journal. And so we're breaking this triad that publishers currently dominate down into manageable little chunks and making it a more equitable, a fairer system, and also one that often puts scientists back in charge.

Preprint review has been growing a lot over the past few years. It is a relatively new thing, so these numbers are not huge, but it is on the rise and there are a lot more services still coming through.

The other thing that preprint peer review has really done is that it's leading to a much stronger formal recognition for preprints. We've got doctoral schools allowing preprints as part of their graduation requirements instead of having to have an accepted paper. EMBO, for example, allow you to use peer reviewed preprints in fellowship applications. Plan S, as I'm sure you've all seen the recent statement from Plan S regarding publishing, the preprints are very much fitting into that.

But what I've not spoken about really is preprints and Open Science, and this is where I think preprints really shine through. So this is some really interesting, very recent data, looking at the cost of Open Access. So this is the total revenue from article processing charges broken down by publisher.

In total, these five publishers have made over a billion dollars in article processing charges. Now remember, at the same time, they're also charging libraries unknown amounts of money, which I will strongly suggest is also probably higher than that number there, for subscription agreements. So this is just one stream of revenue for the publishers. Springer Nature making a lot of money, but they're all making a huge amount of money from article processing charges.

And these are things that, like I said at the start, a lot of low-income countries cannot afford. I'm in the UK and I can't afford to pay these fees. And the other problem with this is that these fees are increasing every year. So you're seeing data from 2015 up to 2018 here, and across all of those five publishers, their fees are increasing. And so this problem is only going to get worse.

The other thing we've seen in regards to Open Science and Open Access is that both the European Council and the US government have produced, this year they've produced two very similar kind of statements. So the EU produced a statement calling for transparent, equitable, open access to scholarly publications, particularly a format in which it is not for profit, and it has, importantly, no cost for authors or for readers.

And I would argue the system I present here, the Publish Review Create System, supported by preprints, is this system that the EU are calling for. The US have also called for a system, very similar wording, so freely available, publicly accessible by default, without any embargo or delay. And again, the system I've spoken about today fits that very well with preprints at the centre of that.

I think that's my last slide. Yeah. So if you do want to reach out and ask questions, you can chat with me or Jessica at ASAPbio. Here's a list of our board members, and you can get our slides if you want them. Thank you.