Commonplace

Rapid sharing of COVID-19 research

Jessica Polka[1], Dave O'Connor[2], Daniela Saderi[3], Richard Sever[4], Richard Wilder[5], Catherine Ahearn[6]

[1]ASAPbio,
[2]The UW Medical Foundation; Professor of Pathology & Laboratory Medicine at the University of Wisconsin,
[3]Co-Founder and Director of PREreview and Outbreak Science,
[4]Co-Founder of bioRxiv and medRxiv, Cold Spring Harbor Laboratory,
[5]General Counsel and Director of Business Development at the Coalition for Epidemic Preparedness Innovations,
[6]Knowledge Futures Group

Introduction

Preprints, open peer review, and the rapid sharing of interim research findings have the potential to accelerate the process of scientific discovery. In research on SARS-CoV-2, speed is paramount, and researchers are using these new tools as never before.

In the following conversation, held on March 31 among Richard Wilder (General Counsel and Director of Business Development at the Coalition for Epidemic Preparedness Innovations), Dave O'Connor (The UW Medical Foundation (UWMF) Professor of Pathology & Laboratory Medicine at the University of Wisconsin), Richard Sever (Co-Founder of bioRxiv and medRxiv, Cold Spring Harbor Laboratory) and Daniela Saderi (Co-Founder and Director of PREreview and Outbreak Science), we examine the use of preprints, rapid peer review, and informal channels to hasten communication of SARS-CoV-2 research.

If not for these channels, how would research and public health progress? As Richard Sever has pointed out, we have a natural experiment in the form of the 2003 SARS outbreak, which took place before preprints were widely used in biomedicine, and "93% of the papers written about the epidemic appeared after the epidemic had ended" in traditional journals. (COVID-19 papers are of course appearing in journals as well, often made publicly accessible by publisher commitments to open access, after accelerated peer review processes).

In May, many universities are carefully reopening their research programs, allowing researchers to filter back into labs. If, in Arundhati Roy's words, "the pandemic is a portal," the long process of recovery will offer a choice to scientists in other fields: whether to restore old practices or, having seen a new model with COVID, move their own fields in the direction of change.

—Jessica Polka, Executive Director, ASAPbio

Click here to view the recording of this webinar.

Conversation

March 31, 2020

Jessica Polka: Welcome everyone to this ASAPBio and KFG webinar on rapid communication of COVID-19 research. We are grateful to be joined by three wonderful panelists today to discuss the various ways in which they and their organizations are having conversations about research.

 This webinar is being produced by the Knowledge Futures Group, which builds technology for the production, curation, and preservation of knowledge and services for the public good. As for ASAPBio,

we are a scientist-driven non-profit working to promote innovation and transparency in Life Sciences communication.

I'd now like to invite one of ASAPBio's board members Richard Wilder who is also the General Counsel and Director of Business Development at the Coalition for Epidemic Preparedness Innovations to say a few words about the importance of rapid research sharing in the context of the coronavirus pandemic.

Richard Wilder: Thank you very much Jessica, I really appreciate the opportunity to be here and I'm proud to be a board member of ASAPBio. I think it's an organization in the promotion of pre-preints that has really pushed a lot in order to ensure that the output from scientific research is rapidly disseminated.

And yes, it's very important with respect to the response to COVID-19: I think we've all seen as the virus emerged and as it spread across the planet that it has moved very rapidly. It has moved in a way I think sometimes that gets ahead of public authorities and resources to be able to deal with it. For us that are involved in research and development activities to come up with interventions to stop it, or to slow it, being a fast to be able to come up with those interventions is critical. We need to be able to match the speed of the virus that we're trying to deal with.

And so, from CEPI's perspective, the Coalition for Epidemic Preparedness Innovations, we have a mantra that we use of pursuing 'speed, scale, and access'.

We're an organization that was established about three years ago now and was announced at the World Economic Forum, the purpose of which was to rapidly develop vaccines against a number of infectious diseases that have epidemic potential, as well as building up platforms that can be used to much more rapidly develop, and bring into existence, vaccines.

Our target was 16 weeks from the identification of new pathogen to have a vaccine that's beginning clinical trials. And with respect to COVID-19, we've been able to achieve that, and even beat that time frame. One of the things that is essential, to be able to be successful if you have a scientific undertaking for which speed is of the essence, is to be open and transparent about what you're doing and how you're doing it.

 One of the things that we have benefited from in responding to the COVID-19 outbreak is the openness and transparency of those that are developing gene sequences for the pathogen itself. Originally coming out of China and being published in an open format - both in terms of the technology that's used as well as what people are enabled to do with those gene sequences. So one could then download

and make of them to begin the process of developing vaccines, developing drugs, developing diagnostics, and so on.

And so openness with respect to access, and broad-ness in terms of permission as to what you can do with that data, with that information, is extremely important. And as we stand up our projects, as we have done since we were founded three years ago - and what we've been doing with respect to COVID-19 as I mentioned - access is very important and when we think about access, what comes first to mind of course is access to the vaccines once developed.

We are intended to be an organization that is developing vaccines that will be accessible and available where they're needed most to address outbreaks, to address epidemics, and now in this case to address pandemics. And so access to the vaccine itself, in terms of ensuring that we achieve the scale necessary to address the problem, and methods of distribution [are] very important as well. So, in that connection, making sure that we have full buy-in to what we need to do in terms of ensuring access requires openness and transparency on our part in terms of what we're doing and how we're doing it, and how we're setting up our arrangements and so on.

It also requires openness and transparency not only amongst the developers that we're funding. We have eight, and probably in the next couple of weeks let's say we have 10 projects up and running, for the development of vaccines against the virus, the SARS-CoV-2 virus that causes COVID-19. And so, we'll have up to 10 projects up and running, and for all of those projects we do require openness with respect to the data that's generated, including clinical trial data, and openness with respect to the publications and speed, including reprints to ensure that the results of the work that is being undertaken using our funding is disseminated broadly as we see it sort of close to home for the benefit of those projects that we're funding.

But there's also a number of other projects that are developing vaccines against the virus, that are developing therapeutics against the virus diagnostics and so forth, for which that same information, that same body of information would be important to be able to move forward and if we're going to be successful in our goal of developing a vaccine within 12 to 18 months of that original identification of a pathogen being able to do so in parallel rapidly is going to require that everyone that we're working with has full and fair access to the data that's being generated and information that's being generated.

Just the last thing I'd say in closing, just as a point of comparison, is that historically the vaccine development work has run to six years to ten years to develop a vaccine because you do things in the normal course, and we're not cutting corners in what we do, but what we're doing is there's a lot of work at risk where we're doing things in parallel, that would normally do in sequence.

And in order for us to be successful in meeting requirements of speed, scale, and access, I think openness and the ability of all the institutions that are involved to broadly share information and not just about the development of vaccines and the projects themselves, and what comes out of them, but what's happening on the ground in terms of epidemiology, what's happening with respect to different policy positions and how they're being implemented around regulatory approvals and so forth, need to be broadly rapidly, and preferably in real-time, communicated out to this community that we are all part of as we're addressing the COVID-19 disease.

So with that Jessica I'll stop my introduction and I know there's an opportunity for questions as we go along, but look forward to the rest of the discussion, so thank you very much.

Jessica Polka: Thank you so much. I just want to introduce also my colleagues Victoria Yan and Catherine Ahearn from the Knowledge Futures Group who will be helping with asking questions as well. So, I just want to flow over to Victoria now who will introduce our panelists.

Victoria Yan: Thank you so much Jessica. Just a reminder for everybody, this will be recorded and following that we will have the Q & A session. To join that later on, please feel free to unmute yourself and unmute on video. First we will have our speaker sharing the researchers' perspective. We have Dave O'Connor from the Department of Pathology at the University of Wisconsin. So, please tell us a little bit about yourself and your research on COVID-19 and how this particular research community is communicating their research during the current situation.

Dave O'Connor: Sure, thank you, Victoria and thank you to everyone who's listening in. As Victoria said, my name is Dave Connor, I run a research group here at the University of Wisconsin-Madison, and we have been involved in a large co-laboratory that we've helped set up called the CoVen which began back in the middle of January when we recognized that this was likely going to be a significant threat that was going to require significant collaboration between different types of stakeholders in order to do the best sort of research possible. My lab works on a couple of different things. First, we do animal model work, and animal models are really important because we can control the dose and the timing and the strain that we use, and we can follow animals longitudinally over time, but they're also restricted access meaning you need specific sorts of facilities in order to conduct these studies, and there aren't that many of them. And what that means is that it is a sort of both a privilege and an exclusive privilege in order to be able to the sort of work. And so, if we're going to design a study, we want to make sure we can get as much out of that study as possible.

To do that, you need a lot of different people with different opinions and different view points. We started assembling a group of people who worked either in animal models or clinicians or in different types of research - from cell biology to transportomics to neurobiology to aerobiology (the study of aerosols), and many, many more to think about how we should be designing studies collaboratively to get the most out of them. And we focused most of our attention on non-human primate and ferret studies so far, and we now have upwards of 100 people participating regularly in our calls and in our discussions.

To do this, we initially use Slack as our discussion forum and that has some real advantages: It's fast, it's easy; most people know how to do work with it right now. But I'll tell you, the downside to Slack is that it's not visible on the open web, and that actually bothers me because I worry that in five or ten years from now, a lot of these discussions won't be fully captured in a public way. I don't think we have a better alternative yet, but this is something that has me concerned.

We have twice weekly phone calls as part of this co-laboratory and that's been quite successful at bringing people to the table. Even if they can't make every call they can try to join it when it makes sense. And we've been able to start some of our studies, and then we followed what we did back in 2016 when the Zika outbreak emerged, which is that we're sharing our data from our non-human primate studies in near real-time at Open Research at LabKey.com.

So that means that we're putting raw datasets in Excel format, we're putting full summaries of our experiments, we're putting a narrative about what we're learning *as we learn it* and making that available because especially in non-human primate work, it's important not only to know what's going well, but it's also important to know what's *not* going well, because this is an exclusive, expensive, and ethically challenging space to be in, we want to make sure that if we make mistakes that no one else goes and repeats those mistakes, and the best way to do that is to make as much of that information available as quickly as possible.

So, that's one arm of what we're doing. The other arm is that we are involved in the sequencing of viruses. And that itself is also an interesting space because as many of you undoubtedly are aware, there has been a real effort, initiated in influenza surveillance, to make global sequences of viruses available and certainly the SARS-CoV-2 has benefited from that experience and the infrastructure that existed for flu.

 The challenge with that is that when you look at something like NextStrain, you see huge heterogeneity in terms of sampling coverage and where this data is available from. So you have small academic labs like mine contributing data alongside public health departments, which on one hand is a great example of grassroots collaboration, but it also suggests that possibly this should be something that has a little bit more structure to it. So for example, we believe right now that there is a significant

operate going on in many US cities, including New Orleans, and yet as of this morning there were no sequences in a GISAID from Louisiana.

So, the plus of self-aggregation is: it's great and it's grassroots and it's real time. But there are some real challenges in making sure that every affected constituency gets represented because if you have an accident of geography where certain communities are really well-represented you can end up with decisions being made that aren't reflected in some of the wider constituencies that are also affected, and as someone who spends a lot of time on HIV issues I'm particularly worried about what this means for the surveillance and tracking in sub-Saharan Africa and other resource-poor settings where there isn't going to be as much data that becomes available quite as quickly. So with that, I will stop and reserve any other comments for the question and answer period. Thank you.

Victoria Yan: Thank you for sharing that. Next we have Richard Sever who is a cofounder of the preprint server bioRxiv and medRxiv to tell us about how by bioRxiv and medRxiv are working with the large amounts of preprints submitted on the SARS-CoV-2 situation.

Richard Sever: Hi, my name is Richard Sever and I'm co-founder bioRxiv and medRxiv at Cold Spring Harbor Laboratory. I want to talk a little bit about the preprints that we've been dealing with at bioRxiv and medRxiv. And just to be very clear, by what I mean by a pre-print, I mean early sharing of research before it goes through and is certified by the peer-review process. This isn't like post-hoc archiving - this is the sort of like 'hot off the printer' papers, so to speak, written by scientists.

We run two prevent servers: bioRxiv, which has been going for seven years now, focused on biological sciences, which is now close to 80,000 papers and a new initiative medRxiv, which unfortunately it was very timely in its launch last year. It's a smaller server but it's focused mainly on the health sciences and clinical results. And the key distinction here being that for clinical research, there are additional concerns you have, so we have enhanced screening and ethics procedures.

So the background to why you would want to post a preprint is I think is exemplified by this data, provided by Stephen Royal, which shows the delay to dissemination that you get in the traditional publication process. The blue curve shows the submission to publication times of journals in PubMed, and you can see the medians about seven or eight months, and that range goes out to two to three years. But if you post on bioRxiv your paper can be read within 24 to 48 hours, typically. So that's a huge time-saver to dissemination and obviously in the aspirations that this translates to discovery and quite the bioRxiv hub in San Francisco has done some back-of-envelop calculations suggesting that if you can get everybody to post pre-prints, you could speed up scientific discovery five-fold in 10 years,

because you have the aggregate effect of all those time-savings, which would essentially have a geometric effect.

This is potentially far, far more important in a pandemic, when you really want communication and discovery to happen as fast as you possibly can. If you compare the 2003 SARS epidemic, 93% of the papers written about the epidemic appeared after the epidemic had ended. And you contrast that with the SARS-CoV-2 pandemic in 2020, and we just posted the thousands of pre-print on bioRxiv and medRxiv, in the midst of the pandemic, which is a really incredible contrast between SARS 1 and SARS 2. These papers span the basic virology, the molecular biology of the virus, structural studies of the proteins involved for example, to immunology, epidemiological study models of the R0, Public Health, and we're now beginning to see reports of drug trials. So it's the whole spectrum of academic research in this area.

This just shows what has happened since January across bioRxiv and medRxiv. The lighter bars are the medRxiv posts and the darker bars are bioRxiv. So you can see none until January and then we're getting probably on average about 30 papers a day, but I think you can look at that distribution and say it does not make sense to talk about an average. Yesterday, we posted about 100 papers in a day on medRxiv, and 100 papers is about what we generally get across all fields in bioRxiv and yet today we had 100 on medRxiv on COVID-19 alone. It's quite striking the distribution of these papers. And among them, there's some very important work. A couple, just to highlight in basic biology, you see this just came out recently from Nevan Krogan and Jaime Fraser and a bunch of other people looking at the SARS-CoV-2 protein interaction, which is a very important paper that gives a lot of starting points for development of new therapeutic targets. Meanwhile on medRxiv, you see papers like this on serological ways to detect seroconversion, which is obviously something that now we want to have as fast as possible so that you can identify people who have had the virus and hopefully have some immunity to it.

One of the things that has changed in the course of the pandemic is the amount of attention that these papers are getting beyond the normal readership. So this is just another example from medRxiv: this paper looked at the stability of the virus on different surfaces like cardboard, metal, etcetera, and you can see that this was picked up by 300 news outlets. These numbers a little bit out-of-date, but the PDF alone has already been downloaded more than 600,000 times. These are not all scientists so this is something to think about.

When we look at this, this is the Rxivist, a third party site that ranks papers by the number of times they've been downloaded, you can see the top ten papers on bioRxiv are all COVID-19 related papers. So we're doing a few things in the midst of this pandemic to change the way we handle things. One of the things we want to do is increase discoverability so we've created this summary page where you only see the COVID-19 related preprints. These are actually annotated by individuals; early in the

pandemic, the word COVID-19 didn't exist. So these are actually COVID-19 papers even before it was called COVID-19. So no, this isn't a search result, these are manually selected papers.

And then one of the things that you need to think about in the midst of the pandemic again is are there things that you should think about a little bit differently? We have general criteria at bioRxiv and medRxiv including, obviously, we've don't want to post nonsense pseudoscience plagiarized work. You may miss something under the normal circumstances, this doesn't really matter. Does this matter when there's increased public attention on preprints? This is something we think about a lot. The do-no-harm, considerations we've had from the inception of medRxiv like being cautious about dual-use research, vaccine safety, disease transmission and toxicity claims. We wouldn't want to post papers that said cigarettes don't cause cancer, or that vaccines cause autism, but there are other things that we should be worried about in the course of the pandemic when they're may be many more eyes and more bias from the general public. So this may come up in the course of conversation. Conspiracy theories: is that something that, if people willfully or accidentally misinterpret discoveries, is that something we should be worried about? Drug-availability: I think we've all seen what's happening with chloroquine. I suspect that the president suggesting it's an effective cure probably does more harm than a medRxiv saying that, but it is something that one needs to think about.

So, in light of some of these concerns, we have made some changes. We now have enhanced declarations on bioRxiv, which now make it much, much more clear that it's a preprint and what that is. Because of COVID-19, members of the general public are looking at preprints much more than they did before. On medRxiv we've always had a very stringent screening process, but because of some of these concerns we've introduced a dedicated COVID-19 work flow on bioRxiv, so that papers go through a dedicated group of screeners who are kind of thinking about these questions that we might have. And so, there are a large number of papers than normal that we will return to and say we think that this paper should go through peer review first in some cases.

One of the things we do want to do as well as to encourage preprint commenting. This is the now infamous, uncanny preprint about the virus, which led to a number of conspiracy theories, which I should stress the authors were keen to point out that they were not themselves conspiracy theorists. And then maybe some complained about that, but the interesting thing that happened because it was a pre-print, was that within hours of the preprint being posted there were many, many comments by renowned expert biologists around the world pointing out the flaws in this paper — we receive 50 lengthy comments within one day.

The offers then apologized for some of the inferences that they might have to make people make and the paper was formally withdrawn within two days. And so I think in many respects that was an unprecedented speed of self-correction within the academic community. But it is something to think about in terms of how we examine and peer review these papers. As well as encouraging commenting,

we honestly do want to link out (as we do with all papers on bioRxiv and medRxiv) to discussions amongst scientists about these papers. So we have track back links to blogs, preprint discussion sites, Twitter conversations about the papers, and things like TRiP, the new initiative from eLife and Review Commons to review preprints much faster than normally. But I think it behoves us as a consequence of this to think about what peer review should look like in a pandemic. So this is a slide that I took from another talk that I gave where I talk about what do preprints mean in terms of how we think about peer-review when we think about which papers we do it, who does the peer review, and when?

And I think a pandemic does pose these questions and I think we really need to think about preprints this is really the perfect time to hand over to Daniela because I think she's going to talk about one of the ways that we might do this, so thank you.

Victoria Yan: Thank you so much, Richard. So yes, this is the perfect segue to introduce our next speaker, Daniela Saderi who is a co-founder of PRE-Review and also co-developed Outbreak Science, which is a platform that structures rapid peer review of outbreak related preprints, which happened to very timely will launch last December. So, please tell us a little bit more about your work Daniela.

Daniela Saderi: Thank you so much for organizing this event first and foremost and also for inviting me to speak today, to Jessica and everybody else, and thank you to the audience for being here. I'm Daniela Saderi and I'm the co-founder and Project Director at PRE-Review, which operates as a non-profit through fiscal sponsorship of Code for Science and Society. And in the past year and a half, we've being collaborating with another non-profit called Outbreak Science to develop new tools to allow for the rapid review of preprints in the context of an outbreak. Little did we know that when we launched we would be in the middle of a pandemic, but I also want to mention that Michael Johansson, the founder of Outbreak Science and our collaborator was supposed to be here, unfortunately he couldn't make it, but I want to note his contribution from the get-go.

Okay, so full disclosure: I am not an outbreak scientist in any way. My training was in neuroscience, but when I first met Michael, who is actually a legitimate outbreak scientist, he showed me some of the plots that kind of convinced me even more than what we were doing at PRE-Review was very important.

These are plots from the *New York Times* showing the number of new cases each week throughout 2013 and 2015 outbreaks of Ebola in three African countries. And if we align to the same time frame the number of publications that came out in response to the outbreak, we can see that has shifted over time. And this is something that I'm stressing more, but Richard already covered. Presumably, all of this research has not really contributed to the response to these outbreaks - at least when they happened in the peak of outbreaks in those countries.

And what Michael and collaborators also showed in an article that was published about the Zika and the Ebola outbreaks is that the preprints that were posted were actually out more than 100 days before the publication of the same manuscript that went through general organized peer review.

And, however, if we look at the number of preprints with respect to the number of publications - I think this was a combination of Zika virus and Ebola, I'm not entirely sure, it was about less than 5%. So what was happening is that the number also of preprints that came out were very slow but nonetheless were very early.

What we are assisting on today, and again Richard has shown this, that the number of preprints related to the Coronavirus pandemic is is unprecedented. This plot expands a little bit on the bioRxiv preprints. We see that from January 15th, every week we have an increased number of preprints that are posted online across different servers with medRxiv having the biggest contribution.

And so what this means is that we have this incredible amount of information that is coming out with the potential to really speed up science and discovery.

What drove us towards building these tools - Outbreak Science rapid peer review - is that we wanted to provide an extra layer of feedback that could be rapidly implemented during an outbreak by scientists who are very busy. And so, when we were designing this tool starting last year, we convened a group of researchers together and were like, "Okay, so how can we come up with a structured review that could provide a certain level of information feedback rapidly in the midst of an outbreak?"

And so, on rapid review, readers can read the reviews that others have made of preprints. Because we launched in January, most of the preprints that have been commented on are coronavirus-related preprints. The researchers themselves can fill out these rapid reviews and they can also request feedback, which I think is a very interesting feature that we would like to bring more to the user of the other researchers so that we can see which one which one of these preprints actually needs more revision.

So this is a screenshot of the platform and we have ORCID IDs ready to log-in so researches with an ID can login and can request or review themselves preprints by coping and pasting DOI or an arXiv ID from a preprint of interest and this covers several preprint servers. We have a search bar that can allow the user to search through the content of the platform, which would be preprints that have already been either requested for review or have been reviewed themselves. And infectious diseased tagging including the 2019 coronavirus.

The review itself displays on top of the preprint itself and it's a series of 11 questions with two optional open-ended questions and the combined reviews of different authors, in this case three, can be displayed in this visualization. The idea is that it could provide a quick understanding of the feedback

aggregated across different reviewers. This works both on the website but also in the context of the preprint servers themselves, if the user installs the available extension for Chrome or Firefox. So while they are surfacing or reading through different preprints on medRxiv for instance, or any other preprint server, they might see the number of requests that the preprint server receives or see the number of reviews, and open the window themselves to review or request to review.

We have a public open API that we're hoping other websites and other stakeholders can use with our collaborations to integrate this tool such that it doesn't require an installation of the extension, and we also are trying to really in the context of this outbreak to outreach to researchers. We understand that they are really, really busy, so we made a call to action for an outbreak scientists to rapid peer-review three coronavirus preprints on the website, from their extensions.

We are also trying to organize public conversations on Zoom. So like this one, and we have done this before, different preprints and other topics that in this particular case, we're speeding up this process and partnering up with JMIR Publications and PLOS and hopefully other journals coming up, to organize preprint virtual journal clubs that are hosted by associates that we pair with experts that can guide the discussions around an emerging reprint on the coronavirus topic. And so, if you click on this form, we are currently asking and crowdsourcing what preprints should discussed.

So if you're interested, please pitch. On April 14th we will have the discussion which anyone can join and also, anyone can request our support for these discussions. There is also another form. We invite everybody to do that. It's free and we are happy to help with all the logistics.

 So I want to thank the advisory committee at PREreview: Sam Hindle, Monica Granados from the Leadership Team, Naomi Penfold, Lenny Teytelman, as well as all the contributors for this project. Michael Johansson from Outbreak Science, Georgia Bullen for help with all the user-research and reviews, and Sebastien Ballesteros for developing the actual tool, as well as funding from the Wellcome Trust and you all for your attention.

Victoria Yan: Thank you so much, Daniela and to all the speakers that have spoken so far. Next I am going to introduce Catherine who will lead our round table discussion with some questions.

Q&A Session

Catherine Ahearn: Hi everyone, thank you so much for your talks. We will kick off this portion of the Q & A to bring in some other points and some of the attendees with their questions. Just a reminder to the attendees that this session is being recorded, and to make a note: if you're asking a question, whether or not you would prefer to ask yourself or if you'd like us to to posit the question to the speakers. I've been following some of the Q & A that's been going on in the comments so far, and it

seems like a lot of the questions really revolve around weighing the need for collaboration, openness, transparency, and speed at this time with the potential danger of publishing something that's inaccurate or harmful. And so I was wondering if more of you wanted to weigh in on that and maybe talk more about what precautions you've taken to balance that as much as possible.

Richard Sever: So yeah, I think it's an important thing to think hard about, as I mentioned in the talk, now that there are more eyes on preprints. I think it's also something that one has to put in perspective. This is an issue that we have with the web, right?

So, this exist anyway. I sort of joked in the past that I don't think that the harm that something like medRxiv could do is anything compared to the bombardment of misinformation on the web as a whole: direct-to-consumer advertisements that I see whenever I watch the sports game, this kind of thing. So I think one has to think there's a much bigger picture here about misinformation and so I think being transparent about what things are and what they aren't, is absolutely critical.

So the chloroquine paper, for example, people were asking whether that paper should be on medRxiv. Well, the important thing to note there is A) the information was already public, anyway, because the author had put it on their institutional website, it was then posted on medRxiv, and then very shortly afterwards it appeared in a peer review journal. There's something vague going on about the extent to which the peer review actually occurred but it did appear.

So this goes back to the notion that if you close the door to something in your little area, that will not enable the information to get out anywhere else; this is somewhat naive. So I think what you need to do is think about what's the most responsible process that you can have in your own screening, and also what you can put in place to inform people about scientific research, and the occasional fallibility of it.

It's not just preprints. We've seen a number - I can think of three or four papers - that have appeared about COVID-19 in the peer review literature be they letters or peer review papers in medical journals. So we have to have this conversation, and actually Kathleen Hall Jamieson, the author of *Cyber War,* and myself, Veronique Kiermer from PLOS, and Marcia McNutt, the president of the National Academy of Sciences**,** wrote an article in [PNAS](#) recently, which I would urge people to look at, that talks about the need to have signals around trustworthy information and the importance of people understanding that lack of a signal is itself a signal.

So I think for preprints, one of the things that is important is to explain that this is preliminary. It's primarily intended for experts and it needs to be discussed, and try and have that discussion publicly in in a transparent way, and I think that's the way forward. And actually, I've been pretty encouraged with the way that most journeys are handling this information.

It was interesting with the... With the uncanny paper, I was kind of internet conspiracy theorists who were spreading mis-information journalists themselves. It was amazing how few of them picked up on that story. The journalist then covered the story about the misinformation. But so I think I've been very encouraged. I saw Carl Zimmer from The New York Times tweeting a paper on medRxiv the other day and underneath his tweet in brackets he put with the caveat, "this hasn't been peer reviewed yet." So I think transparency is important here, and we just have to understand you can do things fast or you can do them thoroughly. And there are pros and cons, but for the most part in the midst of a pandemic, I think most people are pretty clear that the pros of disseminating information in this way, far, far out weigh the cons.

Dave O'Connor: I'm a big fan of medRxiv, but I have to say that I find a little bit of that framing a tad disingenuous, if I'm being honest, because there's a lot of attention seeking that you use in the metrics that are at the bottom of each paper. So it's hard to say that it's really for a technical audience first and foremost, when you have media mentions and blog posts and things like that as part of the way that each paper is presented to the audience. And so, I guess I would question is that really the right framing for how a paper that's in bioRxiv should be accredited for getting its attention.

Richard Sever: I think there's a couple of things there. One of the things I should point out is we do put the metrics there. The metrics are generated by third parties, the attention metrics, they're primarily in part for the benefit of authors because authors like to have, and are encouraged increasingly by their employers to provide, metrics beyond just the impact factor of their paper. So we can have a big discussion about whether the all metrics are a good or bad thing in general. One of the things that we do is: we have never ranked papers like that. If you go to bioRxiv, we don't have a 'this is the most read paper.' We don't rank in any way. Those metrics are primarily for authors to use as evidence of their productivity and attention on their work. I'm not convinced that many people really pay too much attention to them. Where they are valuable is that we seek to aggregate all conversations around this work.

So that's why we aggregate the tweets and the links to blogs, etcetera, is so that you can go to a paper and if it's not being peer reviewed, you can see that somebody on Outbreak Science has written something about it, there are three blog posts by virologists about it, it's been covered in this news, so that you have a way to see the conversation happening around it. So our motivation for all those metrics is not to trump anything. If it were, then we wouldn't put them on one page; we'd create additional pages and we'd be tweeting the hell out of them and things like that. We wouldn't bury them where they're actually quite hard to find.

So yeah, so I don't think I would agree with that, frankly. It's a difficult one, as I say, there's a debate about metrics, but authors love the fact that they can get evidence that their papers being downloaded and being looked at. There's a reason why everybody's obsessed with their H-index and all these things. So I think the idea that we can get away from these types of numbers is very difficult, particularly when we're in the midst of a conversation about the fact that current examinations based on a single number - the impact factor or citations - and people think that's a bad thing. We need to have much more multi-dimensional measures of impact. And if you want to have those things, you can't have them and not have them for the reasons that you're saying. Because one of the things that you mentioned is one reason why we don't make a big deal about them: they're not highly privileged on the site in terms of visibility.

Catherine Ahearn: I'm a bit struck at how quickly this conversation veered into two things that are just foundationally flawed within the scholarly communications landscape anyway, even if we weren't in the middle of a pandemic, and I was thinking about this earlier actually, when you were speaking Richard, you showed a slide that said, "What should peer review look like in a pandemic?" And you had 2020 crossed out. So, this is again a question for everybody, what are really the differences between, in your opinion, what peer review should look like in a pandemic, versus in 2020 or into the future? What are those things that we can change that maybe we're realizing for the first time or realizing more clearly now that we're in this crisis, but that really are indicative of much larger problems that were there all along?

Richard Sever: Well, let me just say one thing here, and probably Daniela is really the person to talk about this, but I think one of the things that is interesting in in the context of a pandemic is this big question about what happens to papers after their preprint? The numbers for medRxiv and bioRxiv are pretty consistent: if you look after a couple of years, 70% of papers end up formally published in journals, so if they're preprints, most of the time they ultimately become journal articles. Of course the 70% is probably a little bit low because you get false negatives because titles and author lists change and some papers can take five years to end up in journal. So, that said, it's closer to 80%. That means that 20% of papers aren't ending up in journals. Now, what I'm kind of interested in is in the midst of something like a pandemic, what does that percentage look like if it takes eight months typically to end up in a journal, and you're writing a paper about the state of an epidemic in Wuhan on January the 30th. Is there any meaning to having that paper formally published on December 1st of 2020 in the 'journal of epidemiological whatever'?

So I think that one of the things that's kind of interesting is it forces us to have that conversation. We've always known that there's a small number of papers, there's a category of papers for which traditional peer review doesn't make sense, and as a consequence of a pandemic, it's shines light on

the fact that there may be a number of papers for which the traditional peer review doesn't make sense. I would argue that something like the approach that Daniela's talking about, which is much more real-time, much more about trying to do this continually and quickly, makes sense. And so what does that kind of 70% figure look like in the context of a pandemic, and what does it mean for us thinking about what peer review should look like in future?

Daniela Saderi: Yeah, thank you Richard. I would also like to pitch in to say that, it's really true what you presented in the last slide. Also, the questions of 'who is doing it, when are we doing it a lot, and what' are really the things that have motivated us, and the peer review before even we started with this. Thinking more about the context of an outbreak and how peer review works, and changing the reason why we really try to engage all researchers into these conversations with the idea that obviously you have to try to have mechanisms to build trust and moderate a bit of the conversations. But what we would love to see with both peer review, and then with the more rapid layer of reviewing in the context of outbreaks, is that those comments and those reviews can then speed up the peer review process itself. And so there are a few journals that are right now trying the model in which they do use comments from trusted platforms to speed up their own process. And I have to say that I have seen a lot of quicker peer review processes happening right now in the context of a pandemic. There is definitely no comparison with how quickly someone can post a preprint, but I think that there is a workflow that I can see, and as a kind of addition to this process, I can also see the advantage of having multiple voices and multiple eyes on the same manuscript rather than having two or three other viewers selected over a much slower timeframe, obviously by journals. So, what we're also trying to do with virtual discussions is to bring together those discussions and create a review that can be then used by the publication system itself to speed of the process. So we at Peer Review would definitely love to see more participation by a more diverse group of peer reviewers, and I think in the context of the pandemic, we're trying to reason through how busy are the researches themselves, and how can we make this process so that they don't have to spend a week writing a report?

 And the last thing I want to say, by having multiple people commenting and giving expert feedback on a preprint, we can also break down what peer review means. We don't need to write an entire report. It could be that a person who doesn't work with outbreaks but has experience with modeling can comment on just a section and so then when we puzzle this together, we have a very robust crowd-sourced scientific feedback on it.

Richard Sever: So I just wanted to follow up on something Daniela said, because I think she makes a very important point about the diversity of representation in the review. In scientific publishing, we simultaneously have these scenarios where 1) Everybody complains about review burden and everybody's got 'review exhaustion' and it's harder and harder to get papers reviewed, but

simultaneously, 2) we have data coming out that the massive under-utilization of vast swathes of the potential peer review population, like the Global South is hugely under-represented: most peer reviews come from North America and Western Europe. So we are complaining that we haven't got enough reviews while ignoring this big population of potential reviewers. We really don't utilize our postdocs anywhere near as much as we could. I'm not necessarily saying that we should, but that is a population of experts who are largely not called upon. And one of the things that I had a discussion with somebody recently in the context of the COVID-19 pandemic, is that there are lots of labs that have had to shut down. I'm sure that the PIs (primary investigators) are still trying to do everything they can and probably have less time on their hands to do the peer review, but there's gonna be a lot of postdocs who would have been spending X hours per day in a lab, and now aren't. So there's a huge potentially under-utilized brain trust there for peer view within a pandemic and maybe that could potentially give us some lessons for moving forward beyond the pandemic.

Catherine Ahearn: Yeah, thank you for that note. I also just want to interject here briefly to thank Richard Wilder for joining. I'm told that he has to drop off at 1 o'clock, so I just wanted to pause briefly and say thank you, and we appreciate the contribution.

Actually, I wanna pull in Jo from AfricArXiv who posed an interesting question in the Q & A. I will unmute you Jo to allow you to talk. Do you want to pose your question to the panelists and the attendees?

Jo Havemann: Sure, thank you. I'm Jo, co-founder of AfricArXiv, a preprint repository, which was also in the slides by Daniela for the African scholarly community — so region specific. My concern was, we also put together in the preprint that I shared in the discussion on the Q & A is that African scholarly research output is hidden and basically invisible on the western landscape, so to say, on search engines like Scopus and the ones we use to search and scrape information. But I think with the platforms like PREreview and other digital tools, we now have a chance to really make this a global effort and also across language barriers.

So I'd like to ask the panelists if you've seen initiatives to address this, in particular in the Global South, not only scraping the back of literature that's available on Ebola and Zika and other epidemics that keep crippling the continents or washing over the continent, which the rest of the world's ignoring over and over again - but now this is a global effort. Africa has a lot of lessons learned to share and they're ready to share it if we as westerners are ready to listen. I think I'm a bit emotional about this because all of us, I've been working for three weeks nonstop, and again, I can only say: please have a look at this preprint I've shared and I'd like to collaborate on all possible levels. Yeah, if anybody wants to jump into the discussion.

Daniela Saderi: Thank you for that question, yeah, so I think that this is definitely a really important point and the fact that we at PREreview have tried to put at the center of our mission to engage under-represented communities in scholarship and the context of peer review of preprints.

And I think that you're completely right. What we can control from our side is that we have several things in place that we're trying to organize in-person, but at this point it will be a lot of virtual events, to highlight the contributions of communities of scientists that are under-resourced and under-represented so partner with with AfricArXiv is something we have in the pipeline; we have some funding specifically for that. So I hear your call and I think that we really need to, especially not just in the context of an outbreak, but there is so much to learn from the research that happens not just in the Europe and the US. When it comes to peer review, just changing and shifting this needle to who are the gatekeepers of scholarly evaluation and publication is really important, and if we don't change that, there is really no point in moving towards open. And so there is a whole discussion of what open transparency means without actually addressing the accessibility problem and who is at the table to participate. So I don't have an answer to you, but all I can say is that at PREreview all I can say is that we're really trying hard to make that a center of our mission.

Jo Havemann: Yeah, thanks very much and I'm very much looking forward to collaborating. This question also goes to Richard with medRxiv and bioRxiv, and all the data that you have, which is also, to quite a decent extent, contributed by African scholars. I always encourage people to share your research output wherever it seems feasible for yourself, be it AfricArXiv, bioRxiv, medRxiv, whichever archive... The important thing is to make it accessible and I think it's as platform providers to ensure that the research output is discoverable across platforms. I think we already have some tweaks and solutions around this, of making our services interoperable and the question - I wrote you guys an email in the course of the webinar - so please, let's be in touch.

I don't have a specific question other than the ones that I've posed but I think it's important for the stakeholders to come together to provide all the possible services to the community, to make the endeavor of bringing the literature to light that's currently hidden — just the sheer number of research being private and we need to make sense of it in an inclusive way, and also across language barriers. So I'd also like us to focus, not so much on English but of course primarily, but also look into other language systems and then across the continent because it's a global pandemic, so we have to also cater to other languages.

Richard Sever: I think what's encouraging here, in terms of the discovery, is the emergence of the third party discovery and indexing tools. I think that historically on the biomedical side of things,

there's been PubMed, which has a very traditional way of doing things and aggregating content but the emergence of things like Meta so many people using Google Scholar, and I was just talking to a group yesterday from CiteNet that seems to be like a whole bunch of new literature discovery tools which all have flexible ways of deciding what they do and don't index. I know Center for Open Science indexes all the stuff on the Center of Open Science posted archives, as well as bioRxiv archive and medRxiv. So I'm encouraged there. There's been one single root for people for a long time in discovery through PubMed, but actually in a lot of fields people use Google Scholar more and I really predict that, as people get better and better at using AI and other tools to give you push alerts to the kind of content that you're interested in, then those third party services will be able to make decisions about what they index - the universe that is indexed by Google Scholar is somewhat different from the universe that is indexed by Semantic Scholar, that's somewhat different from the universe indexed by Microsoft Academic. And I think we'll see more things like this and with the deluge of information from people, there'll be some kind of competition for these services to give you good alerts. And so they will want to index things like AfricArXiv as well as bioRxiv, and as well as traditional journals. So I'm hopeful that that's a good way of of surfacing some of this information on the language issue. That's a challenge, that Daniela and I've had conversations about this very, very recently. You probably will have seen the paneling initiative from Humberto Debat and Richard Abdill which is basically using Google Translate to translate sites so you can apply that to bioRxiv, medRxiv, or any other site.

I think that the concern I have, and Daniela knows this because we were in a meeting talking about it, is that in some ways, you worry with automatic translation services because their translation efficiency is based on the popularity of the number of documents in that language. So, the worry is that you end up only translating into languages where people already speak quite good English, and the under-represented groups, there aren't good translation services because there aren't enough training set for those languages. I mean it's a challenge the scientific community faces as a whole, with English becoming the *lingua franca*, it's difficult to know how to solve it for the people who need it most, which is the languages that have the least attention on them right now.

Jo Havemann: The solution is already there and we also described that or address that in the preprint we shared. Basically, there must be a mix of manual labor and manual translations mixed with artificial intelligence, and better than the ones provided Google Translate because there are better translational services out there, and we need to liberate those to make research more accessible and translatable. But of course, we also need a proof check, but the check point for quality assurance of translations is also needed for peer review, there are misconceptions within English because we're dealing with non-native of the speakers of the scientific community with *lingua franca*. So yeah, the concern is an expert challenge but it's one that can be addressed and that needs to be addressed. Let's please keep talking about these things because we have solutions at our fingertips; we just need to utilize them.

Catherine Ahearn: Thank you, Joe. I also wanna bring in Sarvenaz Sarabipour. You had a few questions about the use of preprints by funders and the long-term effects of the increased using cultural shift toward preprints during this time. Would you like to come in and ask your question to the panel?

Sarvenaz Sarabipour: Yes, I have three questions. Basically, I was talking that funders would pay more attention to this outbreak and how preprints are used and wondering whether funders have reached out to the preprint servers and other conversations between major funders and servers on the future and financial sustainability of the service.

My second question was, on other research of course medRxiv came recently and I think it had a few hundred preprints before the pandemic, but I'm wondering if other wider research in cancer and diabetes and cardiovascular are maybe preprinting faster.

And the last one was that of course, we see this wonderful trend of academic research labs trying to bring the basic research to applications such as the diagnostics or therapeutics, and I'm wondering in pharmaceutical research would, if you anticipate them to start picking up the trend and become more competitive by potentially preprinting because there's massive money spent behind close doors in those companies, and they usually publish late and I don't think they're usually preprints. So thank you.

Richard Sever: I think with funders the short answer is yes, in terms of appreciating preprints, you've had statements from for example the Welcome Trust mandating that science in the course of the outbreak should be hosted on preprint servers. Of course, Chan Zuckerberg initiative have a similar mandate for their investigators, and the Michael J. Fox Initiative as well mandating preprints. , I would love it if all funders would do this. Michael Eisen, John Ingalls, and I wrote a paper in PLOS Biology about this which we call *Plan-You* suggesting all funders shoudl do this to achieve immediate free open access research. So yeah, in terms of funding, bioRxiv is funded by the Chan Zuckerberg Initiative for which we're eternally grateful and they're great, we only have good things to say about the funders at this point. You also mentioned other areas, I think in clinical medicine, you you mentioned cardiology and some of those areas, it's a new experience for clinical researchers in the way that it was for biology researchers when we launched this in bioRxiv in 2013. I think what will be interesting to see is whether the awareness of preprints as a consequence of COVID-19 means that you get an earlier adoption then you would otherwise have had.

In physics, computational science, and other areas on bioRxiv, you see these waves of adoption in different communities and so it will be interesting to see whether some of those waves in infectious disease, for example, and epidemiology start coming higher and faster because there's greater awareness of preprints. And then your final question about pharma: so there is a group could Open Pharma that are trying to push dissemination of preprints in the pharma. I think there is some enthusiasm for some things, but on the other hand, commercial companies have secrets and so getting them to divide the secrets earlier or at all is a different thing. But I think I've definitely heard enthusiasm for pharma among pharma representatives for early dissemination of research as preprints that they would normally disseminate as research articles, I think that makes sense. So there are some legal concerns that they have about claims that they can make, but there's enthusiasm there, but the question of whether they will start divulging stuff that they never normally don't divulge, I think is a different question.

Daniela Saderi: Super quick - just for the funder question. I also want to answer another question that is tangential, so our tools have been funded by the Open Research Fund at the Wellcome Trust, we don't currently have funding to continue, so this is more of a pitch to funders if they're here! But I think it is important to also look at what are things that the funder should put attention to, and definitely like preprint servers and other initiatives should be relevant obviously in these contexts. Thanks for that question.

Catherine Ahearn: I have a question, actually maybe Dave would be in the best person to answer this. I'm wondering about other research or communication tools that researchers are using aside from preprints, even before you get to the point of publication, are there ways that people are communicating that have come into play more so in these times of the pandemic or even the decision-making process around what modalities are more appropriate based on the types of things that you're working on or communicating about?

Dave O'Connor: Yeah, so it's a lot of this actually. Even before the pandemic, I would typically be on go-to-meetings or Zoom calls, three to four hours a day with people from around the world. Part of that was because I was living in Australia last year and initially had to have a lot of calls off timezone with people who were back in North America, so so that's probably ramped up to five or six hours a day. It's really been helped as some of the video conferencing and screen sharing tools have gotten better so

that you can share something on your screen with an audience of people. And that's become a very quick way of having conversations that often end up getting enhanced with screen sharing.

Catherine Ahearn: Thank you, anyone else have experiences with communication tools?

Daniela Saderi: Yeah, so the visual journal clubs that we have run before were all conducted on Zoom, and I have to say that our team has learned a lot from working with Mozilla groups on that. We are currently developing a peer review mentoring program that is entirely based on one-to-one pairing where the experts and the researchers are designed to be virtual and cohort based. So for us, we're really grateful that these tools exist and we are also grateful that more people are using that so that we can have more participation in our virtual journal clubs, and hopefully the peer review mentorship program as well.

Catherine Ahearn: What is your take on how these tools and the ways that research so far has been communicated has really shaped the public's understanding of what's going on and the conversation discourse around it, like if not for preprints and some of these other ways of communicating you guys have mentioned, how would this be different, how would this look different, and how has this really kind of shaped the conversation around what's happening?

Dave O'Connor: Well, I guess I can try to venture answer. I think the public discourse is maybe not as relevant here as the more technical people who are working within a particular field. I've often believed that scientists think their data is more important than it really is, and it makes sense. You think about it all the time, it's your thing, but the reality is you can post your data online and while there's going be some number of people who are going be voyeurs, and who are going to look at it, and are going to be genuinely curious versus 'Oh, I heard this out there I'm gonna click on the link and check it out," it's really a small number of people who are probably gonna be able to benefit from it in a meaningful way and use it in their own research. And I think that we have a very important role in outreach as scientists and as educators. But I would divorce that from the role of putting stuff out in preprints and in our open data sharing. It's fine that people are looking at it, but really that's not the intended audience. The intended audience are the people who are going to be able actually use it in their day-to-day live.

I put on a very different hat and I communicate very differently when I'm doing outreach or education to an audience because I think that you end up with these problematic scenarios where you try to

conflate the two, as evidenced by the chloroquine and the HIV origins of coronavirus. It's a lot easier for something that's inflammatory and provocative to get attention, even if it's poorly understood. Another example would be aerosol versus droplet transmission or fomite transmission of the virus. And I actually view that as almost more of a risk than a benefit at the moment.

The real benefit to me is that I can learn what other people are doing faster and can incorporate that into my thinking and other people can learn from us faster and incorporate it into their thinking, and that is in the issue. To be very blunt about it, I spend a lot of time telling people why some of the stuff that they have misappropriated or understood out of context from some of the preprints don't fully capture what is known about a particular topic and that's not, as Richard said, a feature unique to preprints the same is true if something showed up in a peer review journal. And while I'm against pay walls, the fact is: people have tended to generally ignore things that show up in the Journal of Virology whether they're in PubMed Central or not; people are not ignoring things as much when they're showing up in medRxiv and bioRxiv, and they're getting tweeted around the world really quickly. For better for worse that's just where we are.

Richard Sever: I would agree with what Dave said. It's getting back to this point about the bigger picture of journalism and things like that. There's a joke that Patty Brennan once told at a conference where she said, "data are like pictures of people's kids: the parents think that they're really important and everybody wants to look at them, but it's not clear that everybody else really does." Yet the intended audience for these things is pretty small most of the time. And so, when we think about scientific misinformation and the way to convey credible information, we should be thinking about the people who are aggregating, condensing, and conveying things like journalists and public information campaigns. I had a discussion once with somebody a few months ago, where I was quite blunt where somebody said, "Oh it will be really good if there was a lay summary of every preprint on bioRxiv," and my response was, "I really can't think of anything that would be more of a waste of time. It would take a huge amount of time and effort to get somebody, who would have to be good, quasi-professional writer to do something like that for 7,000 pre-prints."

What you want is a good journalist like Carl Zimmer as I mentioned earlier, Ed Yong, and people like that to reach out to scientists, consult them about the bigger picture, as Dave mentioned, to synthesize this for public and understanding, and what we need to be clear about is that there is this spectrum of information: from highly technical information all the way through to the article in the New York Times by Carl, and explaining to people what these things are. So, explain this is a preprint and that it hasn't been peer reviewed, explain that this is in the Journal of Virology but it's been peer reviewed by three people two years ago before it was published and may still be wrong, but probably won't be. And how different that is from news and views that came out in Nature or an article in *Scientific American*

or in the *New York Times*, and being transparent about all these things and educating people about the missions of these different things, I think is the way forward.

Jessica Polka: Alright, so if there are no further comments on this issue, I just want to take a moment to thank everyone who has participated in this webinar, especially our panelists Dave Richard, and Daniella, as well as all of you attendees for weighing in with your questions and comments. This has been a really rich and interesting discussion.